JOURNAL OF

Personality AND Social Psychology

VOLUME 37 NUMBER 1 · JANUARY 1979

Stressful Life Events, Personality, and Health: An Inquiry Into Hardiness 1
Suzanne C. Kobasa

Executives who stay healthy under stress are distinguished from those who fall ill by their personality characteristics of hardiness.

Interpersonal Attraction in Exchange and Communal Relationships 12
Margaret S. Clark and Judson Mills

Effects of receiving a benefit and a request for a benefit differ depending on the type of relationship.

Person Memory: Personality Traits as Organizing Principles in Memory for Behaviors 25
Reid Hastie and Purohit Anand Kumar

Effects of Amount of Helplessness Training and Internal–External Locus of Control on Mood and Performance 39
Nancy L. Pittman and Thane S. Pittman

Source Credibility in Social Judgment: Bias, Expertise, and the Judge's Point of View 48
Michael H. Birnbaum and Steven E. Stegner

Self-Consciousness, Self-Attention, and Social Interaction 75
Allan Fenigstein

Multidimensional Reaction to Therapeutic Touch in a Hospital Setting 87
Sheryle J. Whitcher and Jeffrey D. Fisher

Effects of Message Repetition and Position on Cognitive Response, Recall, and Persuasion 97
John T. Cacioppo and Richard E. Petty

Effects of Controllable Versus Uncontrollable Factors on Responsibility Attributions: A Single-Subject Approach 110
Daniel Arkkelin, Thomas Oakley, and Clifford Mynatt

Marital Roles, Sex Differences, and Interpersonal Attraction 116
Joseph E. Grush and Janet G. Yehl

Simple Minitheories of Love 124
P. M. Bentler and G. J. Huba

Statistically Combining Independent Studies: A Meta-Analysis of Sex Differences in Conformity Research 131
Harris M. Cooper

Ad Hoc Reviewers 146

Correction to Stephan and Rosenfield 47

Notice on Author Alterations 38

Published monthly by the

AMERICAN PSYCHOLOGICAL ASSOCIATION

ISSN 0022-3514

Editor

Anthony G. Greenwald, *Ohio State University*

Topic Editors

Irwin Altman, *University of Utah*
Robert A. Baron, *Purdue University*
Daryl Bem, *Stanford University*
Eugene Burnstein, *University of Michigan*
Robert B. Cialdini, *Arizona State University*
Reid Hastie, *Harvard University*
Douglas N. Jackson, *University of Western Ontario*
Leslie Z. McArthur, *Brandeis University*

Consulting Editors

Icek Ajzen, *University of Massachusetts—Amherst*
Robert S. Baron, *University of Iowa*
Gordon Bear, *Ramapo College of New Jersey*
Ellen Berscheid, *University of Minnesota*
Michael Birnbaum, *University of Illinois at Urbana-Champaign*
Phillip Brickman, *Institute for Social Research, Ann Arbor, Michigan*
J. Merrill Carlsmith, *Stanford University*
John S. Carroll, *Loyola University of Chicago*
Jerry L. Cohen, *University of Illinois at Urbana-Champaign*
Sheldon Cohen, *University of Oregon*
Rick Crandall, *Texas Christian University*
William D. Crano, *Michigan State University*
Robyn M. Dawes, *University of Oregon*
Kay K. Deaux, *Purdue University*
Marshall Dermer, *University of Wisconsin—Milwaukee*
Edward Diener, *University of Illinois at Urbana-Champaign*
Anthony Doob, *University of Toronto, Toronto, Canada*
Susan V. Eisen, *McLean Hospital, Belmont, Massachusetts*
Baruch Fischhoff, *Decision Research, Eugene, Oregon*
Irene H. Frieze, *University of Pittsburgh*
David Funder, *Stanford University*
David L. Hamilton, *University of California, Santa Barbara*
Ronald D. Hansen, *Oakland University*
Richard J. Harris, *University of New Mexico*
Edward Heisler, *University of Western Ontario, London, Canada*
Donald S. Hinsie, *Veterans Administration Hospital, Brentwood, Los Angeles, California*
Edward E. Jones, *Princeton University*
Martin F. Kaplan, *Northern Illinois University*
Douglas T. Kenrick, *Montana State University*
Norbert L. Kerr, *University of California, San Diego*
Samuel Kessitlis, *University of Illinois at Urbana-Champaign*
Vladimir J. Konečni, *University of California, San Diego*
Robert M. Krauss, *Columbia University*
Ellen J. Langer, *Harvard University*
Mark R. Lepper, *Stanford University*
Melvin Manis, *University of Michigan (Ann Arbor)*
Hazel Markus, *University of Michigan (Ann Arbor)*
Margarett Matlin, *State University of New York College at Geneseo*
Dale T. Miller, *University of British Columbia, Vancouver*
Thomas M. Ostrom, *Ohio State University*
Paul B. Paulus, *University of Texas at Arlington*
Richard Petty, *University of Missouri—Columbia*
Harry T. Reis, *University of Rochester*
Judith Rodin, *Yale University*
Seymour Rosenberg, *Lexington College, Rutgers University*
Ralph L. Rosnow, *Temple University*
Lee Ross, *Stanford University*
Michael Ross, *University of Waterloo, Waterloo, Canada*
Diane Ruble, *Princeton University*
Michael J. Saks, *National Center for State Courts, Williamsburg, Virginia*
Michael F. Scheier, *Carnegie-Mellon University*
Barry R. Schlenker, *University of Florida*
Philip Shaver, *New York University*
Steven J. Sherman, *Indiana University*
Jerome E. Singer, *University of the Health Sciences, Bethesda, Maryland*
Eliot R. Smith, *University of California, Riverside*
Mark Snyder, *University of Minnesota*
Melvin L. Snyder, *University of Texas at Austin*
Lawrence Stricker, *Educational Testing Service, Princeton, New Jersey*
Shelley E. Taylor, *Harvard University*
Abraham Tesser, *University of Georgia*
Bernard Weiner, *University of California, Los Angeles*
Gary L. Wells, *University of Alberta, Edmonton, Canada*
Stephen Worchel, *University of Virginia*
Robert S. Wyer, *University of Illinois at Urbana-Champaign*
Mark Zanna, *University of Waterloo, Waterloo, Canada*
Miron Zuckerman, *University of Rochester*

APA Journal Staff

Anita DeVivo, *Executive Editor*
Ann I. Mahoney, *Manager, Journal Production*
Barbara R. Richman, *Production Supervisor*
Marjorie M. Henry, *Production Editor*
Robert J. Hayward, *Advertising Representative*
Juanita Brodie, *Subscription Manager*

The Journal of Personality and Social Psychology publishes original reports of research, methodology, theory, criticism, and literature review in the fields of personality and social psychology of normal processes and populations. Submitted manuscripts should delineate an important question and provide a clear answer to it. In the case of research reports, this will often require demonstration that the reported results are generalizable beyond a single research setting. Consult the editorial in the January 1979 issue (Vol. 88, pp. 1–7) for a detailed statement of policies and instructions to authors.

Manuscripts: Submit manuscripts in quadruplicate, prepared according to the *Publication Manual of the American Psychological Association* (2nd ed.). Copies should be highly legible and will not be returned to authors unless they are original typewritten copies. Mail manuscripts to the Acting Editor, Clyde Hendrick, Department of Psychology, Box 8433, University of Miami, Coral Gables, FL 33124.

Required supplementary information: Manuscripts cannot be accepted for publication unless the author provides supplementary information in the following categories: (a) a cover letter that indicates the relation (or the lack of relation) of the submitted manuscript to other papers by the same author (a) that are currently submitted, in press, or recently published; (b) one copy of each such cited paper; (c) one copy of a statement of the procedures of the research reported in the submitted manuscript in sufficient detail to serve as a guide for conduct of an exact replication; and (d) one copy of the summary tables (e.g., correlation tables, analysis of variance summaries) for the major analyses reported. If it appears that the supplementary materials in categories (c) or (d) might be bulky, the author should exercise judgment and restraint in selecting what to send. It is not necessary to send supplementary material that repeats verbatim manuscript content.

After a manuscript is accepted, authors will be expected to provide assurance that their data are available in unanalyzed form for reanalyses by other investigators and that their research procedures were consistent with the American Psychological Association's published principles of research ethics, except insofar as may be detailed otherwise in the manuscript.

Copyright and Permission: All rights reserved. Written permission must be obtained from the American Psychological Association for copying or reprinting text of more than 500 words, tables, or figures. Permission is normally granted contingent upon like permission of the author, inclusion of the APA copyright notice on the first page of reproduced material, and payment of a fee of $10 per page, table, or figure. Abstracting is permitted with credit to the source. Permissions and fees are waived for authors who wish to reproduce their own material for personal use; fees only are waived for authors who wish to use their own material commercially. Permission and fees are waived for the photocopying of isolated articles for nonprofit classroom or library reserve use by instructors and educational institutions. Libraries are permitted to photocopy beyond the limits of U.S. copyright law: (1) those post-1977 articles with a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through the Copyright Clearance Center, P.O. Box 765, Schenectady, NY 12301; (2) pre-1978 articles, provided that the per-copy fee stated in the Publishers' Fee List is paid through the Copyright Clearance Center, P.O. Box 765, Schenectady, NY 12301. Address requests for reprint permission to the Permissions Office, APA, 1200 Seventeenth Street, N.W., Washington, D.C. 20036.

Subscriptions: Subscriptions are available on a calendar year basis only (January through December). Nonmember rates for 1979: $60 domestic, $64 foreign, $6 single issue. APA member rate: $20. Write to Subscription Section, APA.

Change of Address: Send change of address notice and a recent mailing label to the attention of the Subscription Section, APA, 30 days prior to the actual change of address. APA will not replace undelivered copies resulting from address changes; journals will be forwarded only if subscribers notify the local post office in writing that they will guarantee second-class forwarding postage.

Reprints: Authors may order reprints of their articles from the printer when they receive proofs.

Back Issues and Back Volumes: For information regarding back issues or back volumes write to Order Department, APA.

Microform Editions: For information regarding microform editions write to any of the following: Johnson Associates, Inc., P.O. Box 1017, Greenwich, Connecticut 06830; University Microfilms, Ann Arbor, Michigan 48106; or Princeton Microfilms, Princeton, New Jersey 08540.

Published monthly in one volume per year by the American Psychological Association, Inc., 1200 Seventeenth Street, N.W., Washington, D.C. 20036. Printed in the U.S.A. Second-class postage paid at Washington, D.C., and at additional mailing offices.

JOURNAL OF
Personality
AND
Social Psychology

Volume 37 January 1979 Number 1

Stressful Life Events, Personality, and Health: An Inquiry Into Hardiness

Suzanne C. Kobasa
University of Chicago

Personality was studied as a conditioner of the effects of stressful life events on illness onset. Two groups of middle and upper level executives had comparably high degrees of stressful life events in the previous 3 years, as measured by the Holmes and Rahe Schedule of Recent Life Events. One group ($n = 86$) suffered high stress without falling ill, whereas the other ($n = 75$) reported becoming sick after their encounter with stressful life events. Illness was measured by the Wyler, Masuda, and Holmes Seriousness of Illness Survey. Discriminant function analysis, run on half of the subjects in each group and cross-validated on the remaining cases, supported the prediction that high stress/low illness executives show, by comparison with high stress/high illness executives, more hardiness, that is, have a stronger commitment to self, an attitude of vigorousness toward the environment, a sense of meaningfulness, and an internal locus of control.

An exceptional number of studies in the last 20 years (cf. Dohrenwend & Dohrenwend, 1974; Gunderson & Rahe, 1974) have suggested that stressful life events precipitate somatic and psychological disease. This article considers the importance of personality as a conditioner of the illness-provoking effects of stress.

During the last decade, investigators have shown that the recent life histories of hospitalized persons contain significantly more frequent and serious stressful events than do histories of matched controls from the general population (e.g., Paykel, 1974) and that Navy personnel who begin a cruise with high stress scores suffer more illness episodes during the months at sea than do sailors who start out with low stress scores (Rahe, 1974). But the possibility of a causal connection between stress and illness is hardly a new idea. Physicians, philosophers, and persons simply

This article is based on the author's doctoral dissertation (Kobasa, 1977), submitted to the Department of Behavioral Sciences at the University of Chicago. The preparation of this manuscript was supported in part by Public Health Service Grant MH-28839-01 from the National Institute of Mental Health. The author wishes to thank Robert R. J. Hilker, James Kennedy, and all of the executives who participated in the study. Special appreciation is extended to Salvatore R. Maddi who supervised the project. Chase P. Kimball, David E. Orlinsky, and Marvin Zonis contributed many useful suggestions as dissertation committee members.

Requests for reprints should be sent to Suzanne C. Kobasa, Department of Behavioral Sciences, University of Chicago, 5848 S. University Avenue, Chicago, Illinois 60637.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3701-0001$00.75

Diary No. 65
Date 2/X/1979
Init No. Librarian
Research
J 853

concerned about staying healthy have long wondered about the etiological significance of life events. The distinctiveness of recent research lies in its attempt to define and measure stress.

In the current studies, a life event is defined as stressful if it causes changes in, and demands readjustment of, an *average person's normal routine*. This definition of stress has relied upon the empirical demonstration (Holmes & Masuda, 1974; Holmes & Rahe, 1967) that there is a general consensus about the degree to which specific life occurrences involve change and require readjustment. This consensus emerged from the ratings of stressfulness of a long list of events made by thousands of subjects varying in age, sex, socioeconomic status, race, cultural background, education, and religion. From this consensus, stressfulness weights for the life events were obtained.

Deriving from this research is the Schedule of Recent Life Events (Holmes & Rahe, 1967), which contains positive (e.g., marriage), negative (e.g., illness of a family member), frequent (e.g., minor traffic violation), and rare (e.g., death of a child) occurrences. The associated Social Readjustment Rating Scale (Rahe, Lundberg, Theorell, & Bennett, 1971) gives the consensual weights for these events (e.g., on a scale from 1 to 100, divorce gets a mean score of 73 and vacation, a score of 13). The stressfulness of a particular person's life is measured in this procedure by finding out what events he or she has encountered and scoring the occurrences according to their consensually quantified weights.

This consensual approach to stress as measurable environmental input has allowed easy collection of large amounts of data and uncomplicated analysis of the statistical relationship between stressful life events and illness onset. The findings of stress research have had dramatic impact in both professional journals and popular magazines. In the journals (e.g., Rabkin & Struening, 1976), one usually finds some caution urged about too quickly concluding a causal relationship between stress and illness on the basis of correlational and methodologically weak studies. This warning is mitigated, however, by a sense of optimism that there is something to this stress and illness connection and a wish for studies that are more sophisticated in design and methods of analysis. But in the popular literature (e.g., Wolfe, 1972), final conclusions have already been drawn. Readers are provided with a self-administered stress test and told the likelihood of illness associated with each total stress score (e.g., a score above 300 means an 80% chance of getting sick). Readers whose stress scores are high are warned that if they wish to remain healthy they should avoid additional stressful encounters, even to the point of shunning such everyday necessities as driving on the Los Angeles Freeway.

This advice seems inadequate, insofar as modern life can be characterized as inherently stressful (Brodsky, 1977; Toffler, 1970). A person might indeed be able to keep from getting married or taking on a new mortgage, but how can one prevent the effects of pollution and overcrowding or of being assigned to a different job when his or her company signs a new contract? Further, increasing levels of stress tend to coincide with increasing opportunities and potential resources (Kobasa, Hilker, & Maddi, Note 1). By avoiding stress, modern persons may be turning away from a chance to better their lives.

Before deciding to avoid stressful life at any cost, one should consider the actual magnitude of the relationship between stress and illness. Although correlations range from .20 to .78, the majority fall below .30, and in Rahe's naval data, the correlations are consistently around .12 (Rabkin & Struening, 1976). Many stress and illness researchers, in their preoccupation with group means, have failed to question the distribution of their data. Variability of both stress and illness scores within groups has been observed to be extreme in several studies (Wershow & Reinhart, 1974). One likely explanation for these data is the presence of subjects with high stress scores who are not getting sick. Oddly, such subjects have been overlooked in the popular and professional literature on stress and illness.

Studying Stressed but Healthy Persons

In contrast, the present study considers how highly stressed subjects who remain healthy differ from those who show illness along with high stress. Studying the individual who undergoes high degrees of stress without falling ill amounts to inquiring about the mediating factors that affect the way one reacts to stress. Holmes and Masuda (1974) and the majority of other stress investigators attempt to draw a direct causal link between the occurrence of stressful life events and the onset of illness by reference to the physiological model of a stress reaction formulated by Hans Selye (1956). Stressful life events are said to evoke "adaptive efforts by the human organism that are faulty in kind or duration, lower 'bodily resistance' and enhance the probability of disease occurrence" (Holmes & Masuda, 1974, p. 68). Holmes and others, however, fail to take into account what Selye goes on to say about individual differences and the stress reaction. In the study described in this article, the more subtle points in Selye's work are referred to, in the attempt to consider factors in the stress reaction that serve to deflect the negative impact of stressful events. Mediators of the stress and illness connection, which probably include physiological predisposition, early childhood experiences, and social resources, as well as the mediator emphasized here, personality, are together responsible for what Selye calls the distinctive way in which each individual "takes to" stressful life occurrences.

The proposition of this study is that persons who experience high degrees of stress without falling ill have a personality structure differentiating them from persons who become sick under stress. This personality difference is best characterized by the term *hardiness*. The conceptual source of the supposition, in contrast to the passive and reactive view of humankind found in most stress and illness work, is a set of approaches to human behavior that Maddi (1976), in his categorization of the major personality theories, calls fulfillment theories. The hardy personality type formulated here builds upon the theorizing of existential psychologists (Kobasa & Maddi, 1977; Maddi, 1975) on the strenuousness of authentic living, White (1959) on competence, Allport (1955) on propriate striving, and Fromm (1947) on the productive orientation. Hardy persons are considered to possess three general characteristics: (a) the belief that they can control or influence the events of their experience, (b) an ability to feel deeply involved in or committed to the activities of their lives, and (c) the anticipation of change as an exciting challenge to further development. Much research has already shown the advantages in behavior of control (e.g., Lefcourt, 1973; Rodin & Langer, 1977; Rotter, Seeman, & Liverant, 1962; Seligman, 1975), commitment (e.g., Antonovsky, 1974; Kobasa & Maddi, 1977; Lazarus, 1966; Lazarus, Averill, & Opton, 1974; Moss, 1973), and challenge (Fiske & Maddi, 1961; Maddi, 1967). In discussing the hypotheses presented below, the implications of theory and research concerning these three general characteristics are extended to considerations of health and illness.

Hypothesis 1. Among persons under stress, those who have a greater sense of control over what occurs in their lives will remain healthier than those who feel powerless in the face of external forces. Following the model proposed by Averill (1973) to explain his laboratory observation that some organisms are not debilitated by stressful stimuli, the highly stressed but healthy person is hypothesized to have (a) decisional control, or the capability of autonomously choosing among various courses of action to handle the stress; (b) cognitive control, or the ability to interpret, appraise, and incorporate various sorts of stressful events into an ongoing life plan and, thereby, deactivate their jarring effects; and (c) coping skill, or a greater repertory of suitable responses to stress developed through a characteristic motivation to achieve across all situations. In contrast, the highly stressed persons who become ill are powerless, nihilistic, and low in motivation for achievement. When stress occurs, they are without recourse for its resolution, give up what little control they do possess, and succumb to the incapacity of illness.

Hypothesis 2. Among persons under stress, those who feel committed to the various areas

of their lives will remain healthier than those who are alienated. Committed persons have a belief system that minimizes the perceived threat of any given stressful life event. The encounter with a stressful environment is mitigated by a sense of purpose that prevents giving up on one's social context and oneself in times of great pressure. Committed persons feel an involvement with others that serves as a generalized resistance resource against the impact of stress (Antonovsky, 1974). Committed persons have both a reason to and an ability to turn to others for assistance in times demanding readjustment.

Although commitment to all areas of life—work, social institutions, interpersonal relationships, family, and self—should be characteristic of highly stressed persons who do not fall ill, one area is singled out as particularly important for health. Staying healthy under stress is critically dependent upon a strong sense of commitment to self. An ability to recognize one's distinctive values, goals, and priorities and an appreciation of one's capacity to have purpose and to make decisions support the internal balance and structure that White and other theorists (cf. Coelho, Hamburg, & Adams, 1974) deem essential for the accurate assessment of the threat posed by a particular life situation and for the competent handling of it.

Hypothesis 3. Among persons under stress, those who view change as a challenge will remain healthier than those who view it as a threat. Persons who feel positively about change are catalysts in their environment and are well practiced at responding to the unexpected. Because they value a life filled with interesting experiences, change seekers have well explored their environment and know where to turn for resources to aid them in coping with stress. They have a predisposition to be cognitively flexible, which allows them to integrate and effectively appraise the threat of new situations. Their basic motivation for endurance allows them to persist even when the new information is exceedingly incongruous and, thereby, maximally provoking of strain and illness (Moss, 1973).

Given the characterization of hardiness contained in the first two hypotheses, it should be clear that the highly stressed but healthy individual is not engaging in irresponsible adventurousness. At the core of the search for novelty and challenge are fundamental life goals that have become, in adulthood, increasingly integrated in a widening diversity of situations (Henry, 1968; Neugarten, 1974).

Although personality is the primary concern of the study reported here, other kinds of variables may well differentiate the highly stressed and healthy from those who have fallen ill under stress. These would include a variety of psychological, social, physiological, and environmental mediators. In order to increase understanding of those who fall ill and stay healthy under high stress, the present study included information concerning various demographic characteristics and perceptions of stressfulness of events.

Method

Overview

The first task was the identification of high stress/low illness and high stress/high illness groups. For this purpose, all members ($N = 837$) of a large subject pool were mailed a stress and illness questionnaire. They were asked to report, by month and year, which of a list of stressful life events and illness episodes they had experienced in the previous 3 years. A total stress score and a total illness score were obtained for each subject by the standard procedure of noting the items checked and multiplying these by their frequencies and consensually defined weights. Subjects were then assigned to one of the two groups or put aside. Assigned to the high stress/low illness and high stress/high illness groups were 126 subjects above the median for total stress and below the median for total illness and 200 subjects above the median for total stress and above the median for total illness, respectively. Discarded subjects included 322 low-stress cases and 22, or all of the female subjects. In further group refinement, 40 high stress/high illness cases whose peak illness score preceded rather than followed their peak stress score and 10 borderline cases from each of the two groups were put aside. Finally, because of a request from the company for which all of the subjects worked that the sample to be administered personality questionnaires be kept at a workable minimum, a set of 100 subjects was randomly selected from the 116 remaining high stress/low illness cases and 100 randomly selected from the 150 remaining high stress/high illness subjects.

Within 3 months of the stress and illness testing, the two groups were mailed another questionnaire. This questionnaire contained several personality tests

along with questions about demographics and perception of life stressfulness.

The data from approximately half of each group were used to test hypotheses concerning group differences in hardiness. Data from the remaining cases in each group were used to cross-validate the results. The division of the groups into test and cross-validation cases was necessitated by the statistical technique relied upon in the study, discriminant function analysis, which has been characterized in previous research by problems of generalizability due to instability of results (cf. Huberty, 1975). Although powerful as a tool for looking at group differences, discriminant function analysis, like regression analysis, has not always provided results that hold up in the making of inferences from sample results to some population, and over repeated samplings. Until a replication of the study is possible, the use of a "holdout sample" must be relied upon for an accurate test of the adequacy of the derived discriminant function that defines group differences.

Subjects

This study required a subject pool both large and stressed enough to obtain sufficiently large groups for study. All of the middle and upper level executives of a large public utility served as the pool from which groups were selected. This utility, in the year prior to the onset of the study, had entertained serious discussion (in executive seminars, company publications, and consultations with the medical department) of the increasing numbers of stressful events faced by executives. These events consisted of changes instituted by the utility itself, like a program of job evaluations that led to some promotions and many more demotions, as well as requirements for readjustment from external sources, such as the federal government's affirmative action demands. These changes, coupled with the expected usual range of personal and family stresses, suggested that stress scores for the subject pool would be generally high.

Demographically, the pool was quite homogeneous. The modal characteristics of the subjects were (a) male gender; (b) 40 to 49 years of age; (c) married, with two children; (d) on the third or middle management level, and having been there for 6 years or more; (e) possessing at least a college degree; (f) wife not working outside the home; (g) usually Protestant, and attending religious services very or fairly often.

Measurement of Stress and Illness

The most frequently used scales in stress and illness research, the Schedule of Recent Life Events and the Social Readjustment Rating Scale (Holmes & Rahe, 1967), were employed in this study. Additions to these scales were made on the basis of pilot testing. The majority of additions were more detailed specifications of the original items, modeled after suggestions from other adapters of the test (Hough, Fairbank, & Garcia, 1976; Paykel, 1974). Each of the most ambiguous events was replaced by two events, one presenting the positive form of the Holmes and Rahe item and the other, the negative version. "Change in financial state," for example, was translated into "improvements in financial state" and "worsening of financial state." These specifications were given the seriousness weights of the items from which they were derived. Other additions to the Holmes and Rahe list were based on a pilot use of the test with 50 randomly selected executives. In response to the question "What other events have you experienced during the past 3 years?", these subjects reported 15 events not found on the original list. Most of these referred to occurrences at work. Seriousness weights were assigned to these additions by the investigator and 20 other judges using the ratio scale judgment procedure of Holmes and Rahe (1967).

The illness items in the stress and illness questionnaire were taken from the Wyler, Masuda, and Holmes (1968) Seriousness of Illness Survey. After consultation with the medical director of the executives' company, 118 of the diseases listed in this survey were chosen as applicable to the pool being tested. Each illness item is characterized by a seriousness weight based on a consensual agreement of numerous and diverse judges (both medical and lay) and is obtained in a manner similar to the derivation of the Social Readjustment Rating Scale. The reliability and validity of this scale as a complete listing of disease syndromes in a form accessible to both laypersons and physicians, and as an accurate set of evaluations of the general seriousness of various distinct illnesses, has been established (cf. Wyler, Masuda, & Holmes, 1970).

Measurement of Personality, Demographic, and Perception Variables

A composite questionnaire, made up of all or parts of four standardized and two newly constructed instruments, was designed to test the three personality hypotheses. The standardized tests were chosen for their theoretical relevance and empirical reliability and validity. All of the instruments are appropriate for use with a sample of executives (i.e., a group of well-educated adult professionals who are relatively free of gross psychopathology).

The *control* dimension was measured through four different instruments. What has been called decisional control or autonomy was measured through the Internal–External Locus of Control Scale (Lefcourt, 1973; Rotter, Seeman, & Liverant, 1962), and the Powerlessness versus Personal Control scale of the Alienation Test (Maddi, Kobasa, & Hoover, Note 2). The latter instrument also provided a way of measuring cognitive control (i.e., the ability to find meaning in stressful life events) in its Nihilism versus Meaningfulness scale. Coping skill (i.e., the availability of responses with which to deal with stressful life events) was measured through the Achievement scale of the Personality Research Form (Jackson, 1974; Wiggins, 1973).

The question of whether these forms of control are accompanied by a need to dominate the behavior of others was addressed through administration of the Leadership Orientation scale of the California Life Goals Evaluation Schedules (Hahn, 1966) and the Dominance scale of the Personality Research Form (Jackson, 1974).

The *commitment* dimension was measured by the Alienation Test (Maddi et al., Note 2). Of relevance here are the alienation versus commitment scores this test yields in five areas of functioning (work, social institutions, interpersonal relationships, family, and self). The degree of consistency across a subject's different forms of involvement was also tested. This was done through the Role Consistency Test, adapted from the Gergen and Morse (1967) Self-Consistency Test to measure the compatibility among the subject's reported five most important life roles.

Measurement of the orientation to *challenge* required the administration of several tests. The degree to which the executive sought out rather than avoided stimulation from his environment was measured through the Preference for Interesting Experiences scale of the Hahn (1966) test and the Vegetativeness versus Vigorousness scale of the Alienation Test. The executive's ability to seek challenge even in the face of potential psychological, social, and biological threat was indexed by Hahn's scale of Security Orientation. Cognitive flexibility versus cognitive rigidity was measured through the Need for Cognitive Structure scale of the Personality Research Form and the ability to persist even in the most challenging environment, through the Need for Endurance scale of the same instrument. Finally, the degree to which an orientation toward challenge was characterized by a general sense of responsibility toward life's demands was measured through the Adventurousness versus Responsibility scale of the Alienation Test.

Subjects were asked about three demographic characteristics: age, job level (third, fourth, fifth, and officer), and number of years spent at current level. Executives were also asked to rate on a scale of 1 (not at all stressful) to 7 (extremely stressful) how stressful they usually think each of the following areas of life is: work, financial concerns, social/community involvements, interpersonal relationships, family, and personal or inner-life concerns.

Results

Stress and Illness Scores

Of the 837 executives who were contacted by mail, 670 (80%) returned completed stress and illness questionnaires. The mean stress score was 399, with a standard deviation of 162, a range from 0 to 2,239, and a median of 306. The mean illness score was 913, with a standard deviation of 1,115, a range from 0 to 6,900, and a median of 550. Comparison of these executives' scores with norms available in the literature (Holmes & Masuda 1974; Wyler et al., 1968, 1970) indicates that on the average, they had experienced a moderate amount of illness (comparable to the threat of life and discomfort associated with having a peptic ulcer or high blood pressure) in the previous 3 years. During this period, they had, on the average, also encountered sufficient change and demands for readjustment in their lives to constitute major life crisis (what Holmes has associated with an 80% likelihood of getting seriously ill in the near future).

A Pearson product-moment correlation of .24 ($p < .025$) was obtained between the total stress and total illness scores. Canonical correlations were also run to determine whether a stronger relationship might be obtained. Stress scores—both the standard weighted scores and simple frequencies—for various areas of life (work, home, community, etc.) were used as the predictor, and illness scores—both weighted and frequencies—for various body systems (cardiovascular, respiratory, etc.) were used as the criterion set. Neither analysis of unweighted stress events and illnesses nor correlations of selected types of stress with selected types of illness, however, significantly strengthened the statistical relationship between stress and illness. The weak, but significant, correlation of .24 is consistent with most of the available research reports (Rabkin & Struening, 1976).

Testing of Group Differences

Completed personality questionnaires were returned by 86% of the high stress/low illness and 75% of the high stress/high illness subjects. Forty "test" subjects were randomly selected from each group for analysis of differences across personality, demographic, and perception variables, and 81 subjects were put aside for cross-validation purposes.

Mean differences between the test subjects of the two groups were evaluated by *t* test. None of the demographic variables and only one of the perception variables yielded significant differences, whereas several of the personality variables did so. Further evaluation of group differences was achieved by a

Table 1
Differences Between High Stress/Low Illness[a] and High Stress/High Illness[a] Executives

Variable	High stress/ low illness		High stress/ high illness		*t* value	Standardized discriminant function coefficient
	M	*SD*	*M*	*SD*		
Control						
Nihilism	196.05	133.61	281.02	169.86	2.49**	.73
External locus of control	5.92	4.10	7.90	4.61	2.03*	.22
Powerlessness	301.15	188.93	388.47	188.44	2.11*	—
Achievement	16.50	2.10	15.12	3.20	−1.20	—
Dominance	14.60	3.26	13.85	4.46	.86	—
Leadership	33.47	7.34	34.63	6.80	.73	.43
Commitment						
Alienation from self	102.35	117.24	219.15	185.77	3.36**	1.04
Alienation from work	181.67	122.04	223.73	175.09	1.22	.43
Alienation from interpersonal	256.02	162.76	316.10	165.24	1.64	—
Alienation from family	158.47	139.02	198.72	144.33	1.27	—
Alienation from social	202.15	100.21	226.95	133.93	.94	—
Role consistency	29.22	6.42	29.50	6.44	.19	.30
Challenge						
Vegetativeness	155.50	140.24	216.27	160.94	1.98*	.99
Security	21.11	6.33	22.19	8.60	.34	.35
Cognitive structure	13.35	2.81	14.10	2.85	1.10	.21
Adventurousness	269.00	164.58	337.54	174.95	1.78*	—
Endurance	15.97	2.35	14.37	3.19	−.96	—
Interesting experiences	34.97	6.83	32.52	7.02	−.92	
Perception of personal stress	3.00	1.21	3.83	1.73	2.46**	.43

Note. For all variables, the higher the number, the greater the degree of the variable observed. Superior hardiness is indicated by higher scores on achievement, role consistency, endurance, and interesting experiences, and lower scores on nihilism, external locus, powerlessness, dominance, leadership, alienation (from self, work, social institutions, interpersonal relationships, and family), vegetativeness, security, cognitive structure, and adventurousness. A subject's scores on all areas of alienation, measured by the Alienation Test, have a possible range of 0 to 1,200. Vegetativeness, nihilism, powerlessness, and adventurousness scores, also from the Alienation Test, may range from 0 to 1,500. External locus has a low of 0 and an upper limit of 23. The scales taken from the Jackson test—achievement, dominance, cognitive structure, and endurance—have a minimum value of 0 and a maximum of 20. The California Life Goals scale—leadership, security, and interesting experiences—may range from 0 to 60. Role consistency has a low of 0 and a high of 40; perception of personal stress can range from 0 to 7.

[a] $n = 40$.

* $p < .05$.

** $p < .01$.

discriminant function analysis done on all the personality variables plus the one perception variable that yielded a significant *t* score. After all data are transformed into standard scores, discriminant function analysis computes a discriminant equation, or a linear combination of weighted variables that produces the greatest statistically derivable distance between the two groups. The larger the weighting or discriminant coefficient of a variable, the more powerful it is as a group discriminator.

Table 1 presents all variables submitted to discriminant function analysis and their summary statistics, including significance of difference between groups established by *t* test. The results demonstrate that high stress/low illness executives can be distinguished from high stress/high illness subjects. The 11 variables for which a standardized discriminant function coefficient is provided in the table combine to form a significant function, with a Wilks's Lambda of .64, significant at the .001 level, and a canonical correlation of .60. The Wilks's Lambda (see Huberty, 1975) is a measure of the original variable's discrimi-

nant power, before it is removed by the discriminant function. The canonical correlation is a measure of the association between the single discriminant function and the set of dummy variables that defines the two group memberships. Given the composition of this function, the general position concerning hardiness is confirmed. High stress/low illness executives are, at least in some ways, more in control, more committed, and more oriented to challenge than are high stress/high illness executives. Looking at the variables that make the most significant contribution to the discriminant equation and that are also responsible for significant mean differences, the high stress/low illness executives are distinguished by their sense of commitment to (or lack of alienation from) self, their sense of vigorousness (as opposed to vegetativeness) about life, their sense of meaningfulness (as opposed to nihilism), an their internal (as opposed to external) locu of control. One perception-of-stress variabl joins the important personality discrimina tors—high stress/low illness executives fin the personal sphere of their lives significantl less stressful then do the high stress/hig illness subjects.

Several personality variables appear to con tribute to the discriminant equation eve though they do not yield significant mean dif ferences by t test. Although it is risky t interpret the discriminant coefficients of thes variables, it is worth noting that the direc tion of mean difference is in general consisten with the hypotheses concrening hardiness a an insulation against illness.

Along with their analytic function, the dis criminant coefficients also serve a classifica tory and validation purpose. Multiplying a

Table 2
Differences Between All High Stress/Low Illness[a] and All High Stress/High Illness[b] Executives

Variable	High stress/ low illness		High stress/ high illness		t value
	M	*SD*	*M*	*SD*	
Control					
Nihilism	217.01	141.03	262.23	161.77	1.82*
External locus of control	5.38	3.90	7.51	4.60	2.05*
Powerlessness	314.95	190.65	364.20	189.50	1.82*
Achievement	16.40	2.35	15.95	3.03	−1.02
Dominance	14.35	3.38	14.27	4.19	−.13
Leadership	33.89	6.56	34.53	6.29	.54
Commitment					
Alienation from self	113.87	122.66	180.56	150.65	2.97**
Alienation from work	180.61	129.22	201.68	159.69	.89
Alienation from interpersonal	276.53	159.95	309.64	151.54	1.30
Alienation from family	168.93	145.83	200.99	144.62	1.35
Alienation from social	205.39	98.37	219.28	123.45	.76
Role consistency	28.80	6.78	29.59	6.84	.71
Challenge					
Vegetativeness	160.04	133.34	215.26	153.54	1.82*
Security	21.52	6.83	22.63	7.43	.95
Cognitive structure	13.16	2.77	13.75	2.94	1.26
Adventurousness	276.88	165.58	330.84	158.91	1.78*
Endurance	15.01	2.95	14.53	3.15	−.96
Interesting experiences	34.29	6.31	32.80	7.00	−1.57
Perception of personal stress	2.72	1.31	3.61	1.79	3.48**

[a] $n = 86$.
[b] $n = 75$.
* $p < .05$.
** $p < .01$.

subject's scores on the discriminating variables by the associated coefficients allows one to predict the likelihood of the subject's membership in each of the groups. As a test of the classificatory power of the derived discriminant equation, its coefficients are (a) reapplied to the test subjects used to derive the equation and (b) applied to the scores of the "holdout" cases. The discriminant function presented in Table 1 shows predictive capability both internally and externally. Applying the unstandardized versions of the discriminant function coefficients to the raw scores used to derive the function, 78% of the "test" cases are correctly classified (80% of the high stress/low illness executives and 75% of the high stress/high illness subjects). This significantly correct ($p < .025$) classification is matched in the external cross-validation. Using the unstandardized coefficients on the raw data from the "holdout" subjects, 35 hits (77%) and 21 hits (60%) in the high stress/low illness and high stress/high illness groups, respectively, are realized ($p < .05$).

These cross-validation results offer support for the stability and generalizibility of the results obtained through discriminant function analysis. An examination of the statistics for the entire sample (test cases plus holdouts), which are notably similar to those of the test cases in Table 1, illustrates the strength of the discriminant analysis. Table 2 presents the mean values, standard deviations, and t values for the full high stress/low illness and high stress/high illness groups.

Discussion

This study of persons who do not fall ill despite considerable stress suggests that personality may have something to do with staying healthy. Using the five most significant discriminators of the high stress/low illness executives (i.e., the variables that contribute to the discriminant equation and produce significant ts) one can speculate on what happens when the hardy individuals meet a stressful life event—how they evaluate the threat posed by the event and cope with it.

A male executive having to deal with a job transfer will serve as an example. Whether hardy or not, the executive will anticipate and experience the changes that the transfer will bring about—learning to cope with new subordinates and supervisors, finding a new home, helping children and wife with a new school and neighborhood, learning new job skills, and so on. The hardy executive will approach the necessary readjustments in his life with (a) a clear sense of his values, goals, and capabilities, and a belief in their importance (*commitment* to rather than *alienation* from *self*) and (b) a strong tendency toward active involvement with his environment (*vigorousness* rather than *vegetativeness*). Hence, the hardy executive does more than passively acquiesce to the job transfer. Rather, he throws himself actively into the new situation, utilizing his inner resources to make it his own. Another important characteristic of the hardy executive is an unshakable sense of meaningfulness and ability to evaluate the impact of a transfer in terms of a general life plan with its established priorities (*meaningfulness* rather than *nihilism*). For him, the job transfer means a change that can be transformed into a potential step in the right direction in his overarching career plan and also provide his family with a developmentally stimulating change. An *internal* (rather than *external*) locus of control allows the hardy executive to greet the transfer with the recognition that although it may have been initiated in an office above him, the actual course it takes is dependent upon how he handles it. For all these reasons, he is not just a victim of a threatening change but an active determinant of the consequences it brings about. In contrast, the executive low in hardiness will react to the transfer with less sense of personal resource, more acquiescence, more encroachments of meaninglessness, and a conviction that the change has been externally determined with no possibility of control on his part. In this context, it is understandable that the hardy executive will also tend to perceive the transfer as less personally stressful than his less hardy counterpart.

The mechanism whereby stressful life events produce illness is presumably physiological. Whatever this physiological response is, the personality characteristics of hardiness may cut into it, decreasing the likelihood of

breakdown into illness. Needless to say, description of the actual nature of physiological mechanisms and their links to personality and stressful occurrences will have to wait for more sophisticated stress research. Until then, however, two alternative explanations of the present results should be considered.

It could be argued that there is a spurious factor at work in the subjects' completion of questionnaires, one that determines what they say about personality, stress levels, and illness experience. Mechanic (1976) has discussed the possible distortion of stress data by a variable that he calls illness behavior. From Mechanic's perspective, one could suggest that the high stress/high illness executives have not really undergone any physiological breakdown but that they simply want to act and be treated as if they are sick and thereby withdraw from a life situation experienced as too stressful. The best evaluation of this position would come from a study that employs physiological rather than self-report measures of illness and monitors physiological states before and after the occurrence of a stressful life event. Until such studies are available, reliance must be placed on data already collected. Mechanic's explanation seems much more likely to apply to the self-report of mild, vague symptoms than to report of more serious illness. For example, headaches and indigestion might well be reported by subjects engaging in illness behavior. But it seems unlikely that definite illnesses requiring medical diagnosis, such as heart attack, cancer, detached retina, and even hypertension would be erroneously reported. Illnesses of the latter sort are responsible for high illness scores on the self-report instrument used in the present study. It is therefore not likely that the results obtained can be explained completely by the concept of illness behavior.

Another alternative explanation would suggest that the observed personality differences between the high stress/low illness and the high stress/high illness executives result from the latter group having experienced illness. Presumably, the high stress/high illness executives were not more alienated, external in beliefs concerning locus of control, and uninterested in change before the illnesses occurred. Their responses to the personality questionnaire may reflect the psychological debilitating effects of illness. A prospective longitudinal study in which stress and personality scores at Time 1 are used to predict illness scores at Time 2 is obviously necessary. But the present study provides some basis for regarding the alternative explanation as less likely than that offered here. Although not discussed in this report, subjects found to be low in stress and high in illness showed personality scores midway between the high stress/low illness and high stress/high illness groups (Kobasa, 1977). It is of specific relevance to the alternative explanation that the low stress/high illness subjects were lower on nihilism, alienation from self, vegetativeness, and external locus of control than were high stress/high illness subjects. This finding indicates that personality questionnaire responses indicative of lack of hardiness are not merely a response to illness.

Continued work in the area of stress is essential. There is still much to be learned about the role of personality and other mediators in the connection between stressful life events and illness or health. Should studies incorporating physiological measures and employing prospective designs confirm the present findings, then instead of merely warning persons to avoid stressful lives, social scientists will be able to illuminate ways of developing the personality characteristics that can aid in a productive and healthy life lived in the full complexity of modern, urbanized, industrialized societies.

Reference Notes

1. Kobasa, S. C., Hilker, R. R. J., & Maddi, S. R. *Remaining healthy in the encounter with stress*. Paper presented at the meeting of the American Medical Association Congress on Occupational Health, St. Louis, Missouri, September 1977.

2. Maddi, S. R., Kobasa, S. C., & Hoover, M. *The Alienation Test: A structured measure of a multidimensional subjective state*. Manuscript submitted for publication, 1978.

References

Allport, G. W. *Becoming: Basic considerations for a psychology of personality*. New Haven, Conn.: Yale University Press, 1955.

Antonovsky, A. Conceptual and methodological problems in the study of resistance resources and stressful life events. In B. S. Dohrenwend & B. P. Dohr

renwend (Eds.), *Stressful life events: Their nature and effects*. New York: Wiley, 1974.

Averill, J. R. Personal control over aversive stimuli and its relationship to stress. *Psychological Bulletin*, 1973, *80*, 286–303.

Brodsky, C. M. Long-term work stress in teachers and prison guards. *Journal of Occupational Medicine*, 1977, *19*, 133–138.

Coelho, G., Hamburg, D. A., & Adams, J. E. (Eds.). *Coping and adaptation*. New York: Basic Books, 1974.

Dohrenwend, B. S., & Dohrenwend, B. P. (Eds.). *Stressful life events: Their nature and effects*. New York: Wiley, 1974.

Fiske, D. W., & Maddi, S. R. (Eds.). *Functions of varied experience*. Homewood, Ill.: Dorsey Press, 1961.

Fromm, E. *Man for himself*. New York: Holt, Rinehart & Winston, 1947.

Gergen, K. J., & Morse, S. J. Self-consistency: Measurement & validation. *Proceedings of the 75th Annual Convention of the American Psychological Association*, 1967, *2*, 207–208. (Summary)

Gunderson, E., & Rahe, R. (Eds.). *Life stress and illness*. Springfield, Ill.: Charles C Thomas, 1974.

Hahn, M. E. *California Life Goals Evaluation Schedules*. Palo Alto, Calif.: Western Psychological Services, 1966.

Henry, W. E. Personality change in middle and old age. In E. D. Norbeck, D. Price-Williams, & W. M. McCord (Eds.), *The study of personality: An interdisciplinary appraisal*. Chicago: Holt, Rinehart & Winston, 1968.

Holmes, T. H., & Masuda, M. Life change and illness susceptibility. In B. S. Dohrenwend & B. P. Dohrenwend (Eds.), *Stressful life events: Their nature and effects*. New York: Wiley, 1974.

Holmes, T. H., & Rahe, R. H. The Social Readjustment Rating Scale. *Journal of Psychosomatic Research*, 1967, *11*, 213–218.

Hough, R. L., Fairbank, D. T., & Garcia, A. M. Problems in the ratio measurement of life stress. *Journal of Health and Social Behavior*, 1976, *17*, 70–82.

Huberty, C. J. Discriminant analysis. *Review of Educational Research*, 1975, *45*, 543–598.

Jackson, D. N. *Personality research form manual*. Goshen, N.Y.: Research Psychologists Press, 1974.

Kobasa, S. C. *Stress personality and health: A study of an overlooked possibility*. Unpublished doctoral dissertation, University of Chicago, 1977.

Kobasa, S. C., & Maddi, S. R. Existential personality theory. In R. Corsini (Ed.), *Current personality theories*. Itasca, Ill.: Peacock, 1977.

Lazarus, R. S. *Psychological stress and the coping process*. New York: McGraw-Hill, 1966.

Lazarus, R. S., Averill, J. R., & Opton, E. M., Jr. The psychology of coping: Issues of research and assessment. In G. V. Coelho, D. A. Hamburg, & J. E. Adams (Eds.), *Coping and adaptation*. New York: Basic Books, 1974.

Lefcourt, H. M. The function of the illusions of control and freedom. *American Psychologist*, 1973, *28*, 417–425.

Maddi, S. R. The existential neurosis. *Journal of Abnormal Psychology*, 1967, *72*, 311–325.

Maddi, S. R. The strenuousness of the creative life. In I. A. Taylor & J. W. Getzels (Eds.), *Perspectives in creativity*. Chicago: Aldine, 1975.

Maddi, S. R. *Personality theories: A comparative analysis* (3rd ed.). Homewood, Ill.: Dorsey Press, 1976.

Mechanic, D. Stress illness and illness behavior. *Journal of Human Stress*, 1976, *2*, 2–6.

Moss, G. E. *Illness, immunity, and social interaction*. New York: Wiley, 1973.

Neugarten, B. L. The middle years. In S. Arieti (Ed.), *American handbook of psychiatry*. New York: Basic Books, 1974.

Paykel, E. S. Recent life events and clinical depression. In E. K. Gunderson & R. H. Rahe (Eds.), *Life stress and illness*. Springfield, Ill.: Charles C Thomas, 1974.

Rabkin, J. G., & Struening, E. L. Life events, stress, and illness. *Science*, 1976, *194*, 1013–1020.

Rahe, R. H. The pathway between subjects' recent life change and their near-future illness reports: Representative results and methodological issues. In B. S. Dohrenwend & B. P. Dohrenwend (Eds.), *Stressful life events: Their nature and effects*. New York: Wiley, 1974.

Rahe, R. H., Lundberg, U., Theorell, T., & Bennett, L. K. The Social Readjustment Rating Scale: A comparative study of Swedes and Americans. *Journal of Psychosomatic Research*, 1971, *51*, 241–249.

Rodin, J., & Langer, E. J. Long-term effects of a control-relevant intervention with the institutionalized aged. *Journal of Personality and Social Psychology*, 1977, *35*, 897–902.

Rotter, J. B., Seeman, M., & Liverant, S. Internal vs. external locus of control of reinforcement: A major variable in behavior theory. In N. F. Washburne (Ed.), *Decisions, values, and groups*. London: Pergamon, 1962.

Seligman, M. E. P. *Helplessness*. San Francisco: Freeman, 1975.

Selye, H. *The stress of life*. New York: McGraw-Hill, 1956.

Toffler, A. *Future shock*. New York: Random House, 1970.

Wershow, H., & Reinhart, G. Life change and hospitalization—A heretical view. *Journal of Psychosomatic Research*, 1974, *18*, 393–401.

White, R. W. Motivation reconsidered: The concept of competence. *Psychological Review*, 1959, *66*, 297–333.

Wiggins, J. S. *Personality and prediction: Principles of personality assessment*. Reading, Mass.: Addison-Wesley, 1973.

Wolfe, S. W. Avoid sickness—How life changes affect your health. *Family Circle*, May 1972, pp. 30; 166–170.

Wyler, A. R., Masuda, M., & Holmes, T. H. Seriousness of Illness Rating Scale. *Journal of Psychosomatic Research*, 1968, *11*, 363–375.

Wyler, A. R., Masuda, M., & Holmes, T. H. Seriousness of Illness Rating Scale: Reproducibility. *Journal of Psychosomatic Research*, 1970, *14*, 59–64.

Received January 9, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 1, 12-24

Interpersonal Attraction in Exchange and Communal Relationships

Margaret S. Clark
Carnegie-Mellon University

Judson Mills
University of Maryland

Communal relationships, in which the giving of a benefit in response to a need for the benefit is appropriate, are distinguished from exchange relationships, in which the giving of a benefit in response to the receipt of a benefit is appropriate. Based on this distinction, it was hypothesized that the receipt of a benefit after the person has been benefited leads to greater attraction when an exchange relationship is preferred and decreases attraction when a communal relationship is desired. These hypotheses were supported in Experiment 1, which used male subjects. Experiment 2, which used a different manipulation of exchange versus communal relationships and female subjects, supported the hypotheses that (a) a request for a benefit after the person is aided by the other leads to greater attraction when an exchange relationship is expected and decreases attraction when a communal relationship is expected, and (b) a request for a benefit in the absence of prior aid from the other decreases attraction when an exchange relationship is expected.

This research is concerned with how the effects of receiving a benefit and a request for a benefit differ depending on the type of relationship one has with the other person. Two kinds of relationships in which persons give benefits to one another are distinguished, exchange relationships and communal relationships. The stimulus for this distinction was Erving Goffman's (1961, pp. 275–276) differentiation between social and economic exchange.

In the present theorizing, the term *exchange relationship* is used in place of Goffman's term *economic exchange* because many of the benefits that people give and receive do not involve money or things for which a monetary value can be calculated. A benefit can be anything a person can choose to give to another person that is of use to the person receiving it.[1]

In an exchange relationship, members assume that benefits are given with the expectation of receiving a benefit in return. The receipt of a benefit incurs a debt or obligation to return a comparable benefit. Each person is concerned with how much he or she receives in exchange for benefiting the other and how much is owed the other for the benefits received.

Since all relationships in which persons give and receive benefits are social, another term is needed to describe relationships in which each person has a concern for the welfare of the other. The term *communal* seems to be the most appropriate. The typical relationship between family members exemplifies this type

This research was supported by a grant from the National Science Foundation.

Requests for reprints should be sent to Margaret S. Clark, Department of Psychology, Carnegie-Mellon University, Schenley Park, Pittsburgh, Pennsylvania 15213.

[1] Benefits are not the same as rewards, when the term *rewards* refers to "the pleasures, satisfactions, and gratifications the person enjoys" (Thibaut & Kelley, 1959, p. 12). The receipt of a benefit usually constitutes a reward, however rewards occur for reasons other than the receipt of a benefit. For example, the rewards that a parent receives from a newborn infant would not fall within the definition of a benefit, since the infant does not choose to give them to the parent. The present theorizing is not concerned with "dependent" relationships in which one person receives benefits from another but does not give benefits to the other.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3701-0012$00.75

of relationship. Although it might appear to an observer that there is an exchange of benefits in communal relationships, the rules concerning giving and receiving benefits are different than in exchange relationships.

Members of a communal relationship assume that each is concerned about the welfare of the other. They have a positive attitude toward benefiting the other when a need for the benefit exists. They follow what Pruitt (1972) has labeled "the norm of mutual responsiveness." This rule may create what appears to an observer to be an exchange of benefits, but it is distinct from the rule that governs exchange relationships whereby the receipt of a benefit must be reciprocated by the giving of a comparable benefit. The rules concerning the giving and receiving of benefits are what distinguish communal and exchange relationships, rather than the specific benefits that are given and received.

From the perspective of the participants in a communal relationship, the benefits given and received are not part of an exchange. The attribution of motivation for the giving of benefits is different from that in an exchange relationship. In a communal relationship, the receipt of a benefit does not create a specific debt or obligation to return a comparable benefit, nor does it alter the general obligation that the members have to aid the other when the other has a need. In a communal relationship, the idea that a benefit is given in response to a benefit that was received is compromising, because it calls into question the assumption that each member responds to the needs of the other.

Experiment 1

The first study reported here was based on the assumption, similar to that made by Kiesler (1966) in her study of the effect of perceived role requirements on reactions to favor-doing, that the giving of a benefit will decrease attraction if it is inappropriate for the type of relationship one has with the other. A benefit given in response to a benefit received in the past or expected in the future is appropriate in an exchange relationship but is inappropriate in a communal relationship. A benefit given specifically because it fulfills a need is appropriate in a communal relationship but not in an exchange relationship.

If two people have an exchange relationship and one person benefits the other, it is appropriate for the other to give the person a comparable benefit. The receipt of a benefit under these circumstances should lead to greater attraction. On the other hand, if two people have a communal relationship and one person benefits the other, it is inappropriate for the other to give the person a comparable benefit, since it leaves the impression that the benefit was given in response to the benefit received previously. The other is treating the relationship in terms of exchange, which is inappropriate in a communal relationship.

When a communal relationship does not yet exist but is desired, the receipt of a benefit should have the same effect as when a communal relationship is assumed to exist. A benefit from the other after the other has been benefited should reduce attraction if there is a desire for a communal relationship with the other. If an exchange relationship is preferred, the receipt of a benefit after the other is benefited should result in greater attraction. Experiment 1 was conducted to test these hypotheses.

The predictions concerning communal relationships might seem contrary to what would be expected from equity theory (Adams, 1963). On the basis of equity theory, one might expect that a benefit from another following aid to that other would increase liking in *any* relationship, because it would reduce inequity. However, the predictions are not inconsistent with a recent discussion of equity theory (Walster, Walster, & Berscheid, 1978). According to Walster, Walster, and Berscheid:

> Another characteristic of intimate relationships, which may add complexity, is that intimates, through identification with and empathy for their partners, come to define themselves as a *unit*; as *one* couple. They see themselves not merely as individuals interacting with others, but also as part of a partnership, interacting with other individuals, partnerships, and groups. This characteristic may have a dramatic impact on intimates' perceptions of what is and is not equitable. (pp. 152–153)

In Experiment 1 the desire for a communal relationship was manipulated by using un-

married males as the subjects and having the part of the other played by an attractive woman, who was described as either married or unmarried. It was assumed that people desire communal relationships with attractive others, but only with those available for such relationships. It was further assumed that the unmarried woman would be considered available for a communal relationship, whereas the married woman would not. Thus, it was assumed that the male subjects would desire a communal relationship with the attractive, unmarried woman but would prefer an exchange relationship with the attractive, married woman.

Method

Overview. Under the guise of a study of task performance, unmarried male college students worked on a task while a television monitor showed an attractive woman working on a similar task in another room. When the subject completed the task, he was awarded 1 point toward extra credit for finishing on time and given the opportunity to send some of his excess materials to the other, who supposedly had a more difficult task. Following receipt of the aid, the other completed her task and was awarded 4 points. Some of the subjects received a note from the other that thanked them (no-benefit conditions), whereas some received a note from the other that thanked them and gave them one of her points (benefit conditions). After receiving the note, the subjects were given information indicating that the other was either married (exchange conditions) or unmarried (communal conditions). Shortly thereafter, liking for the other and expectations concerning a future discussion with the other were assessed.

Subjects. The subjects were 96 unmarried, male students in introductory psychology courses who participated in order to earn extra course credit. They were randomly assigned to one of the four experimental conditions: exchange–benefit, exchange–no benefit, communal–benefit, communal–no benefit.

Procedure. Upon arriving for the experiment, the subject was greeted by the experimenter and told that another subject had already arrived and was in an adjacent room. The subject was seated so that he could see a video monitor that showed an attractive female. The experimenter mentioned that the person appearing on the monitor was the other subject, Tricia. The experimenter told the subject that before starting she would explain some things about the studies in which the subject would be participating, which she had already mentioned to Tricia.

The experimenter said that the first study was actually one of two she would ask them to participate in that day. Each of the two studies took less than half an hour. Although the studies were unrelated, she was asking people to participate in both of them during one session. The first study involved having both subjects work on a vocabulary task. She suspected that people's approaches to solving this task varied when certain conditions were changed. In the condition to which he and Tricia had been randomly assigned, they would be able to see each other over closed-circuit television but not be able to talk to each other directly. To enhance credibility, there was a portable television camera in the room pointing at the subject. Through the use of videotape, what appeared on the monitor was the same for every subject. When the subject asked why he was watching the other person on the monitor, which typically happened, he was told that in the past it had been found that when people worked separately on these tasks in the same room their performance was often affected by the presence of the other person. This might have happened because people could talk to one another or because they could see one another, and the experiment was designed to separate those variables.

The experimenter explained that the second study would be quite different from the first, since it would involve much more contact with the other person. She would bring both participants into one room and ask them to talk over things that they had in common. She was interested in the way in which holding common interests helped people to get to know one another. The experimenter mentioned that in the past, people who had participated had sometimes gotten to know each other quite well.

Vocabulary task. The experimenter next pointed to a batch of letters printed on small cards in front of the subject and said that his task in the first study was to form 10 different four-letter words from the letters. She went on to say that there were two versions of this task and one was more difficult than the other. She had flipped a coin to determine which task each of them would be working on, and the subject would have the easier one, while Tricia would have the more difficult one. The subject's task was easier because he had 55 letters ranging from A to Z, whereas Tricia had only 45 letters ranging from A to L. The subject was to time himself on the task with a stopwatch. Although he and Tricia were not permitted to speak to each other directly, one type of communication was permitted: They would be allowed to send and request letters. Simple forms were available for this purpose. The experimenter would come into the room from time to time to see if the subject had any messages he wanted to send and would deliver them for him.

The experimenter explained that time was a factor in the first study and that an individual's motivation to do well on a task would obviously affect the time taken to complete the task. Therefore, she was taking precautions to insure that motivation to do well was kept at a high level. She would award points toward a possible extra credit for finishing the task in a minimum amount of time. The subjects were not told exactly what that amount of time would be or how the conversion to credit would be made, but just to finish as quickly as they could. The experimenter mentioned that the award-

ing of points to maintain motivation would obviously not be necessary in the second study.

The subject was told to start his stopwatch, begin working, and stop the watch when he had finished. The experimenter picked up another stopwatch and left, saying she would give Tricia the watch and start her on her task. The 55 letters that the subject had to work with allowed him to complete the task within 10 minutes. When approximately 10 minutes had elapsed, the experimenter returned and asked if the subject had any letters he wanted to send to Tricia. All subjects gave the experimenter some letters for Tricia. At this point, the experimenter looked at the subject's stopwatch and told him that he had finished the task in time to get 1 point toward extra credit. As she did so, she also filled out a form indicating the time the subject had taken to complete the task and that he had earned 1 point. She explained that if Tricia finished her task in time she would earn 4 points, since her task was more difficult. The experimenter went on to say that since she allowed participants in the study to send and request letters, she also allowed them to share points they earned. Thus, Tricia could send the subject some of her points if she wanted to do so. The experimenter left the room saying she would give Tricia the letters. Tricia continued to work on her task for about 5 minutes and then finished. The experimenter handed Tricia a form similar to that given the subject earlier. Tricia smiled, wrote a note on a slip of paper, folded it, and gave it to the experimenter.

Benefit manipulation. Within a few moments, the experimenter returned to the subject's room and turned off the monitor. She mentioned that Tricia had completed her task within the necessary time and had received 4 points. She handed the subject a folded note that she said Tricia had asked be given to him. In the no benefit condition, the note said, "Thanks for sending the letters." In the benefit conditions, the note said, "Thanks for sending the letters. The experimenter said it would be OK to give you one of my points. She said she would add it onto the points you've already earned before the end of today's session." Which message the note contained was unknown to the experimenter at the time she handed it to the subject. This was accomplished by having the experimenter pick the note out of a container of folded notes of both types.

Relationship manipulation. The experimenter told the subject that there was one more thing to be done before getting on to the next study. She said she was going to give Tricia some questionnaires to fill out and would then get some more forms for the subject. In the exchange conditions, the experimenter said:

> Tricia is anxious to get on to the next part of the study, since she thinks it will be interesting. Her husband is coming to pick her up in about half an hour and she wants to finish before then.

In the communal conditions, she said:

> Tricia is anxious to get on to the next part of the study, since she thinks it will be interesting. She's new at the university and doesn't know many people. She has to be at the administration building in about half an hour and she wants to finish before then.

Dependent measures. The experimenter then left the room for approximately 5 minutes. When she reappeared she brought two forms, mentioning that these were the forms she had told the subject about. She reminded the subject that the second study involved having the participants talk over things they had in common with each other. She said that before starting it was necessary to get some idea of what their expectations were in order to control for them, since they would vary from person to person. The subject was asked to fill out a form indicating what he expected the interaction would be like and, in addition, another form indicating what his first impressions of the other person were. The experimenter said that these forms would be kept completely confidential and left the room while they were filled out.

The first-impressions form, which was given to the subject on top of the form concerning expectations about the discussion, asked him to rate how well 11 traits applied to the other, on a scale from 0 (extremely inappropriate) to 20 (extremely appropriate). The traits were *considerate, friendly, insincere, intelligent, irritating, kind, open-minded, sympathetic, understanding, unpleasant,* and *warm.* The subject was also asked to indicate his degree of liking for the other, on a scale from 0 (dislike very much) to 20 (like very much). The other form asked the subject to indicate how friendly, spontaneous, relaxed, enjoyable, and smooth he expected the discussion to be, on scales from 0 to 20.

Suspicion check. After the subject had completed both forms, the experimenter casually mentioned that there was something more to the study and asked whether the subject had any idea what it might be. The responses of eight persons indicated suspicion of the instructions, and they were not included as subjects. Four persons thought the experiment was designed to test reactions to the note, one thought the points might have something to do with the ratings, two thought Tricia was a part of the experiment, and one questioned whether Tricia was actually married. Four of these persons were run under the exchange–benefit condition, two under the exchange–no-benefit condition, one under the communal–benefit condition, and one under the communal–no-benefit condition. In addition, 12 other persons were not included as subjects. Six could not finish the task within the 10 minutes allowed, one did not read the note before filling out the forms, two failed to follow instructions when filling out the forms, one discovered the concealed videotape recorder beneath the monitor, and two were married.

Finally, the true purpose of the experiment was fully explained, and the subject promised not to discuss it with anyone.

Results

A measure of liking for the other was calculated by summing the scores for each of the 11 traits and the direct measure of liking on the impressions questionnaire. The scores for the favorable traits and for the direct rating of liking were the same as the subject's ratings. The scores for the unfavorable traits were obtained by subtracting the subject's ratings for those characteristics from 20. The means for the experimental conditions for the measure of liking are presented in Table 1.

From the hypotheses, it would be expected that the scores on the measure of liking would be greater in the exchange–benefit condition than in the exchange–no-benefit condition and would be lower in the communal–benefit condition than in the communal–no-benefit condition. From Table 1 it can be seen that the results are in line with the predictions. An analysis of variance of the measure of liking revealed that the interaction between type of relationship and benefit was significant, $F(1, 92) = 8.35$, $p < .01$. Neither of the main effects approached significance. A planned comparison indicated that the difference between the exchange–benefit condition and the exchange–no-benefit condition was significant, $F(1, 92) = 4.17$, $p < .05$. A second planned comparison indicated that the difference between the communal–benefit condition and the communal–no-benefit condition was also significant, $F(1, 92) = 4.37$, $p < .05$.

A measure of anticipated pleasantness of the discussion was calculated by summing the ratings for the questions concerning how friendly, spontaneous, relaxed, enjoyable, and smooth the subjects expected the discussion to be. As expected, the means for this measure were greater for the exchange–benefit condition than for the exchange–no-benefit condition and were lower for the communal–benefit condition than for the communal–no-benefit condition. However, the interaction between the type of relationship and benefit was not significant. The main effects were also not significant.

Table 1
Means for the Measure of Liking in Experiment 1

Relationship	Benefit from the other: Benefit	Benefit from the other: No benefit
Exchange	193	176
Communal	177	194

Note. The higher the score, the greater the liking. Scores could range from 0 to 240. $n = 24$ per cell.

Discussion

The results of Experiment 1 provide support for the hypothesis that when a communal relationship is desired, a benefit following prior aid decreases attraction. When the attractive woman they had aided was unmarried (communal conditions), the unmarried male subjects liked her less when she gave them a benefit than when she did not. When the attractive woman who was aided turned out to be married (exchange conditions), she was liked more when she gave the subject a benefit than when she did not. The results for the married woman are consistent with the hypothesis that the receipt of a benefit after the other is benefited leads to greater attraction when an exchange relationship is preferred.

It might be thought that the lower liking of the unmarried woman in the benefit condition than in the no-benefit condition could be explained by a difference in the anticipation of future interaction, a variable which has been shown to influence liking in previous research (Darley & Berscheid, 1967; Mirels & Mills, 1964). Such an interpretation would make the assumption that the repayment by the unmarried woman with whom the subject wished to interact suggested that she did not wish to interact with him. However, this assumption is reasonable only if the future interaction is of the kind referred to here as a communal relationship.

If the future interaction involved an explicit exchange of benefits, there is no reason the repayment would have suggested that the unmarried women did not wish to interact with the subject. Since an interpretation in terms of the anticipation of future interaction requires a distinction between different types of interaction similar to the distinction between communal and exchange relationships, it is not an alternative explanation but essentially the same interpretation in somewhat different language.

A different interpretation might be suggested that has to do with the role relationships of males and females. It might be argued that the male subjects subscribed to a "traditional" rule that males should give gifts to females, who should gracefully accept them and not attempt to repay their benefactor, and the unmarried woman who gave them a point violated this rule. Such an interpretation would assume that the males did not apply the same rule to their relationships with a woman who is married, which does not seem consistent with traditional values concerning the role of men vis-à-vis women. If the interpretation is restricted to the relation of men and women in romantic relationships, then it is not an alternative explanation, since romantic relationships are communal relationships.

That the desire for a communal relationship was induced by creating a situation in which there was a possibility of a romantic relationship with an attractive member of the opposite sex was not fortuitous. A romantic relationship that might lead to the development of a family relationship through marriage is a particularly appropriate situation for the study of communal relationships, since relationships between family members are the most typical kind of communal relationships. However, the distinction between communal and exchange relationships is not restricted to romantic relationships with members of the opposite sex. The same effect should occur in situations in which a communal relationship, such as friendship, is desired or expected with a member of the same sex.

It was assumed not only that the other was perceived as available for a communal relationship in the communal conditions but also that she was regarded as an attractive partner for such a relationship. If the other is unattractive, a communal relationship with her should not be desired even if she is available for such a relationship. People do not desire communal relationships with people they dislike. An exchange relationship should be preferred with an unattractive other, and thus a benefit from such a person after he or she has been aided should lead to greater attraction.

Since the effect found in the first study involves the assumption that the benefit that the person received from the other is perceived as a response to the previous benefit that the other received, it should not occur if the other had not received something of value from the person. The receipt of a benefit when the other has not been aided previously should lead to greater liking when a communal relationship is expected or desired. The rule in communal relationships is to respond to a need rather than to reciprocate benefits. The giving of a benefit when no prior help has been received is appropriate for a communal relationship if there is or might be a need for the benefit.

Some evidence that the receipt of a benefit when the other is not aided previously produces greater liking when there is a communal relationship than when there is not is provided by the study by Kiesler (1966). She found that a partner on a cooperative task was liked more when he shared his winnings with the subject than when he did not share, whereas an opponent on a competitive task was liked about the same when he shared and did not share. Partnership on a cooperative task should create the expectation of a communal relationship, and in Kiesler's study the subject always lost, so the partner did not receive any aid from the subject prior to benefiting the subject. However, the possibility that the subject's losing could have been construed as a kind of aid for his opponent in the competitive conditions complicates the comparison of the cooperative and competitive conditions of Kiesler's study.

Experiment 2

The distinction between communal and exchange relationships also has implications for reactions to a request for a benefit. If it is true that in an exchange relationship any benefit given by one member to the other creates a debt or obligation to return a comparable benefit, a request for a benefit from another after one has been given aid by that other creates an opportunity to repay the debt. Thus, such a request following aid should be appropriate in an exchange relationship. Since it provides an opportunity to eliminate any tension caused by the presence of the debt, it should increase liking for the other.

The idea that the recipient of a benefit will like his or her benefactor more if he or she can return the benefit has been expressed before (Mauss, 1954). Several studies have shown that recipients of benefits like the donor more if they are able to repay the benefit than if they are not able to repay, whether the opportunity is provided by the donor's specifically requesting that the other repay the benefit (Gergen, Ellsworth, Maslach, & Seipel, 1975) or whether the opportunity to repay is provided but repayment is not specifically asked for (Castro, 1974; Gross & Latané, 1974).

In a communal relationship, to request a benefit after having given another person aid is inappropriate. It may imply that the original aid was not given with the intent of satisfying a need but rather with the expectation of receiving something in return, which may be taken as an indication that the other does not desire involvement in a communal relationship. Assuming that beginning or maintaining a communal relationship with another is desirable, such an implication should be frustrating and therefore result in decreased liking.

If one has not been previously aided by another and there is no opportunity to aid the other in the future, a request from that other is inappropriate in an exchange relationship. In an exchange relationship, a person who has not been aided by another should like the other more when he or she does not ask for a benefit than when he or she does ask for a benefit.

However, requesting a benefit in the absence of prior aid is appropriate in a communal relationship. Such a request implies that the other desires a communal relationship and, assuming that beginning or maintaining such a relationship is desirable, it should result in increased liking. Jones and Wortman (1973) suggest that asking another for a benefit is a way of conveying that we think highly of them. They say, "This tactic is likely to convey that we feel good about our relationship with the target person, since it is not customary to ask people to do favors for us unless our relationship is a relatively good one" (Jones & Wortman, 1973, p. 13).

The implications of the distinction between communal and exchange relationships for reactions to requests for a benefit in the context of prior aid or lack of prior aid were investigated in Experiment 2. A situation was arranged in which aid could be given to the subject and, later, a different benefit requested from the subject. The anticipation of a communal relationship was created by leading the subject to believe that the other wanted to meet people and that she would be discussing common interests with the other. The anticipation of an exchange relationship was created by not mentioning anything about the other wanting to meet people, implying that the other was very busy and leading the subject to believe that she would be discussing differences in interests with the other.

The following hypotheses were tested in the second experiment:

1. A request for a benefit from another after the person is aided by the other leads to greater attraction when an exchange relationship is expected.

2. When a communal relationship is expected, a request for a benefit after the person is aided by the other decreases attraction.

3. A request for a benefit in the absence of aid from the other decreases attraction when an exchange relationship is expected.

4. When a communal relationship is expected, a request for a benefit in the absence of aid from the other increases attraction.

Method

Overview. Under the guise of a study of task performance, female college students worked on a task while a television monitor showed another female working on a similar task in another room. Some of the subjects were told that the other was married, had a child, and lived far from the university, and that she and the subject would be discussing differences in interests in a second study (exchange conditions). Other subjects were told that the other was new at the university and did not know many people, and that she and the subject would be discussing common interests in a second study (communal conditions). The other female finished the task, received 1 point, and gave the subject aid on her supposedly more difficult task or did not give the subject aid. The other female then requested a point from the subject or did not request a point. Finally, the subject's liking for the other and expectations concerning the future discussion with the other were assessed.

Subjects. The subjects were 80 female, introductory psychology students who received extra credit toward their course grade for their participation. They were randomly assigned to one of the eight experimental conditions: exchange-aid-request, exchange-aid-no request, exchange-no aid-request, exchange-no aid-no request, communal-aid-request, communal-aid-no request, communal-no aid-request, and communal-no aid-no request.

Procedure. Upon arriving for the study, the subject was greeted by the experimenter and told that the other subject scheduled to participate at the same time had already arrived and was waiting in another room for the experiment to begin. The experimenter explained that it would take a little time for the equipment to warm up before the experiment could begin.

Relationship manipulation. In the communal conditions, the experimenter casually stated that the other person was anxious to begin because:

> She thinks it will be interesting. She's new at the university, doesn't know many people, and she's interested in getting to know people.

In the exchange conditions, the experimenter casually said:

> She wants to finish soon. Her husband is coming by to pick her up, then they have to pick up her child and go home to Columbia (a city some distance from the university).

The experimenter said that the first study was actually one of two short, unrelated studies they would be asked to participate in that day. The rationale for the first study was the same as in Experiment 1. The experimenter went on to say that the second study would be quite different from the first. In the communal conditions she continued:

> What we're going to do is bring you both into one room. We want you to talk over common interests. We're interested in finding out how people get to know one another. We try to create a relaxed atmosphere, and actually, in the past we've found that some of the people have gotten to know one another quite well.

In the exchange conditions she continued:

> What we're going to do is bring you both into one room. We want you to talk over differences in interests. We're doing this because most people avoid talking about differences in interests and we're interested in getting people's reactions to doing so.

Vocabulary task. After the subject had signed an experimental consent form, the procedure for the vocabulary task was explained in the same manner as in Experiment 1, except that the subject was told that she would be performing the more difficult task while the other person would be performing the easier task. The experimenter pointed out that since the subject's task was the more difficult one, she would have a chance to earn 4 points toward the extra credit, whereas the other person, Tricia, only had a chance to earn 1 point, since she had the easier task. As in Experiment 1, the experimenter mentioned that the awarding of points to maintain motivation obviously would not be necessary in the second study.

After the same instructions concerning the stopwatches as in Experiment 1, the subject was left alone to work on the task for a short time. With the 45 letters the subject had it was impossible for her to finish the task in that time. Subjects typically finished between five and seven words. During the time the subject was working, the experimenter, who always wore a lab coat so that changes in clothing over days could not be detected, could be seen on the monitor starting Tricia on her task and then leaving the room. Tricia finished her task easily. After a short time the experimenter reentered the other room, and Tricia could be seen pushing some extra letters to the front of the table. At this point the experimenter stepped in front of the camera, blocking the subject's view of the other so that the subject could not see whether the other handed the letters to the experimenter. Finally, the experimenter could be seen leaving the room, and Tricia sat back in her chair.

Aid manipulation. Shortly thereafter, the experimenter reentered the subject's room and said that Tricia had finished her task and received 1 point. The experimenter turned off the monitor, commenting that it wouldn't be needed any more. In the aid conditions she said, "Tricia asked me to give you these letters," and handed the subject some letters. In the no-aid conditions there was no mention of the letters. The experimenter then left the room, telling the subject she would be back

Table 2
Means for the Measure of Liking in Experiment 2

Relationship	Aid from & request for benefit from the other: Aid–request	Aid–no request	No aid–request	No aid–no request
Exchange	173	149	149	173
Communal	156	191	179	177

Note. The higher the score, the greater the liking. Scores could range from 0 to 240. $n = 10$ per cell.

shortly. In approximately 3 minutes, she returned, and regardless of whether the subject had finished (none in the no-aid conditions did, whereas most in the aid conditions did), she told the subject she had done well enough to receive the 4 points toward extra credit. She filled out a form indicating this and handed it to the subject. The experimenter said that was all there was to the first study, except for filling out a form if the subject wanted to request any points from Tricia. Before the subject had an opportunity to look at the form, the experimenter said she would check with Tricia and see if she wanted to fill out a form and left the room for approximately 2 minutes.

Request manipulation. When she appeared, the experimenter handed the subject a folded note, supposedly from Tricia, in which a box was checked indicating either that she wished to request 1 point from the subject (request conditions) or that she did not wish to request any points (no-request conditions). The experimenter was unaware of how the form was checked, having drawn it from a container of folded forms checked in both ways. If the subject wished to fill out a form to request points from the other, the experimenter took it.

Dependent measures. Next, the experimenter reminded the subject that the second study would involve having both subjects talk over common interests (communal conditions) or differences in interests (exchange conditions). Before starting the study it was necessary to get some idea of what their expectations were about the forthcoming interaction in order to control for those expectations, since they might vary from person to person. Therefore, she was asking the subject to fill out two forms indicating what her first impressions of the other person were and also what she expected the discussion to be like. The subject was told that these forms would be kept completely confidential.

The first-impressions form, which was given the subject on top of the form concerning expectations about this discussion, was the same as the first-impression form used in Experiment 1. The other form asked the subject to indicate how friendly, spontaneous, strained, enjoyable, and awkward she expected the discussion to be, on scales from 1 to 20.

Suspicion check. The experimenter left and watched through a one-way mirror until she could see that the subject had finished the forms. She then waited approximately 30 additional sec and reentered the room. As she picked up the forms, she casually said that there was something more to the study than she had mentioned before and asked the subject if she had any idea of what it might be. The responses of eight persons indicated suspicion of the instructions, and they were not included as subjects. Five persons suspected that the other person was not actually in the other room, one thought that the other's task had been intentionally made easy, one thought that the request did not actually come from the other person, and one thought that neither the aid nor the request came from the other person. All five people who suspected that the other person was not actually in the other room had been in another study by a different experimenter, in which they were told that another person participated but later found out that there was no such other person. Three of the eight persons were run under the exchange-aid-request condition, two under the exchange-aid-no-request condition, one under the exchange-no-aid-request condition, one under the communal-aid-request condition, and one under the communal-aid-no-request condition. In addition, four other people were not included as subjects. Two requested points from the other person, one did not read the request for a point before filling out the questionnaires, and one asked so many questions that the experimenter could not leave the room before appearing on the monitor.

Finally, the true purpose of the experiment was fully explained, and the subject promised not to discuss it with anyone.

Results

A measure of liking for the other was calculated in the same way as in Experiment 1. The means for the experimental conditions for the measure of liking are presented in Table 2. An analysis of variance of the measure of liking revealed that the main effect of type of relationship was significant, $F(1, 72) = 6.21$, $p < .05$. Neither of the other main effects nor any of the two-way interactions were significant. The three-way interaction between type of relationship, aid, and request was significant, $F(1, 72) = 12.73$, $p < .001$.

From the first hypothesis, it would be expected that scores on the measure of liking would be greater in the exchange–aid–request condition than in the exchange–aid–no-request condition. As can be seen in Table 2, the difference was in the predicted direction. A planned comparison indicated that this dif-

ference was significant, $F(1, 72) = 4.03$, $p < .05$.

From the second hypothesis, it would be expected that liking would be less in the communal–aid–request condition than in the communal–aid–no-request condition. As can be seen in Table 2, the difference was as predicted. A planned comparison indicated that this difference was significant, $F(1, 72) = 8.60$, $p < .01$.

From the third hypothesis, it would be expected that liking would be less in the exchange–no-aid–request condition than in the exchange–no-aid–no-request condition. As can be seen in Table 2, the difference was as predicted. A planned comparison indicated that this difference was significant, $F(1, 72) = 4.07$, $p < .05$.

From the fourth hypothesis, it would be expected that liking would be greater in the communal–no-aid–request condition than in the communal–no-aid–no-request condition. As can be seen in Table 2, the means for these two conditions were very similar. A planned comparison indicated that the difference between these two means was not significant.

Another way of looking at the results is to compare the aid and the no-aid conditions. It would be expected that liking would be greater in the exchange–aid–request condition than in the exchange–no-aid–request condition. As can be seen in Table 2, this expected difference was obtained. A planned comparison indicated that the difference was significant, $F(1, 72) = 3.99$, $p < .05$. It would be expected that liking would be less in the communal–aid–request condition than in the communal–no-aid–request condition. As can be seen in Table 2, this expected difference was obtained. A planned comparison indicated that the difference was marginally significant, $F(1, 72) = 3.93$, $p < .06$. It would be expected that liking would be less in the exchange–aid–no-request condition than in the exchange–no-aid–no-request condition. As can be seen in Table 2, this expected difference was obtained. A planned comparison indicated that it was significant, $F(1, 72) = 4.10$, $p < .05$. Finally, it would be expected that liking would be greater in the communal–aid–no-request condition than in the communal–no-aid–no-request condition. Although the means were in the expected direction, the planned comparison indicated that this difference was not significant.

Table 3
Means for the Measure of Anticipated Pleasantness of the Discussion in Experiment 2

Relationship	Aid from & request for benefit from the other			
	Aid–request	Aid–no request	No aid–request	No aid–no request
Exchange	56	59	58	59
Communal	63	72	70	52

Note. The higher the score, the more positive the subjects' expectations for the discussion. Scores could range from 5 to 100. $n = 10$ per cell.

A measure of anticipated pleasantness of the discussion was calculated by summing the scores on the questions concerning how friendly, spontaneous, strained, enjoyable, and awkward the subjects expected the discussion to be. The scores for the favorable characteristics were the same as the subjects' ratings. The scores for the unfavorable characteristics were obtained by subtracting the subjects' ratings for those characteristics from 21. The means for the experimental conditions for the measure of anticipated pleasantness of the discussion are presented in Table 3.

An analysis of variance of the measure of anticipated pleasantness of the discussion revealed that the main effect of type of relationship was significant, $F(1, 72) = 5.01$, $p < .05$. The main effects of aid and of request and the interactions between type of relationship and aid and between type of relationship and request were not significant. The interaction between aid and request was significant, $F(1, 72) = 7.31$, $p < .01$, and the three-way interaction between type of relationship, aid, and request was also significant, $F(1, 72) = 5.61$, $p < .05$.

As can be seen in Table 3, anticipated pleasantness was approximately the same in all four of the exchange conditions, but there were differences within the communal conditions. Anticipated pleasantness was lower in the communal–aid–request condition than in

the communal–aid–no-request condition. A simple comparison indicated that this difference approached significance, $F(1, 72) = 2.99$, $p < .10$. Anticipated pleasantness was greater in the communal–no aid–request condition than in the communal–no-aid–no-request condition. A simple comparison of this difference was significant, $F(1, 72) = 11.19$, $p < .001$.

Discussion

In general, the results supported the hypotheses concerning reactions to a request for a benefit based on the distinction between communal and exchange relationships. As expected from the hypothesis that a request for a benefit after the person is aided by the other leads to greater attraction when an exchange relationship is expected, it was found that liking for the other was higher in the exchange–aid–request condition than in the exchange–aid–no-request condition. As predicted from the hypothesis that a request for a benefit after the person is aided decreases attraction when a communal relationship is expected, liking was lower in the communal–aid–request condition than in the communal–aid–no-request condition. In line with the hypothesis that a request for a benefit in the absence of aid from the other decreases attraction when an exchange relationship is expected, liking for the other was lower in the exchange–no-aid–request condition than in the exchange–no-aid–no-request condition.

The hypothesis that a request for a benefit in the absence of aid from the other increases attraction when a communal relationship is expected was not supported; there was no difference in liking between the communal–no-aid–request condition and the communal–no-aid–no-request condition. The subjects in the communal–no-aid–request condition may have been somewhat uncertain about the intentions of the other. Although the request may have indicated to the subject that the other wanted a communal relationship with her and consequently led the subject to expect such a relationship, it also may have reminded the subject that the other had not given her aid earlier. This reminder may have raised doubt about whether the other would behave in an appropriate way for a communal relationship. This could explain why the request did not result in increased liking for the other in the communal–no-aid–request condition.

As would be expected from the distinction between communal and exchange relationships, liking was greater in the exchange–aid–request condition than in the exchange–no-aid–request condition, marginally less in the communal–aid–request condition than in the communal–no-aid–request condition, and less in the exchange–aid–no-request condition than in the exchange–no aid–no-request condition. The greater liking in the exchange–aid–request condition than in the exchange–no-aid–request condition could be due to a general tendency for aid to increase liking, as well as to the appropriateness of the request. However, the fact that liking was less in the exchange–aid–no-request condition than in the exchange–no-aid–no-request condition is opposite to what would be expected from a general tendency for aid to increase liking, but follows from the idea that differences in liking are due to the appropriateness of the other's behavior for the type of relationship. That liking was less in the communal–aid–request condition than in the communal–no-aid–request condition is also opposite to the tendency for aid to increase liking and consistent with the effect on liking of the appropriateness of the other's behavior for the type of relationship.

Since the focus was on the interactive effects of type of relationship, a main effect of type of relationship was not specifically predicted. However, the significant main effect that was obtained for liking is understandable in view of the operations used to manipulate type of relationship. Among other things, subjects in the communal conditions may have expected the other to be more similar than did subjects in the exchange conditions.

The results for the measure of anticipated pleasantness of the interaction were also consistent with the distinction between communal and exchange relationships. Anticipated pleasantness was similar in the four exchange conditions, marginally less in the communal–aid–request condition than in the communal–

aid–no-request condition, and greater in the communal–no-aid–request condition than in the communal–no-aid–no-request condition.

The results for the measure of pleasantness can be understood in terms of the assumption that one will anticipate interaction with the other to be more pleasant if one expects a communal relationship than if one expects an exchange relationship. In the exchange condition, subjects were led to expect an exchange relationship by the experimental instructions given in those conditions. Subjects in the communal–aid–request condition should have expected an exchange relationship, because the rules of a communal relationship were violated by the request following the prior aid. Subjects in the communal–no-aid–no-request condition should also have expected an exchange relationship, because the rules of a communal relationship were violated by the other's failure to respond to their needs or to request something from the subject for which the other presumably had a need.

The interpretation of the results for the measure of anticipated pleasantness might appear inconsistent with the fact that liking was not significantly higher in the communal–no-aid–request condition than in the communal–no-aid–no-request condition. However, it is possible that the request for a benefit in the absence of prior aid was sufficient to create an expectation of a communal relationship yet insufficient to increase liking. As mentioned earlier, the request for aid in the communal–no-aid–request condition may have been taken as an indication that the other felt positively toward the subject and thus led the subject to expect a communal relationship with that other, but it may also have reminded the subject that the other had not fulfilled her needs earlier, resulting in ambivalent feelings toward the other.

General Discussion

While it is assumed that the distinction between communal and exchange relationships is made implicitly by most people in their interactions with others, it is not assumed that they are explicitly aware of the distinction or are able to describe how it affects their reactions. Certainly they do not use the terms *communal* and *exchange relationships*. It is also not assumed that everyone makes the distinction in the same way. Some people restrict their communal relationships to only a very few persons, whereas others have communal relationships with a wide circle of others. There are some people who do not make the distinction at all. Some people treat every relationship, even relationships with members of their own immediate family, in terms of exchange.

It is possible for a person to have both a communal relationship and an exchange relationship with the same other, for example, when a person sells something to a friend or hires a family member as an employee. In such instances, a distinction is typically made between what is appropriate for the business (exchange) relationship and what is appropriate for the family or friendship (communal) relationship. Exchange relationships sometimes can develop into communal relationships, such as when a merchant and a customer become close friends or when an employer and an employee marry.

The lack of attention paid to communal relationships in previous research on interpersonal attraction may be accounted for by the fact that almost all of the past research has involved attraction toward persons who are not only previously unknown to the subject but who are not expected by the subject ever to be known in the future. Communal relationships involve an expectation of a long-term relationship, whereas exchange relationships need not be long-term. However, the variables of communal versus exchange relationship and expected length of the relationship are conceptually independent. Exchange relationships may be expected to continue over a long period.

If it is true that treating a communal relationship in terms of exchange compromises the relationship, then exchange theories of interpersonal attraction (e.g., Secord & Backman, 1974, chapter 7) may create a misleading impression about the development and breakup of intimate relationships. The idea that exchange is the basis of intimate relationships may actually have the effect of impairing such relationships. For example, the rec-

ommendation, which seems to be growing in popularity, that prior to marriage a marriage contract be drawn up that specifies in detail what each partner expects from the other, should, if followed, tend to undermine the relationship.

If the theoretical viewpoint of this research is correct, a communal relationship will be strained by dickering about what each of the partners will do for the other. Of course, if one of the partners in a communal relationship is convinced that he or she is being exploited by the other because that person is concerned about the other's welfare while the other is not concerned about his or her welfare, the communal relationship has disintegrated. If this happens in a marriage, there may be attempts to preserve the marriage by changing it into an exchange relationship through dickering.

References

Adams, J. S. Toward an understanding of inequity. *Journal of Abnormal and Social Psychology*, 1963, *67*, 422–436.

Castro, M. A. C. Reactions to receiving aid as a function of cost to donor and opportunity to aid. *Journal of Applied Social Psychology*, 1974, *4*, 194–209.

Darley, J. M., & Berscheid, E. Increased liking as a result of the anticipation of personal contact. *Human Relations*, 1967, *20*, 29–40.

Gergen, K. J., Ellsworth, P., Maslach, C., & Seipel, M. Obligation, donor resources, and reactions to aid in three cultures. *Journal of Personality and Social Psychology*, 1975, *31*, 390–400.

Goffman, E. *Asylums*. Garden City, N.Y.: Anchor Books, 1961.

Gross, A. E., & Latané, J. G. Receiving help, reciprocation, and interpersonal attraction. *Journal of Applied Social Psychology*, 1974, *4*, 210–223.

Jones, E. E., & Wortman, C. *Ingratiation: An attributional approach*. Morristown, N.J.: General Learning Press, 1973.

Keisler, S. B. The effect of perceived role requirements on reactions to favor doing. *Journal of Experimental Social Psychology*, 1966, *2*, 198–210.

Mauss, M. *The gift: Forms and functions of exchange in archaic societies*. Glencoe, Ill.: Free Press, 1954.

Mirels, H., & Mills, J. Perception of the pleasantness and competence of a partner. *Journal of Abnormal and Social Psychology*, 1964, *68*, 456–459.

Pruitt, D. G. Methods for resolving differences of interest: A theoretical analysis. *Journal of Social Issues*, 1972, *28*, 133–154.

Secord, P. F., & Backman, C. W. *Social psychology* (2nd ed.). New York: McGraw-Hill, 1974.

Thibaut, J. W., & Kelley, H. H. *The social psychology of groups*. N.Y.: Wiley, 1959.

Walster, E., Walster, G. W., & Berscheid, E. *Equity: Theory and research*. Boston: Allyn & Bacon, 1978.

Received November 7, 1977 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 1, 25-38

Person Memory: Personality Traits as Organizing Principles in Memory for Behaviors

Reid Hastie and Purohit Anand Kumar
Harvard University

Subjects studied and recalled sentences describing behaviors while performing a laboratory impression-formation task. Recall was high for behaviors that were incongruent with a personality-trait impression for a character, whereas recall was much lower for behaviors that were congruent or neutral with reference to the impression. Set size, the number of congruent and incongruent behaviors attributed to the character, was shown to be a major determinant of this result. The smaller the size of the incongruent set, the higher the probability of recalling an item from the set. There was no tendency for behaviors to cluster by trait category in recall output protocols. This result was interpreted as evidence that a simple analogy to hierarchical noun categories, studied in many verbal learning experiments on organization of memory, did not apply to the present results. Three theoretical analyses—an associative network model, a depth-of-processing model, and a schema model—are reviewed in light of these results.

Psychologists have been interested for a long time in the effects of abstract organizing principles on memory. Sir Frederic Bartlett's (1932) schema theory is one of the earliest efforts to explicate the relations between an abstract structure and the recall of specific facts. Bartlett proposed that memory is not primarily or literally reduplicative or reproductive. Rather, he argued that perceived events are assimilated to mental schemata that have been formed by similar events experienced in the past. Bartlett's classic treatise on remembering is replete with examples of phenomena that he believed evidenced the workings of these schemata.

Researchers have studied the manner in which experimental subjects use well-defined organizational structures in numerous memory tasks. Miller (1956), Mandler (1967), Tulving (1968), Bower (1970), and Rosch (1973) conducted pioneering research on the operation of semantic structures in the recall of words; Bransford and his colleagues (Bransford & Franks, 1972; Bransford & Johnson, 1973) have studied the operation of linguistic and spatial schemata in recognition and recall; and Kintsch (1974), Thorndyke (1977), and Mandler and Johnson (1977) have studied the operations of schemata in memory for prose paragraphs. One consistent conclusion from these experiments is that recall of specific facts is affected by their relation to an abstract organizing principle, theme, or schema.

Psychologists interested in social judgment have also studied organization—the organization of attitudes and beliefs (Abelson et al., 1968; McGuire, 1968), the perception of social groups (Heider, 1958; Insko, Songer, & McGarvey, 1974), and the organization of impressions (Anderson, 1974b; Hastorf, Schneider, & Polefka, 1970; Rosenberg & Sedlak, 1972a). Recently, researchers have attempted to integrate the two traditions to

This research was supported by National Institute of Mental Health Grant MH-28928 to Harvard University. We would like to thank Joan Dobrof, John Porter, and Barbara Signer for their assistance with the research and David Kenny, Steven Penrod, and Shelley Taylor for comments on this manuscript. Final revision of the manuscript was facilitated considerably by Anthony Greenwald and two anonymous reviewers.

Requests for reprints should be addressed to Reid Hastie, Department of Psychology and Social Relations, Harvard University, Cambridge, Massachusetts 02138.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3701-0025$00.75

study the manner in which social information is stored and retrieved from memory. Anderson and Hastie (1974), Sulin and Dooling (1974), D'Andrade (1974), Lewis and Anderson (1976), Cantor and Mischel (1977), Markus (1977), Snyder and Uranowitz (1978), and Rogers, Kuiper, and Kirker (1977) have studied the operation of general organizing principles, themes, or prototypes in the perception and memory of information about people. Several of these researchers have suggested that when information is stored in memory about an individual person, personality-trait dimensions or personality prototypes provide the mental categories for the new information. Thus, inferences about abstract personality traits will organize the perception, storage, and retrieval of information about people.

In the three experiments reported in this article, subjects studied and recalled lists of behavior descriptions. The behaviors in each list were attributed to a single fictional character, along with a brief personality-trait sketch. The behavior descriptions were selected to be congruent or incongruent with respect to the personality sketch, and the focus of the experiments was on differential recall of these descriptions.

Differential recall of specific behaviors varying in congruence to the general personality impression would be taken as evidence that the impression controls the encoding, retention, or retrieval of social information. Results from the research cited above imply that information that confirms or is congruent with the impression will be more accurately recalled than information that is incongruent with the impression. Accordingly, our analysis will focus on the relative recall of congruent and incongruent behavior descriptions.

Experiment 1

Method

Design. Each subject studied and recalled sentences from six 20-sentence lists. Each list was comprised of 12 sentences describing behaviors congruent with a personality trait, 4 describing behaviors that were incongruent with the trait, and 4 describing behaviors that were neutral with respect to the trait. The three types of sentences were distributed uniformly across input serial positions. Each of the six lists of behaviors studied by a subject was associated with a different personality trait. There were 12 different traits in the initial set of experimental materials, and so two sets of 6 traits each were arbitrarily created to produce two replications of the design. Finally, each of the 6 subjects studying one of the 6-trait subsets received the 6 traits in a different order. These orders were arranged in a Latin square plan so that each trait occurred in a different within-session position for each subject. This design was replicated with a second set of 12 subjects.

Materials. The memory experiment outlined above required sets of sentences describing behaviors characteristic of a representative sample of personality traits. First, 12 traits were chosen from the 80 traits studied by Rosenberg and Sedlak (1972b): *intelligent–unintelligent, honest–liar, irresponsible–conscientious, friendly–hostile, aggressive–shy,* and *naive–cynical*). The sample was composed of six pairs of "opposite meaning" traits that spanned the space revealed by Rosenberg and Sedlak's multidimensional scaling analysis. The pairs were presented to 12 pretest subjects with an instruction of the form, "Consider a person who is very *intelligent*; you would expect to see him (her). . . ." The resulting lists of pretest subject-generated behaviors were edited to yield 12 three-to-five-word behavior descriptions for each of the traits. Some examples of behaviors generated as congruent with the intelligent trait are "won the chess tournament" and "attended the symphony concert." In addition, the authors generated 48 "neutral" behavior descriptions. These were sentences describing behaviors that were neutral or that had few or no implications about personality. Some examples of neutral behavior descriptions are "ordered a cheeseburger for lunch" and "took the elevator to the third floor." All behavior descriptions were written to describe specific, easy-to-picture events. Finally, four incongruent behaviors were selected for each trait by sampling from the set of behaviors associated with the opposite-meaning trait from each pair (e.g., four behaviors congruent with the unintelligent trait were selected to be incongruent behaviors for the intelligent trait, and so forth). Some examples of incongruent behaviors for the intelligent trait (actually congruent behaviors for the *un*intelligent trait) are "made the same mistake three times" and "was confused by the television show." This list-generation plan provides a control for idiosyncratic characteristics of particular trait-related sentences. For instance, congruent sentences for the intelligent character presented to one subject also appear as incongruent sentences for the unintelligent character presented to another subject.

During the memory task, each list of trait-related behaviors was preceded by an ensemble of eight trait adjectives presented to induce the unitary appropriate personality impression for the list. These adjectives were selected from Anderson's (1968) list of trait terms to be synonymous or close in meaning to the original trait for each list (e.g., for the intelligent list, *intelligent, clever, bright, smart, quick,*

wise, knowledgeable, and *decisive* formed the adjective ensemble). All of the materials were typed on 12.7 cm × 20.3 cm index cards for display to the subjects.

Subjects. The subjects were 24 undergraduate university students. They were paid $1.50 each for their participation in the experiment. Approximately equal numbers of male and female subjects participated in all of the research reported in this article. In no case was there a statistically reliable effect of the sex variable on measures of memory or judgment, and so there will be no further mention of sex differences in this report.

Procedure. The individual experimental sessions lasted approximately 30 minutes. Subjects were told that they would be asked to form impressions and recall information about six fictional characters. On each of the six experimental trials, the subject performed four tasks. First, the subject read aloud the eight trait adjectives that preceded each list. Second, he or she read aloud 20 sentences, presented at a 5-sec-per-sentence rate, describing behaviors performed by the character. Third, the subject recalled as many of the sentences as possible in writing under free-recall instructions. Finally, he or she rated the fictional character's personality on nine bipolar scales. This rating task was included to emphasize the impression-formation task. The results of all rating tasks were consistent with averaging or adding models for impression formation—the more behaviors presented that were congruent with a particular trait, the higher the rating of the stimulus character on that dimension. Beyond this simple finding, results of the rating task were not related to sentence recall, and so they will not be discussed in this article.

Results

The primary data from the experiment are the proportions of behavior descriptions recalled from the congruent, incongruent, and neutral subsets of each list. A note on the scoring method for the behavior descriptions is necessary. The descriptions were simple phrases such as "won the chess tournament," "attended the symphony concert," and so forth. First, a lenient scoring criterion was employed: An item was scored as correctly recalled if the general sense of the description was preserved (i.e., any changes from the original were substitutions of synonymous words). The data reported in this article were tabulated following this criterion. Second, a strict scoring criterion was adopted: An item was scored as correctly recalled if it differed from the original in at most one important word (and if this deviation did not affect the meaning of the phrase). All analyses were repeated with this criterion, and no discrepancies from the present summary were apparent.

Table 1
Mean Proportions of Behavior Descriptions Recalled in Experiment 1

Input serial position	Item type		
	Congruent behaviors	Incongruent behaviors	Neutral behaviors
1 to 5	.47	.57	.38
6 to 10	.33	.48	.36
11 to 15	.35	.47	.27
16 to 20	.60	.61	.51
Total	.43	.54	.38
n	1728	576	576

The proportions of recalled data are displayed in Table 1, broken down by item type and input serial positions. The single striking effect in the data is the superior recall of incongruent behavior descriptions in comparison to recall of congruent and neutral descriptions, $F(2, 40) = 18.51$, $p < .01$. A multiple comparison test following the Tukey (a) procedure indicated that the incongruent mean recall was reliably higher than the mean proportions recalled from congruent and neutral subsets ($p < .01$), although recall from these latter two subsets did not differ reliably. Twenty-one of the 24 subjects recalled a higher proportion of incongruent behaviors than either of the other two item types. The serial position data show that the superior recall of incongruent behaviors is most pronounced in the middle serial positions of the list and vanishes in the recency portion of the list. A statistical test of the Item Type × Serial Position interaction did not yield a significant result, $F(6, 120) = .98$, *ns,* and so the relationship between recall and serial position has not been clearly established.

The pattern of recall did not vary systematically from the first half of the experimental session to the second. However, the low intrusion rate (recall of nonpresented behaviors), less than one item per trial, did increase slightly over the course of the session. A separate analysis on the order of recall of the three item types to determine whether congruent items were more likely to precede or

follow incongruent items did not suggest that congruent, incongruent, or neutral behaviors were systematically ordered at recall.

An index of recall clustering by item type (congruent, neutral, incongruent) was calculated for each recall protocol (observed stimulus category repetition − expected stimulus category repetition; Bousfield & Bousfield, 1966). An analysis of variance performed on these data revealed no tendency to order recall of behaviors according to the item type (the grand mean of the clustering index was not significantly different from zero, $F(1, 20) = .82$, *ns*.

Discussion

The results of this experiment are evidence for the importance of personality traits as organizing principles in memory for facts about people. Recall of particular behavior descriptions is clearly affected by their relation (congruent, incongruent, neutral) to the salient trait. It is surprising that the incongruent behaviors are so well recalled. Simple versions of schema theories (Bartlett, 1932; Bransford & Franks, 1972; Mandler & Johnson, 1977), organization theories (Bower, 1970; Tulving, 1968), and prototype theories (Cantor & Mischel, 1977; Rosch, 1973) would seem to predict that fitting, congruent, or consistent information would be better recalled than incongruent information. Furthermore, some of these theories would predict that related facts would be recalled together. The failure to observe clustering by trait categories disconfirms this prediction as well.

The present results are related to two phenomena observed in verbal learning research with more traditional list materials. First, it may be that the differential recall of congruent and incongruent behaviors is best viewed as a case of the familiar list-length effect in free recall: As list length increases, the proportion of items recalled decreases (Murdock, 1962). Since there are more congruent behaviors than incongruent ones, proportional recall of congruent items is lower. One problem with this account is the low recall of neutral items that occur with the same frequency as the well-remembered incongruent items. However, it may be that subjects notice incongruent items and class them apart from congruent and neutral behaviors without distinguishing between the latter two types of items. Second, the pattern of recall for the incongruent items resembled the recall of striking "von Restorff" items in certain respects—for instance, the apparent primacy effect in the serial position curve for the incongruent items from early serial positions (Bellezza & Cheney, 1973).

It is important to note that these verbal learning interpretations of the recall results assume that congruent and incongruent items are perceived and mentally categorized by the abstract trait impression and its relation to the individual behaviors. Thus, set size and von Restorff interpretations imply that personality-trait impressions are organizing principles controlling memory for the behavioral information.

Experiment 2

A second experiment was designed to explore the degree to which set size determined the recall of congruent and incongruent behaviors in Experiment 1. Experiments 2 and 3 included conditions in which lists were comprised of equal numbers of congruent and incongruent behavior descriptions. The "pure" effect of the initial trait ensemble on memory, unconfounded with set-size effects, could be assessed under these conditions. These experiments employed the same populations of subjects, task, and materials as Experiment 1, but this time the relative sizes of the congruent and incongruent behavior subsets within each list were varied.

Method

Design. Each subject studied and recalled sentences from eight 16-sentence lists. Each list was associated with personality trait and included four neutral behavior descriptions with varying numbers of congruent and incongruent behaviors. There were four types of lists containing (a) 12 congruent items and no incongruent items; (b) 11 congruent and 1 incongruent items; (c) 9 congruent and 3 incongruent items; and (d) 6 congruent and 6 incongruent items. Each subject studied two lists of each type. Precautions were taken to counterbalance the materials against serial position and list-sequence effects as in Experiment 1.

Materials. The materials from Experiment 1, with the addition of two trait pairs (*generous–stingy* and *calm–nervous*) were used in this experiment. Once again, each list was preceded by an ensemble of eight traits to establish an initial personality impression.

Subjects. The subjects were 24 undergraduate university students paid $1.50 each to participate in the 30-minute experimental session.

Procedure. Instructions and procedures were the same as in Experiment 1.

Results

The primary data from Experiment 2 are the proportions of behavior descriptions recalled. These data are displayed in Table 2 classified according to list type (12/4/0, 11/4/1, 9/4/3, and 6/4/6 congruent/neutral/incongruent behavior subset sizes) and item type (congruent, neutral, and incongruent behaviors). An analysis of variance of the data from the three list types in which all three types of items were presented revealed a large main effect of item type on the proportion of items recalled, $F(2, 36) = 21.43$, $p < .001$. It is clear from these results that the superior recall of incongruent items was almost completely determined by subset size. The fewer incongruent items in a list, the higher the probability of recalling any particular item. An analysis of variance of the proportion-of-incongruent-items-recalled data yielded a significant linear trend across levels of the set-size factor, $F(1, 18) = 10.21$, $p < .01$. The probability of recalling congruent or neutral behaviors did not depend on set size: $F(3, 54) = 2.24$, *ns*, for congruent recall; $F(3, 54) = 1.09$, *ns*, for neutral item recall. Serial position curves were extremely unstable and will not be described for this experiment.

Again, as in Experiment 1, the Bousfield and Bousfield (1966) clustering index was not reliably greater than zero, $F(1, 18) = .79$, *ns*, indicating that subjects were not grouping items by trait categories in recall.

Experiment 3

A third experiment using a new sample of behaviors and four new traits was designed as a systematic replication of Experiment 2. Furthermore, the experiment was designed to obtain stable serial position curves.

Method

Design. Each subject studied and recalled sentences from four 14-sentence lists. Each list was associated with a personality trait and included a mixture of congruent and incongruent behaviors. There were four types of lists containing (a) 13 congruent items and 1 incongruent item; (b) 11 congruent and 3 incongruent items; (c) 9 congruent and 5 incongruent items; and (d) 7 congruent and 7 incongruent items. Each subject studied one list of each type. Precautions were taken to counterbalance the materials against serial position and sequence effects.

Materials. Sentences describing behaviors characteristic of four personality traits were generated by pretest subjects, as in Experiment 1. The four traits, *intelligent, stupid, friendly,* and *hostile,* were selected to span Rosenberg and Sedlak's (1972a) trait space representing poles of their social and intellectual dimensions. A separate sample of 48 subjects rated each of the behavior descriptions on intelligence, friendliness, and imageability scales.

During the memory experiment, each list was preceded by an ensemble of five trait adjectives selected to be close to the key trait in meaning (for example,

Table 2
Mean Proportions of Behavior Descriptions Recalled in Experiment 2

List type	Item type: Congruent behaviors *M*	Congruent behaviors *n*	Incongruent behaviors *M*	Incongruent behaviors *n*	Neutral behaviors *M*	Neutral behaviors *n*
12 congruent/0 incongruent/4 neutral	.56	576	—	—	.48	192
11 congruent/1 incongruent/4 neutral	.55	528	.77	48	.47	192
9 congruent/3 incongruent/4 neutral	.49	432	.61	144	.42	192
6 congruent/6 incongruent/4 neutral	.50	288	.59	288	.42	192

Table 3
Mean Proportions of Behavior Descriptions Recalled in Experiment 3

List type	Item type: Congruent behaviors *M*	*n*	Incongruent behaviors *M*	*n*
13 congruent/ 1 incongruent	.60	624	.79	48
11 congruent/ 3 incongruent	.62	528	.75	144
9 congruent/ 5 incongruent	.60	432	.71	240
7 congruent/ 7 incongruent	.61	336	.67	336

intelligent, knowledgeable, reliable, skillful, persistent, and *scientific*).

Subjects. The subjects were 48 undergraduate university students. They were paid a wage of $2 each to participate in the 1-hour experimental session.

Procedure. The experimental sessions followed the procedure used in the preceding experiments. Subjects studied trait and behavioral information about a character, recalled the behaviors, and then rated their impression of the character. These tasks were repeated for each of four characters.

Results

The pattern of recall data replicated the results of Experiment 2 in almost every detail. These data are displayed in Table 3 classified according to list type (13/1, 11/3, 9/5, 7/7 congruent/incongruent behavior subset sizes) and item type (congruent or incongruent behavior).

Again, a large set-size effect appeared for the recall of incongruent items, such that the larger the set, the less likely that an item was recalled, linear trend $F(1, 47) = 3.80$, $p < .05$. No such effect was apparent in the congruent-item recall data, main effect $F(3, 141) = .35$, *ns*.

A further breakdown of these data yielded the serial position curves displayed in Figure 1. These curves show that the recall advantage of incongruent items over congruent items is most apparent in the central serial positions of the list. Further, when several incongruent items appear, the earliest presented incongruent items have the greatest advantage.

The design of Experiment 3 did not allow a direct analysis of variance test of the reliability of the Serial Position × Item Type × Set Size interaction that is implied by this summary of the serial position data. Twelve pseudosubject units were created by aggregating data from subsets of 4 subjects (in the original sample of 48) to permit an analysis of variance test. For each pseudosubject, the serial position data were combined to yield four Probability of Recall × Serial Position indexes (the first data from true Serial Positions 1 to 3, the second from Positions 4 to 7, the third from Positions 8 to 11, and the fourth from Positions 12 to 14). Thus, an analysis was performed *as if* 12 subjects had each provided data from recall of four serial position locations from four lists (each with a different mixture of congruent and incongruent set sizes). This analysis yielded a significant main effect of serial position on recall, $F(3, 33) = 6.83$, $p < .001$; a significant Item Type × Serial Position effect, $F(3, 33) = 3.04$, $p < .05$; and a significant Item Type × Serial Position × Set Size interaction, $F(9, 99) = 2.70$, $p < .01$—all consistent with our summary of the serial position results.

Again, an analysis of the Bousfield and Bousfield (1966) clustering index produced no evidence of grouping behaviors by trait category in recall, $F(1, 47) = 1.02$, *ns*.

A test of the "pure" trait ensemble effects on recall was performed to answer the question: When congruent and incongruent set sizes are equal, is incongruent-item recall reliably greater than congruent-item recall? An analysis of variance was performed on the proportion of recalled data from equal set-size lists in Experiments 2 and 3 (6 vs. 6 lists in Experiments 2 and 7 vs. 7 lists in Experiment 3). The item (congruent vs. incongruent) effect is significant, indicating that although the absolute difference is small (.09 in Experiment 2 and .06 in Experiment 3) incongruent behaviors are recalled at a higher rate then congruent behaviors, $F(1, 66) = 6.14$, $p < .01$. In Experiment 2, 5 sub-

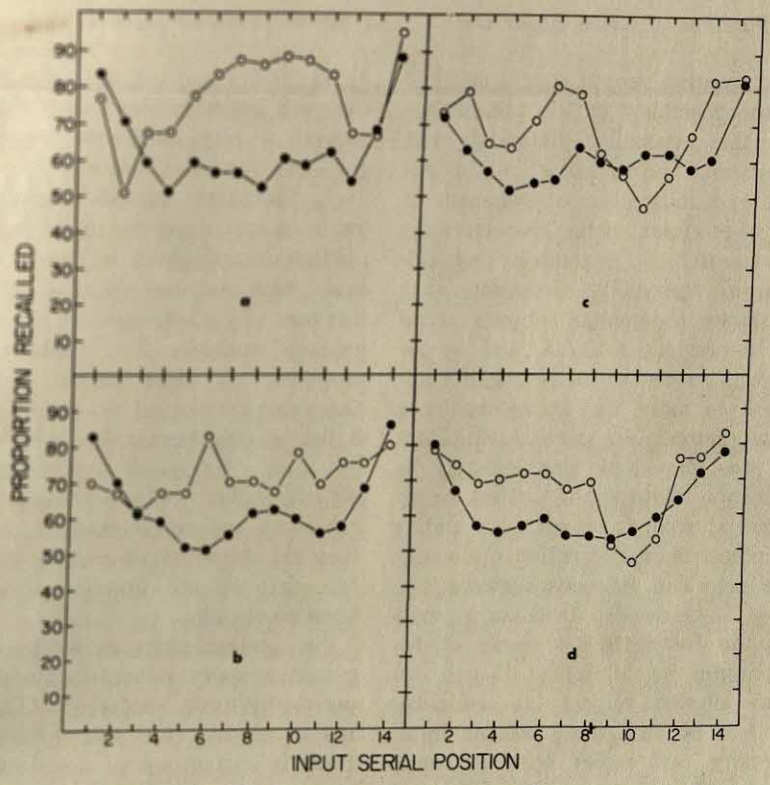


Figure 1. Serial position curves for Experiment 3. (Congruent behavior recall is indicated by filled circles, incongruent behavior recall by unfilled circles. The panels present data from the four set-size conditions in the experiment: [a] 13 congruent items and 1 incongruent item; [b] 11 congruent items and 3 incongruent items; [c] 9 congruent items and 5 incongruent items; [d] 7 congruent items and 7 incongruent items.)

jects recalled equal numbers of congruent and incongruent items, 5 recalled more congruent than incongruent items, and 14 recalled more incongruent items than congruent ones. In Experiment 3, 8 subjects recalled equal numbers, 11 recalled more congruent items, and 29 recalled more incongruent items.

General Discussion

The research reported in this article has established four results about the recall of information about people. First, specific acts that are unexpected or incongruent with reference to a person's general impression are well remembered compared to acts that are unsurprising or congruent. Neutral acts that are uninformative for the impression-judgment task are least well recalled. Second, the primary determinant of the differential recall of congruent and incongruent acts is the relative number of each type of act performed by the person. When the person performs only a few incongruent acts, the relative recall of incongruent items is quite high. When the person performs equal numbers of congruent and incongruent acts, the relative recall advantage of incongruent acts is much smaller. Third, differential recall of congruent and incongruent acts is most pronounced in the central portions of the serial position sequence. Both types of acts are well recalled from early (primacy) and final (recency) serial positions. Fourth, there was no grouping of items in recall by trait categories. Clustering measures revealed no tendency to recall congruent and incongruent acts grouped in any orderly pattern.

Conflicting Findings of Other Studies

We should mention several conditions that may limit the generality of the present results. First, the personality dimension that defines congruence–incongruence in the experimenter's eyes must also be prominent in the subject's perception of the characters. In the present research, we guaranteed this condition by using personality dimensions that have been shown to dominate subjects' social judgments in previous research, and we required subjects to perform a personality-impression task to make the dimensions even more salient. Second, the pace at which information was presented was leisurely in the experiments, allowing the subject to reflect for several seconds on each act and to distribute rehearsal and attention optionally. Third, the retention intervals between presentation of to-be-recalled behavior descriptions and the free-recall test were brief—under 5 minutes for all items. Fourth, the instructions allowed subjects to anticipate the recall tests so that recall was an intentional memory test rather than an unexpected, incidental memory test. Fifth, all memory tests followed free-recall procedures. It may be that the present results would not generalize to cued recalled or recognition test conditions (cf. Cantor & Mischel, 1977).

Several experiments (Bear & Hodun, 1975; Cantor & Mischel, 1977; Greenwald & Sakumura, 1967; Picek, Sherman, & Shiffrin, 1975; Snyder & Uranowitz, 1978; Zadny & Gerard, 1974) have found that information that is consistent with an impression or expectation is better remembered than inconsistent information. Each of these experiments differs from the present research on one or more of the conditions we have listed. Furthermore, several experiments have found that novel (Greenwald & Sakumura, 1967), distinctive (Hamilton & Gifford, 1976), or schema-incongruent (Smith, 1973) information is well remembered. The obvious path for future researchers is to identify the variables that interact with degree of consistency or congruence to produce one pattern of results or another.

A Simple Network Model

Our first cognitive account for these results will begin by describing the interaction between a very long-term conceptual memory store for social information and a relatively short-term episodic memory structure for facts about specific individuals. This distinction is analogous to the distinction between semantic and episodic memory stores proposed by Tulving (1972). Conceptual memory controls the "pattern recognition processes" of identification, comprehension, and categorization of on-going behavior, and it has usually been called implicit personality theory by social psychologists (Rosenberg & Sedlak, 1972a; Schneider, 1973). As behaviors are perceived and comprehended, they are stored in an episodic memory structure that we will describe as a network of associative links.

For present purposes we hypothesize that person memory structures are dominated by personality-trait categories. There are two major reasons for this assumption. First, there is a tradition of describing conceptual person memory, implicit personality theory, as a multidimensional space of trait concepts. Second, the behaviors used as stimulus materials in these experiments were generated by pretest subjects given trait terms as cues. It is difficult to conceive of an alternative standard to the principle of trait congruence that could account for subjects' reliable differential treatment of congruent and incongruent behaviors. It is possible that "behavioral scripts" or schematic images (Schank & Abelson, 1977) are more accurate representations of social categories than trait concepts. However, trait categories seem most appropriate in the present research.

The first analysis of episodic person memory is a network model after the fashion of John Anderson's human associative memory (HAM) theory (Anderson, 1976; Anderson & Bower, 1973; Anderson & Hastie, 1974). This application of HAM is brief, but it would be easy to elaborate the model following the conventions outlined in Anderson's work. First, we suppose that the subject

studying the list of facts about one of our characters produces a mental structure in episodic memory best described as a network of associative *links* and idea *nodes*. We hypothesize that a typical structure encoded in our experimental task has a hierarchical structure of the form depicted in Figure 2. At the highest (entry point) level in the structure are features or ideas such as the individual's proper name, definite descriptions ("the best lawyer in town"), or perhaps a representation of the individual's physical appearance. At the second, intermediate, level of the structure are organizing principles—trait labels in the present experiments. Finally, at the lowest level in the structure are specific facts or attributes of the individual—behavior descriptions in this research. (The HAM analysis actually provides for a much finer-grained analysis of this structure.) Thus, we hypothesize that behaviors are classified into trait categories and stored accordingly during the study phase of our experiments. These assumptions are plausible if we keep in mind that the experimental materials were generated such that each behavior was distinctly and uniquely associated with a single trait category, that incongruent acts were strikingly opposed to congruent acts, and that list presentation was at a leisurely pace so that subjects had time to review each act carefully.

Second, we suppose that during the retention interval, a matter of seconds or minutes in these experiments, links slowly disintegrate or fade from memory.

Third, we postulate a variety of processing rules that operate on the stored memory structure when the signal to recall facts about an individual is given. (a) A search process starts at the highest (proper name) node in the structure and traverses the associative links downwards until it reaches a terminal, behavior description, node. (b) Upon reaching a behavior description, the subject reviews it, to see if that item has been recalled previously, and writes the description on the response form. In our example (Figure 2), the search process would start at the node labeled "James Bartlett," might take the left-hand, "honest trait" path, and then select the path terminating in the behavior description "admitted he caused the accident." (c) If the act has already been recalled, the subject does not write it down. (d) After a behavior has been reviewed, the search process starts over, *from the highest* (entry point) *node*. (e) The probability that the search process chooses a particular link as it passes through a node is determined by

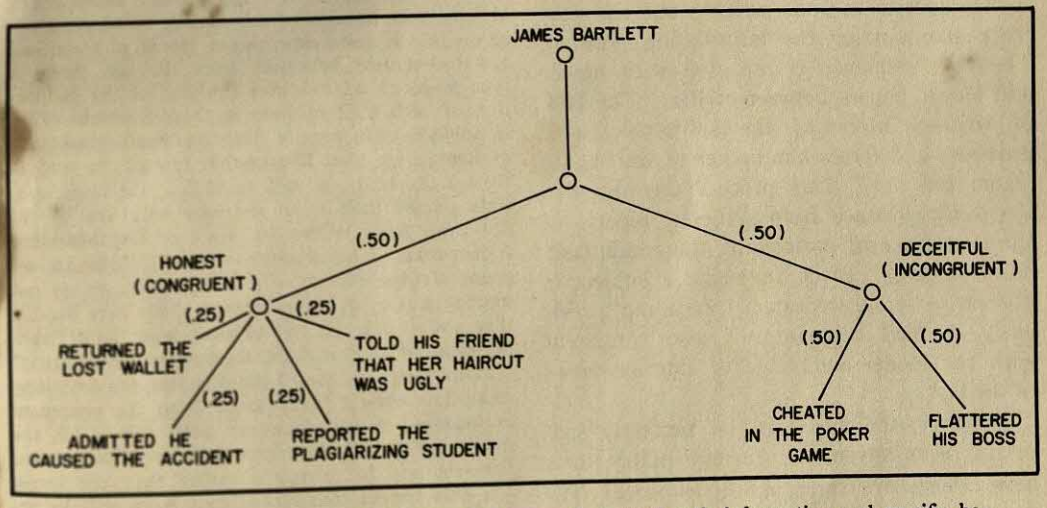


Figure 2. A hierarchical network structure to represent abstract trait information and specific behaviors in long-term memory. (Numbers in parentheses are probabilities of choosing a particular path.)

the number of other links emanating from the node. The probability of choosing a particular path is simply the reciprocal of the number of "departing" links. These traversal probabilities are indicated by the numbers in parentheses in the example in Figure 2. (f) The subject cycles through this retrieval routine for a fixed number of trials. (Twelve to 14 cycles provide the best fits to data from the present research.)

There are several features of this model that may seem implausible to some readers.[1] First, the model will seem too simple, since it relies on a single, orderly mnemonic structure and a small set of processing rules. Of course, the model is not nearly as simple as it appears in this rendering. We urge the interested reader to review the elaborate development in Anderson's (1976; Anderson & Bower, 1973) lengthy theoretical treatises. For the reader who is still dissatisfied, we would argue that the model is complex enough to account for the major features of the data we have presented thus far. Second, the blind, probabilistic search process embodied in the retrieval rules may not appear "intelligent" enough to some readers. It is important to note the pattern of overt behavior that characterizes the recall phase of these experiments. The subject sits quietly for a few seconds, presumably scanning his or her memory for relevant information, and then spends several seconds writing out the sentence summarizing the information recalled. Then the sequence is repeated with longer and longer pauses between writing. The task of writing interrupts the subject's search processes and forces him or her to start again "from the top." This process accords with the weak evidence from subjects' reports of the experience of performing the recall task. The "clogging" effect, in which a subject is unable to avoid repeatedly recalling previously recalled behaviors, is also consistent with the model and subjects' interior views of the task.

However, the model is not perfectly successful even given the current paltry data base. There are three subtle empirical failures. First, the model predicts a set-size effect for congruent-item recall as well as for incongruent-item recall. There is no consistent congruent set-size effect across the three experiments, nor is there a significant effect within any experiment. Second, the model predicts equal recall of congruent and incongruent items when set sizes are equal. Experiments 1 and 2 showed a small, but reliable, recall advantage for incongruent items in equal set-size lists. Third, the simple network model outlined here predicts that incongruent items will tend to be recalled before congruent items in any single output sequence.[2] An analysis of recall output sequences revealed no consistent output priority for incongruent items in any of the three experiments. Thus, we are not overwhelmingly in favor of the simple network model. Its major theoretical function is to show us just how far we can go toward a complete theoretical analysis if we restrict ourselves to very simple associative encoding and retrieval rules and use set size as the primary explanatory construct.

A Depth-of-Processing Model

An alternative account for our experimental results can be developed from Craik and

---

[1] A preliminary remark is necessary about recall of behaviors from the final recency serial positions of the lists in these experiments. Recall of congruent and incongruent behaviors from the last three or four positions was a direct function of the proportion of each kind of item in these positions. That is, subjects were equally likely to recall either type of item, given that it appeared toward the end of the list. Typically, in 90% to 95% of the cases, subjects started their recall sequence with one of the last three items in the list. Both of the theoretical frameworks to be presented below include an account of these recency effects. However, we do not develop arguments about recency effects here because they are not relevant to our analysis of the *differential* recall of congruent and incongruent behaviors.

[2] Obviously, the model predicts that the first item recalled is equally likely to be from the congruent or incongruent sets. However, after this recall, the smaller set, usually incongruent items in the present research, will be exhausted earlier than the larger set. This implies that incongruent items will be recalled earlier than congruent ones. Of course, if set sizes are equal, there would be no differential priority in the output sequence.

Lockhart's (Craik & Lockhart, 1972; Lockhart, Craik, & Jacoby, 1976) levels-of-processing framework.[3] First, certain events in the series of acts are perceptually striking or distinctive. We assume that perceptual distinctiveness is determined by the degree to which an act is congruent with the general impression of a person at the time when the act occurs (Cantor & Mischel, 1977, make a similar proposal in their discussion of recognition memory for traits). Another expression for this characteristic of an event is its *informativeness*. Informative events are "novel," unexpected, or nonredundant with previous information about a person. Crudely speaking, events early in the consequence of acts are more informative than later events; events that disconfirm the current impression of a person are more informative than events that confirm that impression; and events that are extremely inconsistent or incongruent with a prevailing impression are more informative than events that are consistent or only slightly incongruent with that impression. Of course, this list is merely a summary of variables that have been demonstrated to have large effects on the meanings of items in an initial impression (Asch, 1946; Luchins, 1957) or to determine the weights assigned items in information integration tasks (Anderson, 1974b).

Second, distinctiveness or informativeness determines the depth to which an item is processed. The more distinctive or informative an item, the deeper its processing. In the present experimental task, two types of deep processing are particularly important: impression change and causal attribution. We hypothesize that the adjustment of an individual's impression involves special processing of the specific information that leads to that adjustment. We also hypothesize that the subject is very likely to generate causal explanations for distinctive acts. This attribution processing produces deep, durable memory traces in Craik and Lockhart's terms.

Third, deeply encoded acts are less likely than shallowly encoded acts to undergo interference or decay during the retention interval. Finally, deeply encoded acts are retrieved more easily than other acts during the recall phase of the task. Figure 3 summarizes this analysis for the present experimental task. It is clear that this approach can account for the congruence, set size, and serial position results. Again, the primacy results require a further comment. We assume that any initial items are informative and processed deeply. This means that initial congruent items will be well recalled relative to congruent items from the central portions of the sequence, as was observed.

The Craik and Lockhart account, with its emphasis on encoding or perceptual effects, is in the spirit of current models for impression formation and attribution (Ajzen & Fishbein, 1975; Anderson, 1974a; Jones et al., 1972; Kelley, 1973). Furthermore, the Craik and Lockhart approach suggests some additional data analysis. For example, degree of incongruence should be positively related to probability of recall. If we compute correlations between probability of recall and rated extremity of the behavior on the trait dimension relevant to each list across incongruent behavior items, we consistently obtain positive coefficients, although they are not impressively large (Experiment 1, mean $r = .15$; Experiment 2, mean $r = .21$; Experiment 3, mean $r = .07$; none of these values is significantly different from zero).

We should point out that a hybrid model including components of both network and

---

[3] Recently the conceptual analysis underlying the levels-of-processing framework has been challenged (Baddeley, 1978; Nelson, 1977). We retain the levels-of-processing analysis for two reasons. First, acknowledging that the critics have pointed out aspects of the approach that need further conceptual development, that require additional empirical investigation, or that may be incorrect, we still believe that the framework is a useful theoretical analysis. Additionally, we believe that future work on the framework will address all of the critics' objections and yield a more precise and more powerful theoretical structure. Second, everyone agrees that variations in subjects' encoding behaviors are critical determinants of later recall test performance. The Craik and Lockhart framework is the richest current account for variations in processing *quality* as well as *quantity*. Our speculation is that there are differences in the quality of processing afforded incongruent items as compared to congruent items. However, we admit that the present experiments do not provide convincing support for this inference.

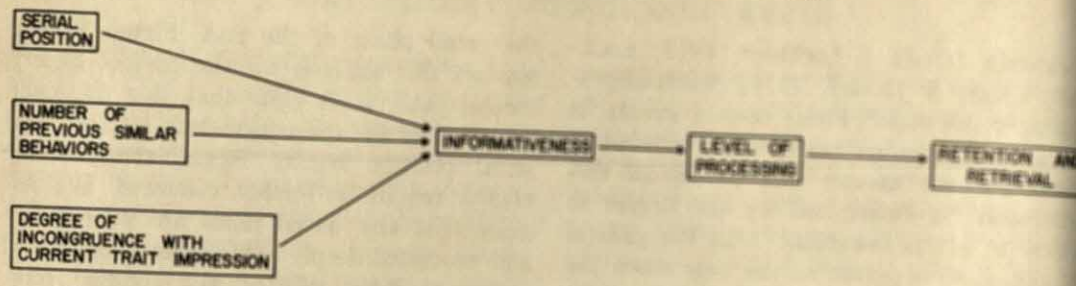


Figure 3. A schematic representation of the variables involved in a depth-of-processing model for the person memory task.

levels-of-processing formulations is easy to imagine. Our preference would be to follow the levels-of-processing analysis of perception and encoding, proposing that deep encoding involves the establishment of more interbehavior links than shallow encoding. Thus, an informative, incongruent act will be linked to more other behaviors than a relatively uninformative congruent act. For instance, we would suppose that causal attribution involves the establishment of a rich network of interitem links between the to-be-explained act and other information about the actor. Retrieval would follow the network model rules of tracing paths through the memory structure and reading out information when new nodes are reached.

Schema Models

A third theoretical account for the present results may be available based on the notion of a schema, frame, or script (Bartlett, 1932; Bobrow & Norman, 1975, Neisser, 1976; Minsky, 1975; Schank & Abelson, 1977). Such an explanation would postulate a mental representation with a central "principle," "terminal slot," or "stereotyped action sequence." Early information (for instance, the trait ensemble in our experiments) would activate or select such a data structure. For the moment, we think that our results present a challenge to schema theorists, because no particular current schema theory appears to provide an account for the pattern of data obtained in these experiments. Of course, we do not doubt that a schema theory can be developed to account for the results of the present research. In fact, Winston (1977) points out that schema models are more powerful than the network models we have outlined. But, at present, no such explicit schema memory model is in the literature.

Conclusion

The present research establishes the significance of abstract trait information in determining the recall of specific behaviors. The effects of congruence or consistency among acts attributed to a person depend on the primacy of the trait impression in perception and memory. There is no way to explain the large, reliable congruence and set-size effects obtained in this research without reference to a standard of consistency such as the abstract trait impression. For the present, it remains to be seen whether a network model, a depth-of-processing model, or some currently unspecified alternative will provide the most satisfactory account of these results. We are also curious as to what extent the neat pattern of results obtained in our laboratory task will generalize to people and settings in the real world.

References

Abelson, R. P., et al. (Eds.). *Theories of cognitive consistency: A sourcebook*. Chicago: Rand McNally, 1968.

Ajzen, I., & Fishbein, M. A Bayesian analysis of attribution processes. *Psychological Bulletin, 1975, 82*, 261–277.

Anderson, J. R. *Language, memory, and thought*. Hillsdale, N.J.: Erlbaum, 1976.

Anderson, J. R., & Bower, G. H. *Human associative memory*. Washington, D.C.: Winston, 1973.

Anderson, J. R., & Hastie, R. Individuation and reference in memory: Proper names and definite descriptions. *Cognitive Psychology, 1974, 6*, 495–515.

Anderson, N. H. Likableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology*, 1968, *9*, 272–279.

Anderson, N. H. Cognitive algebra: Integration theory applied to social attribution. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 7). New York: Academic Press, 1974. (a)

Anderson, N. H. Information integration theory: A brief survey. In D. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 2). San Francisco: Freeman, 1974. (b)

Asch, S. E. Forming impressions of personality. *Journal of Abnormal and Social Psychology*, 1946, *41*, 258–290.

Baddeley, A. D. The trouble with levels: A reexamination of Craik and Lockhart's framework for memory research. *Psychological Review*, 1978, *85*, 139–152.

Bartlett, F. C. *Remembering*. London: Cambridge University Press, 1932.

Bear, G., & Hodun, A. Implicational principles and the cognition of confirmatory, contradictory, incomplete, and irrelevant information. *Journal of Personality and Social Psychology*, 1975, *32*, 594–604.

Bellezza, F. S., & Cheney, T. L. Isolation effect in immediate and delayed recall. *Journal of Experimental Psychology*, 1973, *99*, 55–60.

Bobrow, D. G., & Norman, D. A. Some principles of memory schemata. In D. G. Bobrow & A. Collins (Eds.), *Representation and understanding*. New York: Academic Press, 1975.

Bousfield, A. K., & Bousfield, W. A. Measurement of clustering and of sequential constancies in repeated free recall. *Psychological Reports*, 1966, *19*, 935–942.

Bower, G. H. Organizational factors in memory. *Cognitive Psychology*, 1970, *1*, 18–46.

Bransford, J. D., & Franks, J. J. The abstraction of linguistic ideas: A review. *Cognition*, 1972, *1*, 211–240.

Bransford, J. D., & Johnson, M. K. Consideration of some problems of comprehension. In W. G. Chase (Ed.), *Visual information processing*. New York: Wiley, 1973.

Cantor, N. E., & Mischel, W. Traits as prototypes: Effects on recognition memory. *Journal of Personality and Social Psychology*, 1977, *35*, 38–48.

Craik, F. I. M., & Lockhart, R. S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 1972, *11*, 671–684.

D'Andrade, R. G. Memory and the assessment of behavior. In H. Blalock (Ed.), *Measurement in the social sciences*. Chicago: Aldine, 1974.

Greenwald, A. G., & Sakumura, J. S. Attitude and selective learning: Where are the phenomena of yesteryear? *Journal of Personality and Social Psychology*, 1967, *7*, 387–397.

Hamilton, D. L., & Gifford, R. K. Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgments. *Journal of Experimental Social Psychology*, 1976, *12*, 392–407.

Hastorf, A. H., Schneider, D. J., & Polefka, J. *Person perception*. Reading, Mass.: Addison-Wesley, 1970.

Heider, F. *The psychology of interpersonal relations*. New York: Wiley, 1958.

Insko, C. A., Songer, E., & McGarvey, W. Balance, positivity, and agreement in the Jordan paradigm: A defense of balance theory. *Journal of Experimental Social Psychology*, 1974, *10*, 53–83.

Jones, E. F., et al. *Attribution: Perceiving the causes of behavior*. Morristown, N.J.: General Learning Press, 1972.

Kelley, H. H. The processes of attribution. *American Psychologist*, 1973, *28*, 107–128.

Kintsch, W. *The representation of meaning in memory*. Hillsdale, N.J.: Erlbaum, 1974.

Lewis, C. H., & Anderson, J. R. Interference with real world knowledge. *Cognitive Psychology*, 1976, *8*, 311–335.

Lockhart, R. S., Craik, F. I. M., & Jacoby, L. Depth of processing, recognition and recall. In J. Brown (Ed.), *Recall and recognition*. New York: Wiley, 1976.

Luchins, A. S. Experimental attempts to minimize the impact of first impressions. In C. I. Hovland (Ed.), *The order of presentation in persuasion*. New Haven, Conn.: Yale University Press, 1957.

Mandler, G. Organization and memory. In K. W. Spence & J. A. Spence (Eds.), *The psychology of learning and motivation* (Vol. 1). New York: Academic Press, 1967.

Mandler, J. M., & Johnson, N. S. Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, 1977, *9*, 111–151.

Markus, H. Self-schemata and the processing of information about the self. *Journal of Personality and Social Psychology*, 1977, *35*, 63–78.

McGuire, W. J. The nature of attitudes and attitude change. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (Vol. 3). Reading, Mass.: Addison-Wesley, 1968.

Miller, G. A. Human memory and the storage of information. *IRE Transactions*, 1956, *2*, 129–137.

Minsky, M. A framework for representing knowledge. In P. Winston (Ed.), *The psychology of computer vision*. New York: McGraw-Hill, 1975.

Murdock, B. B., Jr. The serial position effect of free recall. *Journal of Experimental Psychology*, 1962, *64*, 482–488.

Neisser, U. *Cognition and reality*. San Francisco: Freeman, 1976.

Nelson, T. O. Repetition and depth of processing. *Journal of Verbal Learning and Verbal Behavior*, 1977, *16*, 151–171.

Picek, J. S., Sherman, S. J., & Shiffrin, R. M. Cognitive organization and coding of social structures. *Journal of Personality and Social Psychology*, 1975, *31*, 758–768.

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology*, 1977, *35*, 677–688.

Rosch, E. On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language*. New York: Academic Press, 1973.

Rosenberg, S., & Sedlak, A. Structural representations of implicit personality theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6). New York: Academic Press, 1972. (a)

Rosenberg, S. & Sedlak, A. Structural representations of perceived personality trait relationships. In A. K. Romney, R. N. Shepard, & S. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences* (Vol. 2). New York: Seminar Press, 1972. (b)

Schank, R., & Abelson, R. P. *Scripts, plans, goals, and understanding*. Hillsdale, N.J.: Erlbaum, 1977.

Schneider, D. Implicit personality theory: A review. *Psychological Bulletin*, 1973, *79*, 294–319.

Smith, K. H. Effect of exceptions on verbal reconstructive memory. *Journal of Experimental Psychology*, 1973, *97*, 119–139.

Snyder, M., & Uranowitz, S. W. Reconstructing the past: Some cognitive consequences of person perception. *Journal of Personality and Social Psychology*, 1978, *36*, 941–950.

Sulin, R. A., & Dooling, D. J. Intrusion of a thematic idea in retention of prose. *Journal of Experimental Psychology*, 1974, *103*, 255–262.

Thorndyke, P. W. Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, 1977, *9*, 77–110.

Tulving, E. Theoretical issues in free recall. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, N.J.: Prentice-Hall, 1968.

Tulving, E. Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press, 1972.

Winston, P. H. *Artificial intelligence*. Reading, Mass.: Addison, Wesley, 1977.

Zadny, J., & Gerard, H. B. Attributed intentions and information selectivity. *Journal of Experimental Social Psychology*, 1974, *10*, 34–52.

Received August 17, 1977

Notice on Author Alterations

Effective January 1979, the procedure for billing authors for alterations will change. Authors of articles in APA journals are customarily charged for alterations on proofs that result from (a) authors' errors or omissions in the manuscript and (b) changes that result from authors' failure to review the edited manuscript or the proof. Until now, alteration charges have been determined by the number of lines affected by changes requested and have been based on 1974 printing rates. Beginning with the proofs for the first issues of 1979, authors' alterations will be determined not only by the number of lines affected but also by the number of figures and pages affected and will be based on current printers' rates.

This change in alteration billing is being made so that alteration charges will more closely reflect the actual cost of changes requested and may mean larger alteration bills for some authors. Authors are reminded that alteration charges apply only when changes are made on proofs and that *there is no charge* for changes made at the manuscript stage.

Journal of Personality and Social Psychology
1979, Vol. 37, No. 1, 39-47

Effects of Amount of Helplessness Training and Internal–External Locus of Control on Mood and Performance

Nancy L. Pittman
University of Maryland

Thane S. Pittman
Gettysburg College

Based on Wortman and Brehm's integration of reactance theory with Seligman's model of learned helplessness, an investigation was conducted to examine the effects of amount of helplessness training and internal–external locus of control on subsequent task performance and on self-ratings of mood. Subjects were divided into "internal" and "external" groups and were then given either high, low, or no helplessness training on a series of concept-formation problems. After completing a mood checklist, all subjects worked on an anagram task presented as a second experiment by a second experimenter. The results revealed that internals exhibited greater performance decrements and reported greater depression under high helplessness than did externals. In the low helplessness conditions, internals tended to perform better than control subjects, while externals tended to perform worse than control subjects; low helplessness subjects also reported the highest levels of hostility. The results are discussed within the context of Wortman and Brehm's integration of reactance and learned helplessness theories.

Seligman (1972, 1975) has proposed that learned helplessness, a phenomenon originally demonstrated in animal research, may underlie human reactive depression. According to learned helplessness theory, when individuals are exposed to situations in which reinforcements are independent of responses, they gradually develop the expectation that reinforcements cannot be controlled. This expectation of uncontrollability reduces the person's motivation to respond in new situations, interfering with performance on new tasks, and leads to anxiety and a state of depression, because of the perceived lack of control over outcomes.

This article is based on a master's thesis submitted by the first author to the Graduate College of the University of Maryland. The authors thank Mary Faith Tanney, Barry Smith, and Clara Hill, who served on the examining committee, and Susan Adams, who served as the second experimenter.

Requests for reprints should be addressed to Thane S. Pittman, Department of Psychology, Princeton University, Princeton, New Jersey 08540, until July 1, 1979, and to Thane S. Pittman, Department of Psychology, Gettysburg College, Gettysburg, Pennsylvania 17325, after July 1, 1979.

Wortman and Brehm (1975) have proposed an integration of the learned helplessness model with reactance theory. According to reactance theory, individuals who expect to have control (i.e., freedom to obtain an outcome) are motivationally aroused to restore control when it is threatened or reduced. This motivational arousal may be accompanied by hostility and aggression. In integrating the two theories, Wortman and Brehm suggested that mild experiences with helplessness training will result in reactance and improved performance on subsequent tasks in an effort to reassert control, while more extensive experience with uncontrollability will lead to passivity and debilitated performance on new tasks. Several investigators have reported evidence consistent with this hypothesized relationship between amount of helplessness training and subsequent task performance (e.g., Roth & Bootzin, 1974; Roth & Kubal, 1975).

Learned Helplessness and Mood

There is a substantial amount of evidence consistent with the hypothesis that experience

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3701-0039$00.75

with uncontrollable contingencies leads to depression. Miller and Seligman (1973) found that depressives were more likely than nondepressives to view reinforcements as independent from responses. Klein, Fencil-Morse, and Seligman (1976) found that nondepressed subjects given helplessness training on a series of unsolvable problems exhibited impaired performance on a subsequent task and that this impaired performance was similar to the performance of depressed subjects not given helplessness training. Both Gatchel, Paulus, and Maples (1975) and Miller and Seligman (1975) found that subjects subjected to inescapable aversive noise exhibited increases in levels of reported depression, anxiety, and hostility on the Multiple Affect Adjective Checklist compared to their pretest responses. Griffith (1977), using a forced-choice mood scale that paired depression and anxiety adjectives in order to assess the relative balance of depression and anxiety, gave helplessness subjects 10 trials of noncontingent feedback on each of four concept-formation problems. Half of these subjects were also told that they had failed on each problem; the other half were told that they had succeeded on each problem. Results on the mood checklist indicated that noncontingent feedback – failure subjects changed in the direction of increasing depression after treatment, while subjects in the noncontingent feedback – success group changed in the direction of increasing anxiety. Both groups exhibited performance decrements on a posthelplessness solvable anagram task. Finally, Glass and Singer (1972) obtained self-ratings in one investigation that revealed that subjects who perceived no control over an aversive stimulus rated themselves as significantly more helpless, incompetent, and weak than subjects who perceived control over the same stimulus.

While there is considerable support for a relationship between helplessness training and depression, the question of whether different moods are experienced by subjects after different amounts of helplessness training has received less attention.[1] One major purpose of the present study was to assess whether, as Wortman and Brehm (1975) imply, subjects initially respond to uncontrollab by experiencing frustration and/or host but become resigned and depressed as controllability continues.

Learned Helplessness and Locus of Contr

The second major purpose of the pres study was to examine the relationships tween locus of control, amount of helple ness training, and subsequent mood and ta performance. The personality dimension Internal–External (I–E) Locus of Cont (Rotter, 1966) refers to the extent to whi an individual generally perceives reinforc ments as response dependent or controllabl Internal individuals tend to perceive rei forcements to be response dependent, whi external individuals tend to perceive rei forcements as response independent. Give Wortman and Brehm's (1975) suggestion that initial expectations of control affect re sponses to uncontrollability, it might be ex pected that internals and externals would react differently to helplessness training.

Individuals who expect control should be more likely to experience reactance when confronted with uncontrollability than individuals who do not expect control. If internals have greater expectations of control than externals, then internals should also exhibit more pronounced reactance effects when given mild experiences with uncontrollability. It might also be suggested that individuals who expect control will be affected more severely by helplessness training once they realize that they truly have no control. Watson and Baumel (1967) and Petzel and Gynther (1970) have suggested that internals are more uncomfortable with uncontrollability than externals, because of the incongruence between their expectations of controllability and the uncontrollable situation. These predictions were examined in the present

[1] In the important conditions of their study, Roth and Kubal (1975) found general increases in depression, anger, and anxiety as helplessness training increased. Several procedural differences between the Roth and Kubal study and the present investigation are discussed in the final section of this article.

study by exposing groups of internals and externals to one of several levels of helplessness training.

Method

Design

The design was a 2 × 3 between-groups factorial, with three levels of helplessness training on a concept formation task (high helplessness, low helplessness, and no-helplessness control) crossed with two groups of subjects selected on the basis of their extreme internal or external scores on the I–E scale (Rotter, 1966). The proportion of noncontingent incorrect feedback on the concept formation problems was held constant at .5 for all low and high helplessness subjects. A yoked contingent and noncontingent feedback design was not employed, since practice effects could be expected to lead to differential overall proportions of correct and incorrect feedback as a function of the number of problems in the contingent feedback conditions.[2] A review of previous studies that employed similar concept-formation tasks and both contingent feedback (nonyoked) and no-feedback control groups (e.g., Benson & Kennelly, 1976; Griffith, 1977; Hanusa & Schulz, 1977; Hiroto & Seligman, 1975; Klein et al., 1976; Roth & Kubal, 1975; Tenner & Eller, 1977) indicated that contingent feedback either enhanced performance or did not affect performance compared to no-feedback or baseline subjects. A no-feedback procedure was therefore employed in the no-helplessness control conditions of the present study.

Subjects

Ninety students were selected on the basis of their scores on the Rotter I–E scale. Three hundred eighty-five students in introductory psychology were given the I–E scale by their discussion leaders. Subjects with scores above 15 and below 8 on the I–E scale were assigned to the external and internal groups, respectively, and were called and asked to participate in two different learning experiments. The 90 subjects were randomly assigned to one of the three conditions within levels of externality. Each subject was run individually and received two credits for participation in the experiment.

Procedure

A series of five-dimensional stimulus patterns previously used in discrimination learning studies (Levine, 1966) and in other helplessness studies (cf. Hiroto & Seligman, 1975; Klein et al., 1976) were used for the treatment task. Each of the five dimensions had two values: (a) letter (A or T); (b) letter color (black or red); (c) letter size (upper- or lowercase); (d) border surrounding letter (circle or square); and (e) underline (dotted or solid). Two stimulus patterns were shown on each 3 × 5 card. One pattern consisted of a set of values from each of the five dimensions; the other pattern consisted of the complimentary values of the dimensions.

All subjects were given six concept identification problems by the first experimenter, who was seated next to the subject at a table (the first experimenter was unaware of the subject's I–E scale score). Subjects in the helplessness groups were told that for each card, the task was to choose which side of the card contained the "correct" value, arbitrarily set by the experimenter, and that from the experimenter's feedback ("correct" or "incorrect") they should learn the correct answer and choose correctly as often as possible. Subjects in the low helplessness groups were told that for several problems they would be asked to work without feedback, to provide a "baseline." Subjects in the control groups received the same task description as those in the helplessness groups, but were told that the experimenter needed a baseline of choices made without feedback for all of the problems.

All subjects initially received one 10-trial sample problem to clarify the tasks. After the sample problem, all subjects were told that it was very important for the experiment that they work hard, and if they did their best on the task in the experimenter's judgment, they would receive an additional credit. These instructions were given to motivate subjects to attend to the tasks and to try harder. Subjects then received six problems, 10 trials per problem.

High helplessness. Following the sample problem and offer of credit incentive, predetermined noncontingent feedback (example: CICICCICI) was given for each problem; incorrect feedback was always given on the last trial. After each set of 10 trials, subjects were asked to state what they believed was the "correct" value for that set but were given no feedback concerning their answer.[3] They were then told to start the next problem, whose

---

[2] Since the training conditions differed in the absolute amount of feedback, it might be argued that the sheer amount of negative feedback, rather than the noncontingency of feedback, could account for the obtained results. While this interpretation is possible, it is unlikely, since (a) other investigations in which yoked controls were employed have demonstrated helplessness effects (e.g., Glass & Singer, 1972; Hiroto & Seligman, 1975), and (b) noncontingent success has been shown to produce helplessness (Benson & Kennelly, 1976).

[3] Pretesting indicated that subjects became suspicious when they were given incorrect feedback at the end of all six problems, so subjects were simply given no feedback on their final guesses to eliminate suspicion.

value might or might not be the same as that of the previous problem.

Low helplessness. For the first four problems, subjects were asked to choose the correct side and value without any feedback, to provide a baseline for their performance. For the last two problems, contingent feedback was given using the same procedure as in the high helplessness conditions.

Control. Subjects in this condition were asked to guess the correct side for all problems without any feedback, to provide a baseline for guessing for use with other subjects who would receive feedback.

In all conditions, the experimenter recorded all of the subject's responses. No trial was allowed to exceed 15 sec; after 10 sec subjects were told that they had 5 sec left to make their choice.

Mood assessment. The Multiple Affect Adjective Checklist (MAACL) Today Form (Zuckerman & Lubin, 1965) was given to all subjects immediately following the treatments. Subjects were told that their responses would be anonymous and that they were being asked to fill out the checklist because the experimenter wanted to include their current feelings with the results. The MAACL provides measures of three negative affective states: anxiety, depression, and hostility. After reading the standard instructions and completing the MAACL, all subjects received two credits and were thanked for their participation.

Solvable test task. The test task was conducted at a different location by the second experimenter, who was always blind to conditions, immediately following the first experiment. When the subject reported, the second experimenter explained that she was "developing norms for anagram solution times from a cross section of students." The subject's task was to solve a series of anagrams. The experimenter explained that there could be a pattern by which to solve all of the anagrams but that it was up to the subject to figure that out. Subjects were told to give their solution as rapidly as possible, since their performance was being timed. If the subject stated an incorrect solution, the experimenter said, "No, that's not it." If the subject stated the correct word, the experimenter said, "Yes, that's it," recorded the time to solution, and handed the subject the next anagram.

Twenty solvable anagrams were given, each on a 3×5 card. All anagrams had the same letter sequence (e.g., ECUNL–UNCLE) and were obtained from Tresselt and Mayznor (1966). Three measures of anagram performance were obtained: (a) trials to criterion, defined as the solution of three consecutive anagrams in less than 15 sec; (b) number of failures to solve within 100 sec, at which point the trial was ended; and (c) mean response latency for the 20 anagrams.

Following the anagram trials, the first experimenter entered the second experimenter's room, having "forgotten" to give the subject a questionnaire about the first learning experiment. The questionnaire assessed perceptions of task solvability before and after treatment, level of motivation, degree of success on the task, and likelihood that others would be able to solve the concept-formation problems. All subjects were then debriefed by the first experimenter.

Results

I–E Scale Scores

A 2×3 analysis of variance of the I–E scores indicated the expected main effect for the I–E variable, $F(1, 84) = 1,151.44$, $p < .001$, such that subjects in the external conditions had higher scores than subjects in the internal conditions ($M = 18.20$ and 4.20, respectively). Neither the main effect for helplessness nor the two-way interaction was significant, $F(2, 84) = .59$ and 1.16, respectively.

Anagram Performance

A 2×3 analysis of variance of the trials-to-criterion measure indicated a significant main effect for the helplessness variable, $F(2, 84) = 81.97$, $p < .001$, such that subjects in the high helplessness conditions performed worse than subjects in either the low helplessness or the control conditions. The main effect for I–E was not significant. However, the analysis indicated a significant interaction between the helplessness and I–E variables, $F(2, 84) = 14.99$, $p < .001$ (see Table 1). Although internal and external subjects did not differ from each other in the control condition ($F < 1$), they reacted in opposite ways to the low helplessness manipulation. Low helplessness–internal subjects performed significantly better than control–internal subjects, $F(1, 84) = 4.32$, $p < .05$, a reactance effect. Low helplessness–external subjects, however, performed significantly worse than control–external subjects, $F(1, 84) = 5.92$, $p < .05$. In the high helplessness conditions, both internals and externals performed significantly worse than their appropriate controls, $F(1, 84) = 90.41$, $p < .001$, and $F(1, 84) = 40.80$, $p < .001$, respectively; in addition, high helplessness–internal subjects performed significantly worse than high helplessness–external sub-

Table 1
Mean Anagram Performance as a Function of Helplessness Training and Internal–External Classification

Group	High helplessness	Low helplessness	Control
	Trials to criterion		
Internal	16.80	5.66	7.66
External	14.40	10.60	8.26
	No. of failures to solve		
Internal	4.80	.73	1.40
External	3.06	2.13	1.53
	Mean response latency (in sec)		
Internal	51.02	24.84	29.51
External	42.85	37.43	30.93

Note. The number of trials to criterion could range from a minimum of 3 (if the first 3 anagrams were each solved in less than 15 sec) to a maximum of 20 (if criterion was never reached); the number of failures to solve could range from 0 (if every anagram was solved) to 20 (if none of the anagrams were solved); the maximum possible response latency was 100 sec (the point at which a trial was terminated).

jects, $F(1, 84) = 6.23$, $p < .05$. Statistically weaker Helplessness × I–E interactions of similar form were also obtained for the number of failures to solve, $F(2, 84) = 3.75$, $p < .05$, and for mean response latencies, $F(2, 84) = 4.42$, $p < .05$ (see Table 1).

Mood Indexes

Depression. A 2 × 3 analysis of variance of the depression scores from the MAACL yielded a significant main effect for helplessness, $F(2, 84) = 91.86$, $p < .001$, such that subjects in the high helplessness conditions were more depressed than those in the low helplessness and control conditions. The main effect of the I–E variable was also significant, $F(1, 84) = 5.12$, $p < .05$, such that internal subjects reported themselves to be more depressed than external subjects. The interaction between helplessness and I–E was also significant, $F(2, 84) = 6.34$, $p < .01$. While internal and external subjects did not differ from each other in the control and low helplessness conditions (Fs < 1), in the high helplessness condition, internals were significantly more depressed than externals, $F(1, 84) = 17.76$, $p < .001$ (see Table 2).

Hostility. A 2 × 3 analysis of the hostility scores (see Table 2) yielded a significant main effect for helplessness, $F(2, 84) = 43.29$, $p < .001$. There were no other significant effects on this measure. Comparisons of the marginal means indicated that high helplessness subjects were significantly more hostile than controls, $F(1, 84) = 39.07$, $p < .001$ (overall $M = 12.37$ and 6.94, respectively). Low helplessness subjects were significantly more hostile than high helplessness subjects, $F(1, 84) = 8.05$, $p < .01$ (overall $M = 14.83$ and 12.37, respectively).

Anxiety. A 2 × 3 analysis of the anxiety scores (see Table 2) yielded a significant main effect for the I–E variable, $F(1, 84) = 5.32$, $p < .05$, such that internal subjects overall reported more anxiety than external subjects (overall $M = 10.27$ and 9.22, respectively). The main effect for helplessness was also significant, $F(2, 84) = 49.64$, $p < .001$. Since no predictions had been made for anxiety scores, a Newman-Keuls analysis was performed on the marginal means. This analysis indicated that although subjects in the low and high helplessness conditions did not differ, both groups reported significantly greater anxiety than controls ($p < .01$). The

Table 2
Multiple Affect Adjective Checklist Means

Group	High helplessness	Low helplessness	Control
	Depression		
Internal	26.60	12.80	11.93
External	20.86	13.13	12.00
	Hostility		
Internal	12.80	15.73	7.00
External	11.93	13.93	6.87
	Anxiety		
Internal	12.40	11.20	7.20
External	10.80	10.87	6.00

Note. For each mood measure, the higher the score, the greater the mood. Mood scores could range from a minimum of 0 to a maximum of 40 for depression, 28 for hostility, and 21 for anxiety.

interaction on this measure was not significant.

Questionnaire Responses

A 2 × 3 analysis of the question that asked subjects how hard they had worked on the concept-formation task indicated that there were no significant main or interaction effects (all Fs < 1). All groups reported that they had worked hard on the task (on the 7-point scale, where 1 was labeled "extremely hard," the means ranged from 1.93 to 2.47).

The second question dealt with subjects' perceptions of their performance on the concept-formation task. A 2 × 3 analysis of variance of these ratings indicated that there was a significant main effect for helplessness, $F(2, 84) = 51.14$, $p < .001$. There were no significant effects for the I–E variable or for the two-way interaction. Analysis of the marginal means indicated that high helplessness subjects rated themselves as less successful than low helplessness subjects and that the low helplessness subjects rated themselves as less successful than control subjects ($p < .01$, Newman-Keuls).

There were no significant main effects or interactions on the question that assessed whether subjects perceived the concept-formation task to be solvable at the beginning of the task (all Fs < 1), and all groups rated the task as solvable. The fourth question assessed subjects' perceptions of their ability to solve the task after they had worked on it. A 2 × 3 analysis of variance indicated that only the main effect for helplessness was significant on this measure, $F(2, 84) = 24.97$, $p < .001$. A Newman-Keuls analysis of the marginal means indicated that low helplessness subjects believed that they were less able to solve the task than did control subjects ($p < .01$), and high helplessness subjects believed that they were significantly less able to solve the problems than did low helplessness subjects ($p < .01$).

While the fourth question concerned subjects' perception of their own ability to solve the problems, the fifth question asked subjects to what extent they believed that the problems could be solved by someone else. In order to examine the relative decline in belief in one's own and another's ability to solve, a 2 × 2 × 3 analysis of variance with type of question as the repeated factor was performed. In addition to main effects for the type of question and the amount of helplessness training, an interaction between these two variables was obtained, $F(2, 84) = 13.56$, $p < .001$. Control subjects believed that both they and others could solve the problems ($M = 6.30$ for self and 6.23 for others on the 7-point scale), but both low and high helplessness subjects reported that they believed they were less able to solve than others (in low helplessness, $M = 4.13$ for self and 5.67 for others; in high helplessness, $M = 2.70$ for self and 4.43 for others). Although these data should be interpreted with caution, since they were collected at the very end of the experimental sessions, they seem to indicate that subjects in the helplessness conditions were attributing their inability to solve to themselves. High helplessness subjects in particular believed that they could not solve the problems, but tended to believe that others could.

Discussion

Overall Mood Patterns

The mood results were generally consistent with Wortman and Brehm's (1975) prediction that individuals will initially respond to uncontrollability by experiencing reactance but will eventually exhibit helplessness after continued experience with lack of control. In the present study, low helplessness subjects were significantly more hostile than either high helplessness or control subjects, while high helplessness subjects were significantly more depressed than either low helplessness or control subjects. These results are consistent with Wortman and Brehm's model, but they are somewhat discrepant with Roth and Kubal's (1975) finding that increases in helplessness training led to increases in depression *and* anger when the training task was important. There are several possible reasons for this apparent discrepancy. While Roth and Kubal used a

series of 7-point Likert-type scales to measure affect at the end of the second (test) task, in the present investigation moods were measured with the MAACL immediately following the training task. In addition to the possibility that the measuring instruments used in the two studies may have tapped different mood states, subjects' ratings of their moods immediately after helplessness training may differ from their subsequent retrospective reports. Finally, since moderate internals and externals were excluded from the present study but were presumably included in Roth and Kubal's investigation, it is possible that moderates on the I–E scale react differently than their more extreme counterparts.

The mood results also indicated that both high and low helplessness pretreatments produced anxiety, consistent with the findings of Gatchel et al., (1975), Miller and Seligman (1975), and Roth and Kubal (1975). The anxiety data do not, however, provide support for the alternative hypothesis that impaired performance following helplessness training is due to anxiety (Hebb, 1966, as cited in Wortman & Brehm, 1975). High and low helplessness subjects did not differ in reported anxiety but exhibited quite different levels of performance on the anagram test task. Furthermore, anxiety and performance were not significantly correlated (see Table 3).[4]

Locus of Control, Mood, and Performance

For internals, the pattern of performance results was also compatible with Wortman and Brehm's prediction that individuals who initially expect control will attempt to reassert control after mild experiences of uncontrollability but will eventually show helplessness effects when the uncontrollability is extensive. Low helplessness internals exhibited improved performance on the anagram task relative to controls, while high helplessness internals displayed the greatest performance decrements in the experiment. Under low helplessness, internals' hostility was also negatively correlated with failures, such that the more hostile they were, the more anagrams they solved (see Table 3). Under high helplessness, internals' depression, rather than hostility, was most highly correlated with number of failures. For these subjects, the higher their reported depression, the fewer anagrams they solved. That hostility was related to enhanced performance under low helplessness for internals was a finding particularly supportive of Wortman and Brehm's (1975) model.

Table 3
Within-Cell Correlations Between Mood Indexes and Number of Failures to Solve on the Anagram Task

Group	High helplessness	Low helplessness	Control
	Depression vs. no. of failures to solve		
Internal	.74**	.33	.71**
External	.75**	.41	.43
	Hostility vs. no. of failures to solve		
Internal	.14	−.52*	.12
External	−.41	−.43	−.43
	Anxiety vs. no. of failures to solve		
Internal	.39	.34	−.36
External	−.50	.00	−.13

* $p < .05$.
** $p < .01$.

Externals, however, showed a different pattern of performance. As the amount of uncontrollability increased for these subjects, their performance became increasingly worse. This finding is consistent with Wortman and Brehm's notion that individuals who do not initially expect control will not experience reactance and will show helplessness effects after mild experience with uncontrollability. However, externals were unexpectedly similar to internals in their pattern of moods. As would be expected under high helplessness, externals showed substantial depression (although less than internals), and the correla-

[4] Within-cell correlations between the mood indexes and the number of failures to solve are displayed in Table 3. The correlations between moods and the other measures of performance followed a similar pattern but were generally weaker.

tion between externals' depression and number of failures to solve was similar to that of internals (see Table 3). However, under low helplessness, externals were also similar to internals in that they were more hostile than control subjects; but unlike the internals, externals performed worse on the anagram task than controls. Also in contrast to the relationship for internals, the correlation between hostility and failures for externals was unchanged in all three helplessness conditions (see Table 3). Therefore, while mild experience with uncontrollability led externals to become hostile, this treatment did not appear to alter the relationship between hostility and performance.

That both internal and external subjects were hostile after mild experience with uncontrollability could mean that both groups experienced reactance but behaved differently in response to reactance. This interpretation implies that some aspect of the I–E variable other than initial expectations of control may have led to the different behavior patterns that were obtained. For example, internals and externals may differ in their characteristic response to failure to exert effective control, such that internals initially tend to exert more effort whereas externals rapidly give up. This could explain why both internal and external low helplessness subjects expressed hostility, but only internals displayed enhanced anagram task performance. Cohen, Rothbart, and Phillips's (1976) finding that the effects of learned helplessness tend to generalize more easily for externals is also consistent with this interpretation.

Overall, the findings of the present investigation for internals were consistent with Wortman and Brehm's (1975) predicted sequence of both moods and performance following experience with uncontrollability. The performance of externals was also compatible with Wortman and Brehm's model; however, the increased hostility of externals in the low helplessness condition may not be consistent with an expectation interpretation and indicates that it would be useful to determine the actual critical difference between internals and externals that accounts for their different responses to uncontrollability. In the present investigation only total I–E scores from a mas testing were available for analysis; it woul be interesting in future studies to focus o specific factors of the I–E scale. Investigator have found the I–E scale to contain from tw (e.g., Mirels, 1970) to as many as five fac tors (see Phares, 1976, for a review of these studies). Since, as Phares has observed, the ability to derive reliable separate factors from the original scale is limited because of the small number of items (23), it might be useful to employ either an expanded version of the scale or subscales specifically designed to tap expectations of control versus habitual reactions to uncontrollability. The effects of expectations of control on reactions to helplessness training should also be directly studied by employing unambiguous manipulations of expectations instead of individual difference variables.

References

Benson, J. S., & Kennelly, K. J. Learned helplessness: The result of uncontrollable reinforcements or uncontrollable aversive stimuli? *Journal of Personality and Social Psychology*, 1976, *34*, 138–145.

Cohen, S., Rothbart, M., & Phillips, S. Locus of control and generaliyt of learned helplessness in humans. *Journal of Personality and Social Psychology*, 1976, *34*, 1049–1056.

Gatchel, R. J., Paulus, P. B., & Maples, C. W. Learned helplessness and self-reported affect. *Journal of Abnormal Psychology*, 1975, *84*, 732–734.

Glass, D. C., & Singer, J. E. *Urban stress: Experiments on noise and social stressors*. New York: Academic Press, 1972.

Griffith, M. Effects of noncontingent success and failure on mood and performance. *Journal of Personality*, 1977, *45*, 442–457.

Hanusa, B. H., & Schulz, R. Attributional mediators of learned helplessness. *Journal of Personality and Social Psychology*, 1977, *35*, 602–611.

Hebb, D. O. *A textbook of psychology*. Philadelphia, Pa.: Saunders, 1966.

Hiroto, D. S., & Seligman, M. E. P. Generality of learned helplessness in man. *Journal of Personality and Social Psychology*, 1975, *31*, 311–327.

Klein, D. C., Fencil-Morse, E., & Seligman, M. E. P. Learned helplessness, depression, and the attribution of failure. *Journal of Personality and Social Psychology*, 1976, *33*, 508–516.

Levine, M. Hypothesis behavior by humans during discrimination learning. *Journal of Experimental Psychology*, 1966, *71*, 331–338.

Miller, W. R., & Seligman, M. E. P. Depression and the perception of reinforcement. *Journal of Abnormal Psychology*, 1973, *82*, 62–73.

Miller, W. R., & Seligman, M. E. P. Depression and learned helplessness in man. *Journal of Abnormal Psychology*, 1975, *84*, 228–238.

Mirels, H. L. Dimensions of internal versus external control. *Journal of Consulting and Clinical Psychology*, 1970, *34*, 226–228.

Petzel, T. P., & Gynther, M. D. Effects of internal–external locus of control and skill or chance instructional sets on task performance. *Journal of General Psychology*, 1970, *82*, 87–93.

Phares, E. J. *Locus of control in personality*. Morristown, N.J.: General Learning Press, 1976.

Roth, S., & Bootzin, R. R. The effects of experimentally induced expectancies of external control: An investigation of learned helplessness. *Journal of Personality and Social Psychology*, 1974, *29*, 253–264.

Roth, S., & Kubal, L. The effects of noncontingent reinforcement on tasks of differing importance: Facilitation and learned helplessness. *Journal of Personality and Social Psychology*, 1975, *32*, 680–691.

Rotter, J. B. Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs*, 1966, *80*(1, Whole No. 609).

Seligman, M. E. P. Learned helplessness. *Annual Review of Medicine*, 1972, *23*, 407–412.

Seligman, M. E. P. *Helplessness: On depression, development, and death*. San Francisco: Freeman, 1975.

Tenner, H., & Eller, S. J. Attributional components of learned helplessness and facilitation. *Journal of Personality and Social Psychology*, 1977, *35*, 265–271.

Tresselt, M. E., & Mayznor, M. S. Normative solution times for a sample of 134 solution words and 378 associated anagrams. *Psychonomic Monograph Supplements*, 1966, *1*, 293–294.

Watson, D., & Baumel, E. Effects of locus of control and expectation of future control upon present performances. *Journal of Personality and Social Psychology*, 1967, *6*, 212–215.

Wortman, C. B., & Brehm, J. W. Responses to uncontrollable outcomes: An integration of reactance theory and the learned helplessness model. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 8). New York: Academic Press, 1975.

Zuckerman, M., & Lubin, B. *Manual for the multiple affect adjective checklist*. San Diego, Calif.: Educational & Industrial Testing Service, 1965.

Received October 31, 1977 ■

Correction to Stephan and Rosenfield

In the article, "Effects of Desegregation on Racial Attitudes," by Walter G. Stephan and David Rosenfield (*Journal of Personality and Social Psychology*, Vol. 36, No. 8, pp. 795–804), the following corrections should be made in Table 2 on page 800: (a) The correlation between parental authoritarian child-rearing practices and changes in racial attitudes should be −.04, not .04. (b) The correlation between parental opposition to integration and increases in self-esteem should be .16, not −.16.

Journal of Personality and Social Psychology
1979, Vol. 37, No. 1, 48-74

Source Credibility in Social Judgment: Bias, Expertise, and the Judge's Point of View

Michael H. Birnbaum and Steven E. Stegner
University of Illinois at Urbana-Champaign

Mathematical models of source credibility were tested in five experiments in which judges estimated the value of hypothetical used cars based on blue book value and/or estimates provided by sources who examined the cars. The sources varied in mechanical expertise and in bias; they were described as friends of the buyer or seller of the car or as neutral. Individuals judged the highest price the buyer should pay, the lowest price the seller should accept, and the "true" value ("fair" price) of the car. Data indicated that expertise amplifies the effect of the source's bias. This effect is predicted by a scale-adjustment model, in which the source's bias shifts the scale value of the source's estimate. The weight of an estimate depends chiefly on the source's expertise. The weight of an estimate also depends configurally on the other estimates: Judges instructed to take the buyer's point of view give greater weight to the lower estimate, whereas judges who identify with the seller place greater weight on the high estimate. Simple premises about human judgment give a good account of the data.

Social judgments often require the combination of pieces of information provided by sources who vary in credibility (Hovland, Janis, & Kelley, 1953; McGuire, 1968). Rosenbaum and Levin (1968, 1969), Anderson (1971), Birnbaum, Wong, and Wong (1976), and Birnbaum (1976) have proposed and/or tested formal theories of source credibility. The present research extends these developments and argues that the concept of *credibility*, used loosely in early persuasion research to mean "believability," can be profitably decomposed into at least three constructs: expertise, bias, and the judge's point of view.

The *judge* is the person (the subject in the present experiments) who combines information provided by one or more *sources* to make an overall evaluation or judgment. The juror, who decides guilt based on contradictory evidence, the voter, who chooses among candidates who disagree, and the consumer, who evaluates the worth of a product, are examples of people acting as judges.

The *expertise* of the source refers to the perceived correlation between the source's report and the outcomes of empirical verification. Expertise would be expected to depend upon such factors as training, experience, and ability. For example, a doctor whose diagnoses are often confirmed in the post mortem would be considered a more expert source of information about the state of a person's health than would an untrained student.

The *bias* of the source refers to factors that are perceived to influence the expected algebraic difference between the source's report and the true state of nature. For example, a Republican might be considered a biased source of information about a Democrat who is running for office.

The distinction between expertise and bias is like the distinction between regression slope and intercept. For example, a used car salesman might be an expert source of information about the value of his cars; however, the salesman's estimates may be biased upward, since the seller stands to profit by convincing potential

This research was supported in part by a grant from the Research Board, University of Illinois. We are grateful to Marcy Wilhite and Michael Less for assistance and to Samuel Komorita, Barbara Mellers, Clairice Veit, and Robert Wyer for helpful comments on the manuscript.

Requests for reprints should be sent to Michael H. Birnbaum, Department of Psychology, University of Illinois, Champaign, Illinois 61820.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3701-0048$00.75

buyers of the high worth of the car. Similarly, an insurance claims adjuster is an expert source who might underestimate the value of damaged goods on which his or her company is required to pay claims.

The judge's bias is termed the judge's "point of view." Thus, a judge who is a Republican or a Democrat may treat information provided by a Republican source and a Democratic source differently. It is important to maintain the distinction between the bias of the sources and the point of view (bias) of the judge who combines the information from the sources.

Previous Research on Source Expertise

Rosenbaum and Levin (1968, 1969) extended an averaging model of impression formation to account for source credibility effects. Anderson (1971) discussed additive and averaging models of source credibility and theorized, on the basis of previous research (on the set-size effect in impression formation), that averaging models would prove superior. Wyer (1974) revived an additive model for source credibility as a "case against averaging." Lichtenstein, Earle, and Slovic (1975) offered a constant-weight averaging model of numerical cue prediction (each cue can be thought of as a source). Additive and averaging models both predict that the effect of a source's report should increase with increased credibility, but the models make different predictions for the effect of the credibility of one source on the effect of information provided by another source.

Birnbaum et al. (1976) and Birnbaum (1976) conducted three experiments testing three models for the effect of source expertise in information integration. In one experiment (Birnbaum et al., 1976, Experiment 1) undergraduates judged the value of used cars based on two cues: blue book value and an estimate provided by one of three sources who examined the car. The three sources differed in mechanical expertise. For example, how much is a car worth if its blue book value is $500 but an expert mechanic who has examined the car evaluates its worth at $700? In the second experiment, judges rated the likableness of hypothetical persons described by personality-trait adjectives attributed to sources who varied in their length of acquaintance with the person they described. For example, how much would you like a person who was described by an acquaintance of 3 years as *sincere* and by an acquaintance of 3 weeks as *phony?* In a third experiment (Birnbaum, 1976), students were trained with feedback to predict a numerical criterion from single independent cues separately, then asked to predict (without feedback) the criterion from pairs of cues.

The results of the used car, impression formation, and numerical prediction studies form a coherent picture. The effect of a cue varies inversely with the number of cues presented, contrary to the additive models. Furthermore, the effect of a cue varies inversely with the credibility of the other cues, in violation of the constant-weight averaging model (including the models of Rosenbaum & Levin, 1968, 1969, and Lichtenstein et al., 1975, as special cases). The results of Birnbaum et al. (1976) and Birnbaum (1976) were qualitatively consistent with a relative-weight averaging model, in which the effect of information provided by one source is inversely related to the number and credibility of the other sources.

Since the results of the three quite different experiments were consistent with the same general model, it is inductively appealing to theorize that the model will hold across different judgmental domains. Values of the estimates, adjective likableness values, and numerical values of the cues are represented by scale values. Levels of mechanical expertise, length of acquaintance, and cue-criterion validity are represented by changes in weight.

The relative-weight averaging model for used car judgment can be written as follows:

$$R = \frac{w_0 s_0 + w_V s_V + w s_E}{w_0 + w_V + w}, \qquad (1)$$

where R is the judged worth; w_0, w_V, and w are the weight of the initial impression, the blue book, and the source, respectively; and s_0, s_V, and s_E are the scale values of the initial impression (the presumed response in the absence of information), the blue book value, and the source's estimate, respectively.

Purposes of the Present Research

The present research investigates the effects of the source's *bias* and the judge's *point of view* on the information integration process. Judges made evaluations of hypothetical used cars based in part on estimates provided by sources who varied in bias and expertise. Bias was manipulated by stating that the source was either a friend of the buyer or the seller or an independent. The judge's point of view was manipulated by asking the judge to identify with either the buyer or the seller of the car. For example, subjects were asked to judge the *most they would advise the buyer to pay* for a car, given that the blue book value is \$500 and an *expert* mechanic, who is a *friend of the seller*, estimates its value at \$700. The bias of a source may affect either the weight or the scale value of the information provided by the source. Furthermore, the effects of the source's bias may depend on the judge's point of view. The next section shows that certain experimental designs make it possible to distinguish different theories of the effect of source characteristics on weight and scale value.

The first four experiments compare three models of the effects of the source's bias. These relative-weight averaging models are compatible with previous research; however, they make strikingly different predictions for the interaction of the source's expertise and bias. The models can be distinguished by qualitative comparisons that do not require metric assumptions about the dependent variable or global tests of goodness of fit.

The fifth experiment tests a theory of configural effects. Deviations from the relative-weight averaging model obtained in previous research (Birnbaum, 1974; Birnbaum et al., 1976) and in the first four experiments of this article are presumed to depend on the stimulus configuration. Implications of a configural-weight theory (Birnbaum, 1974) are explored in the fifth experiment.

Weight and Scale Value

Figure 1 shows how effects of the source on scale value and weight can be separated and analyzed in a relative-weight model. The three examples show hypothetical predictions assuming that the source affects weight only (A1 & B1), scale value only (A2 & B2), or both (A3 & B3).[1] Panels A1, A2, and A3 plot

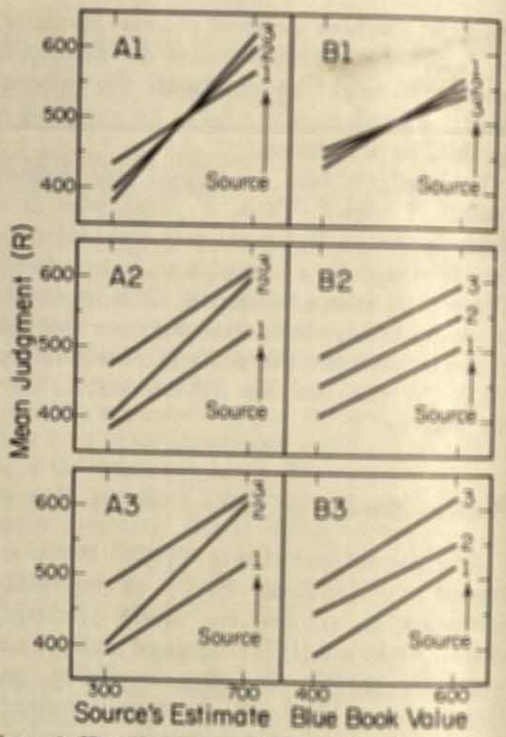


Figure 1. Hypothetical results assuming that sources affect weight only (Panels A1 & B1), scale value of the source's estimate only (A2 & B2), or both weight and scale value (A3 & B3). (Panels A1, A2, and A3 show mean judgments, averaged over blue book value, as a function of the source's estimate with a separate curve for each source. Panels B1, B2, and B3 show mean judgments, averaged over source's estimate, as a function of blue book value with a separate curve for each source. Note that changes in either weight or scale value of the source's estimate can affect the slopes in the A panels. Differences in slope in the B panels are not affected by changes in the scale values of the source's estimates, but depend upon the source's weights.)

[1] In all three examples in Figure 1, scale values for the blue book value were \$400 and \$600, with a weight of 2; the scale values for estimates of Source 2 were always \$300 and \$700, with a weight of 2; the weight of the initial impression was 0. In the first example (A1 & B1), the weights of Sources 1, 2, and 3 were 1, 2, and 3, respectively, and scale values remained fixed at \$300 and \$700. In the second example (A2 & B2), the weights of all three sources were 2; the scale values for Source 1 were \$450 and \$725, for Source 2 they were \$300 and \$700, and for Source 3 they were \$275 and \$550. For the third example, the weights of Sources 1 and 3 were 1, and the weight of Source 2

judged value as a function of the source's estimate with a separate curve for each source. The effect of the source's estimate, ΔR_E, that is, the change in response due to the source's estimate (averaged over blue book value), from Equation 1, will be given by the expression:

$$\Delta R_E = \frac{w}{w_0 + w_V + w} \cdot \Delta s_E, \qquad (1a)$$

where Δs_E is the range of scale values of the source's estimates. Note that the slopes in the A panels of Figure 1 will depend on both weight (w) and the range of scale values, Δs_E. The slopes are proportional to ΔR_E because the abscissa values are constant for all sources; Δs_E and w may or may not vary for different sources.

Panels B1, B2, and B3 of Figure 1 show the response plotted against the blue book value, averaged over the source's estimate, with a separate curve for each source. For the B panels, the effect of blue book value, ΔR_V, is given by the following equation:

$$\Delta R_V = \frac{w_V}{w_0 + w_V + w} \Delta s_V, \qquad (1b)$$

where Δs_V is the range of scale values of the blue book values. Note that if w_0, w_V, and Δs_V are presumed to be independent of the source, the effect of the blue book value (slope) should vary inversely with the weight of the source, w.

Change of weight only. The pattern of results in Figure 1, Panels A1 and B1, is consistent with the results of Birnbaum et al. (1976), who varied expertise (but not bias) of the source. For that experiment, Sources 1, 2, and 3 would represent low, medium, and high expertise, respectively. It was concluded that expertise affects weight, not the range of scale values, since the effects of blue book value (slopes in Panel B1) are lower for sources of higher expertise, as predicted by Equation 1b if expertise affects weight.

Change of scale value only. It seems unlikely that manipulation of the bias of a source, holding expertise constant, would affect only weight and produce the pattern of Panels A1 and B1. If the source's bias affects the scale values of the source's estimates, one would expect main effects of bias and possibly interactions between the bias of a source and the source's estimate. Figure 1, Panel A2, illustrates such a possibility. In this case, Sources 1, 2, and 3 might represent friend of the seller, independent, and friend of the buyer, respectively. At first, the increased slope for Source 2 in Panel A2 might be thought to indicate that Source 2 has greater weight. However, Panel B2 shows the weights are equal, since the curves are parallel. Parallelism implies that the effect of blue book value is independent of the source, indicating (by Equation 1b) that the sources have equal weights.

Change of weight and scale value. Panels A3 and B3 of Figure 1 indicate a pattern in which the source affects both weight and scale value. The curve with the steeper slope in Panel A3 also has the flatter slope in Panel B3.

These examples illustrate that in order to separate the influences of the source on weight and scale value, one must examine not only the effect of a source's estimate (A panels) but also the effect of another cue such as the blue book value (B panels). The effect of the source's *estimate* depends on *both* the source's weight and the range of scale values (Equation 1a and A panels). The effect of blue book value, however, provides an unambiguous constraint on the source's weight (Equation 1b and B panels). By comparing Panel A with Panel B one can tease apart the effects of source variables such as bias on weight and scale value.

Experiments 1–4: Three Models of Source Bias

Figure 2 shows three models of source bias (on the left) and an important prediction of each (on the right). All of the models are relative-weight averaging formulations and can be represented by lever and fulcrum models. The lever is an analog computer that can be used to make predictions for the three models. In each case, the scale value of the source's estimate is represented by the *location* along the lever where the weight is placed,

was 2; the scale values for Source 1 were \$450 and \$850, for Source 2 they were \$300 and \$700, and for Source 3 they were \$150 and \$550.

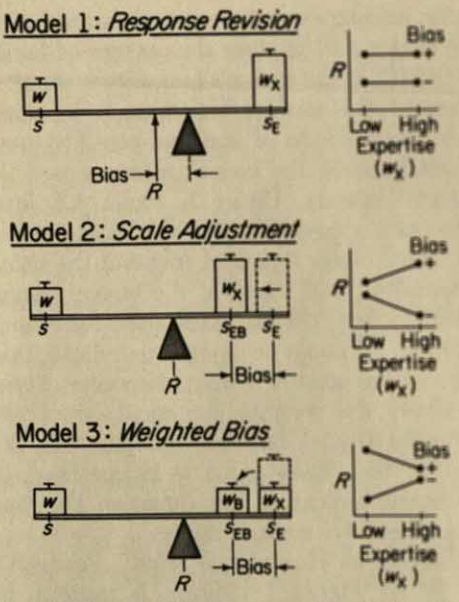


Figure 2. Three models of source bias. (Scale value, s, corresponds to a point along the lever; mathematical weight, w, corresponds to physical weight placed on the lever. Apart from the effects of bias, the response [R] would be the center of gravity, fulcrum balance. Model 1 predicts no interaction between bias and expertise. Model 2 assumes that bias affects the scale value of the source's estimate and predicts that the effect of bias increases with the increasing expertise of the source. Model 3 assumes that the source's bias and estimate are two pieces of information that are averaged to form an overall assessment. Since the relative weight of bias decreases as expertise increases, Model 3 predicts that the effect of bias diminishes with increasing expertise.)

weight corresponds to *weight*, and the response is represented by the *center of gravity* (location of the fulcrum at equilibrium).

In each model, the scale value and weight of the other information (e.g., blue book value) are s and w, the scale value of the source's estimate is represented by s_E, and the weight of the source's expertise is w_X. To represent the initial impression, the plank alone is presumed to have a weight of w_0 with a center of gravity at s_0. The change in response due to bias is shown in each figure by a comparison of solid and dashed symbols (weights or arrows). The direction of bias illustrated in the figure is negative, consistent with the source being a friend of the seller.

Model 1: Response Revision

According to the first model, bias produces a shift in the response, rather than affecting weight or scale value. As shown in Figure 2, this can be represented by the lever if the response is assumed to be the weighted average of a bias-free response, R^*, and a response effect of bias, b_B. A special case of this model can be written as follows:

$$R = \left(\frac{w_{R^*}}{w_{R^*} + w_B}\right)R^* + \left(\frac{w_B}{w_{R^*} + w_B}\right)b_B, \quad (2)$$

where w_{R^*} and w_B are the weights of the bias-free response and the bias adjustment, respectively, which are assumed to be constant, and R^* (fulcrum, or balance point not considering bias) is given by Equation 1:

$$R^* = (w_0 s_0 + w_S + w_X s_E)/(w_0 + w + w_X).$$

This model predicts no interaction between bias and any of the other factors. In particular, the effect of bias is predicted to be independent of expertise, as shown in the upper right section of Figure 2.

A more complicated version of Model 1 would allow the source's weight (w_X) to depend on both expertise and bias. The changes in weight due to bias and expertise can be estimated from the effects of blue book value (as in Figure 1). This more general model still predicts that once the weights have been estimated in this fashion, the residual effect of bias will be independent of expertise. The response revision model represents the judgment as a two-stage process in which the average value is first computed without paying attention to bias, then bias is averaged together with this implicit response.

Model 2: Scale Adjustment

The second model assumes that the bias of the source causes a shift in the value of the information provided by the source. Thus, the scale value of information from a biased source is adjusted "prior" to the integration process. Figure 2 depicts this process by showing that the source's weight is *not* placed at the scale value of the estimate, s_E, but instead at a bias-corrected value, s_{EB}, which depends on both bias and estimate. A special case of this model, which assumes that $s_{EB} = s_E + b_B$, can be written:

$$R = \frac{w_0 s_0 + w_S + w_X(s_E + b_B)}{w_0 + w + w_X}. \quad (3)$$

This scale-adjustment model predicts that the effect of bias will be greater for sources of high expertise than for sources of low expertise, since expertise (w_X) multiplies the bias correction (b_B).

A more complicated version of this model allows the weight of the source to depend on both expertise and bias and allows the corrected scale value to depend on both estimate and bias. This general version of Model 2 can be written as follows:

$$R = \frac{w_0 s_0 + w s + w_{XB} s_{EB}}{w_0 + w + w_{XB}}, \quad (4)$$

where the subscripts of w_{XB} and s_{EB} suggest that the weight of a source depends on both expertise (X) and bias (B), and scale value depends on both estimate (E) and bias (B). The more general model does not require estimate and bias to combine additively to produce scale value.

Model 3: Weighted Bias

The third general theory assumes that the source's estimate and the source's bias are both pieces of information that must be integrated to form an overall evaluation. The scale value of the source's estimate, s_E, receives a weight that depends on the source's expertise, w_X. However, the scale value of the source's bias, b_B, has a weight that depends only on bias, w_B. Thus, as the expertise increases, the source's estimate receives greater relative weight, and the source's bias receives reduced relative weight. Model 3 can be written:

$$R = \frac{w_0 s_0 + w s + w_X s_E + w_B b_B}{w_0 + w + w_X + w_B}. \quad (5)$$

Model 3 predicts that the effect of bias will be *inversely* related to the expertise of the source: As w_X increases, the relative weight of bias, $w_B/(w_0 + w + w_X + w_B)$, decreases. Model 3 can also be extended to allow the weight of the source's estimate to depend on bias, w_{XB}, or to allow the bias correction to depend on estimate and bias. Model 3 asserts that as a source grows in expertise, the judge places more weight on what the source *says* and therefore less relative weight on the correction for bias.

In sum, the three models predict that increasing expertise either increases the effect of bias (Model 2), decreases the effect of bias (Model 3), or does not interact with bias (Model 1). These differential predictions, shown on the right in Figure 2, are tested in the first four experiments.

Method

Instructions

The task was to judge the values of hypothetical used cars based on *blue book value* and/or *estimates* of value provided by sources who varied in *bias* and *expertise*. The sources of the estimates were described as people who attempted to judge the "true" value of the cars, based on a 30-minute inspection and test drive. Their relationship with the buyer or seller and their expertise in judging the value of automobiles were specified.

Source expertise. Separate paragraphs discussed the training and mechanical skill of the sources, who were described as low, medium, or high in expertise. The low-expertise source was described as a competent person who drives a car regularly and has purchased cars for himself. The medium-expertise source was a person who has taken some classes in auto shop and can make some repairs himself. The high-expertise source was described as an expert mechanic whose hobby is the repair and modification of sports cars.

Source bias. Each source was described as a friend of the buyer, a friend of the seller, or an independent. It was explained that the buyer's friend would be expected to be sensitive to his friend's desire to get a good value for his or her money and might emphasize the car's bad points. The independent was characterized as a neutral person with no relationship with the buyer or seller and no reason to under- or overestimate. The seller's friend would be concerned with the seller's desire to get as much money as possible for the car. He might be optimistic about the car and emphasize its good points.

Blue book value. The blue book value was described as a standard "fair" price that is determined by such factors as year, model, make, and mileage. It was remarked that blue book value is widely relied upon by businesses that deal in large numbers of cars, but that it would not describe individual cars.

Procedure and Designs

Each test booklet contained three pages of instructions, 20 warm-up trials, and 293 randomly ordered test trials. The test trials were constructed from the following designs:

Source estimate. A $3 \times 3 \times 5$ (Bias $\times$ Expertise $\times$ Estimate) factorial design generated 45 trials on which the only information presented was the estimate of a source whose expertise and bias were specified. The three levels of the source's bias were friend of the buyer, friend of the seller, and independent. The three

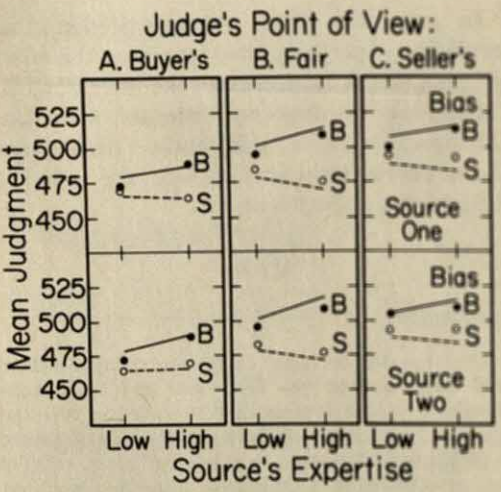


Figure 3. Mean judgment of value as a function of the source's expertise with a separate curve for each level of the source's bias. (Points are empirical means, and lines are predictions based on the scale-adjustment model, Equation 4. Solid points and lines are for friends of the buyer [B]; open circles and dashed lines are for friends of the seller [S]. Upper and lower rows of panels are for the first and second sources, respectively. Panels A, B, and C represent different points of view for the judge, Experiments 1–4.)

levels of source expertise were high, medium, and low. The five levels of estimate were \$300, \$400, \$500, \$600, and \$700.

Source estimate and blue book value. The entire $3 \times 3 \times 5$ source-estimate design was factorially combined with the four levels of blue book value: \$350, \$450, \$550, and \$650. The resulting $(3 \times 3 \times 5) \times 4$, (Bias × Expertise × Estimate) × Blue Book Value, design yielded 180 trials on which the information consisted of a source's estimate and a blue book value.

Two source estimates. Two separate $2 \times 2 \times 2$ source-estimate designs were factorially combined to generate 64 trials ($[2 \times 2 \times 2] \times [2 \times 2 \times 2]$) with estimates from two sources. For both sources, the two levels of bias were friend of the buyer and friend of the seller; the two levels of expertise were high and low. The levels of estimate for the first source were \$400 and \$600; for the second source, they were \$300 and \$700.

In addition, there were four trials in which only the blue book value was presented: \$350, \$450, \$550, and \$650.

Judge's Point of View and Research Participants

All four experiments used the same test trials and warm-ups, but three different points of view were given to different groups of judges, who were instructed to identify with either the buyer, the seller, or an independent. The judges were 121 undergraduates at the University of Illinois who received extra credit in an introductory psychology course. A small number of additional students failed to follow instructions or complete the task and were excluded.

Experiments 1 and 2: Fair price. The judges were instructed to imagine that they had been hired by an "independent, unbiased assessor to estimate the 'true' value of many cars." They were told to neither overnor underestimate "true worth." Experiment 1, with 51 participants, was conducted a year before the others. Experiment 2, with 19 judges, was a replication of Experiment 1 conducted contemporaneously with Experiments 3 and 4. The data for Experiments 1 and 2 were analyzed separately and were virtually identical; consequently, the data were pooled.

Experiment 3: Buyer's price. The buyer's task was to estimate the highest price for which he or she would recommend buying a car. The 26 judges in Experiment 3 were instructed to imagine that they were acting as the agent of someone who would be buying used cars and to ask themselves, "What is the *maximum* amount I would advice *paying* for each car?"

Experiment 4: Seller's price. The seller's task was to estimate the lowest price for which he or she would recommend selling a car. These 25 judges were instructed to imagine that they were acting as the agent of someone who would be selling used cars and to ask themselves, "What is the *minimum* amount I would advise *accepting* for each car?"

Results

Figure 3 shows the interaction between expertise and bias for the two-source design, plotted for comparison with the predictions on the right in Fig. 2. Mean judgments of value are plotted as a function of the source's expertise, with solid points for the friend of the buyer (B) and open circles for the friend of the seller (S). As expected, judged values are greater when the source is a friend of the buyer than when he is a friend of the seller. The upper panels are for the first source, averaged over levels of both estimates and over expertise and bias of the second source. Lower panels are for the expertise and bias of the second source.

Panels A, B, and C show the results for the buyer's point of view (Experiment 3), the independent, "fair" point of view (Experiments 1 & 2), and the seller's point of view (Experiment 4), respectively. Comparing the three panels shows that the mean judgments are greater for estimates of the "lowest selling price" (Panel C) than they are for the "highest buying price" (Panel A).

All six panels show a divergent interaction: As expertise increases, the difference due to the source's bias increases. Divergence is also characteristic of plots made for individual judges. Analysis of variance tests of the inter-

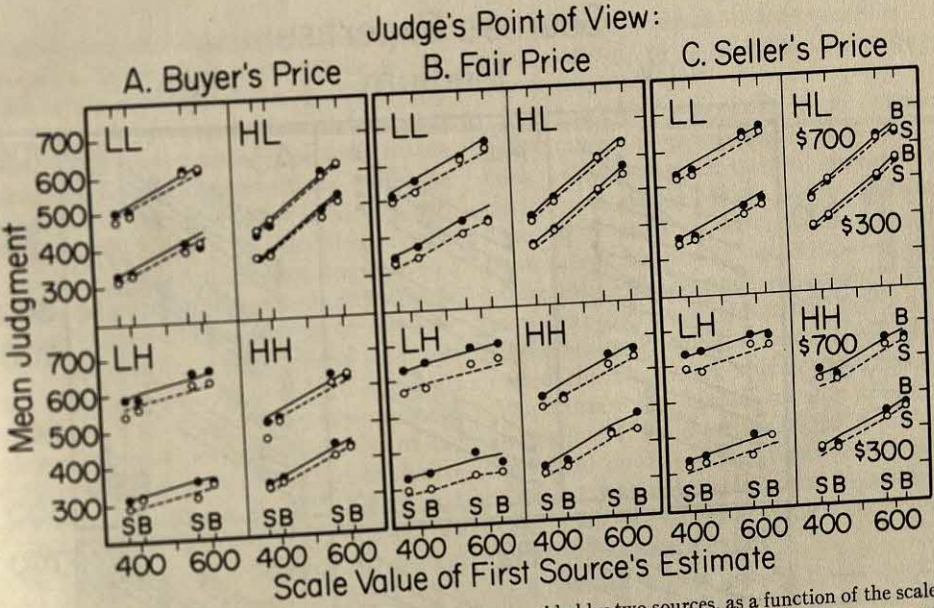


Figure 4. Mean judgment of value based on estimates provided by two sources, as a function of the scale value of the first source's estimate. (The four abscissa values represent scale values for estimates of $400 or $600, by a friend of either the buyer [B] or seller [S]. Upper panels represent data for second sources of low expertise [L]; lower panels are for second sources of high expertise [H]. Open circles and solid points represent second sources who are friends of the seller and friends of the buyer, respectively. Lines represent theoretical predictions based on scale-adjustment theory, with dashed lines for friends of the seller, Experiments 1–4.)

action between bias and expertise of the second source yielded $F(1, 25) = 12.7$, $F(1, 69) = 36.6$, and $F(1, 24) = 10.2$ for buyer's, fair, and seller's points of view, respectively. This divergence is predicted by the scale-adjustment model (Model 2 of Figure 2), which assumes that the source's bias causes a shift in the scale value of the information he provides. The solid and dashed lines plotted on Figures 3–5 represent predictions for Model 2 of Figure 2 (Equation 4). Model analyses will be discussed in a later section.

Figure 4 shows the results for all 64 two-source combinations, with a separate set of four panels for each point of view of the judge. Each panel within Figure 4 contains letters representing the expertise of the first and second sources, respectively. Thus the upper, left panel of each figure represents the results for two low-expertise sources (LL); the upper right panel shows the results for a high-expertise first source and a low-expertise second source (HL). The abscissa is spaced according to the scale value of the first source's estimate, derived from Equation 4. The letters S and B on the abscissa show the scale values for the first source's estimates (either $400 or $600), provided by a friend of the seller (S) or buyer (B), respectively. The four separate curves within each panel are for levels of bias and estimate of the second source. Solid points and lines are used for second sources who are friends of the buyer; open circles and dashed lines are used for friends of the seller.

There are four important results in Figure 4 that are common to all points of view, characteristic of the majority of individual subjects, statistically reliable, and (importantly) relevant to evaluation of the models. First, the effect of a source's estimate is greater for sources of higher expertise. Within each set of panels, proceeding from the upper left to the upper right corresponds to an increase in the expertise of the first source. Since the first source's estimate is on the abscissa, the increase in slope represents an increase in the effect of this estimate. The vertical separation between the two curves of the same type (either solid or dashed) represents the effect of the second source's estimate, which was

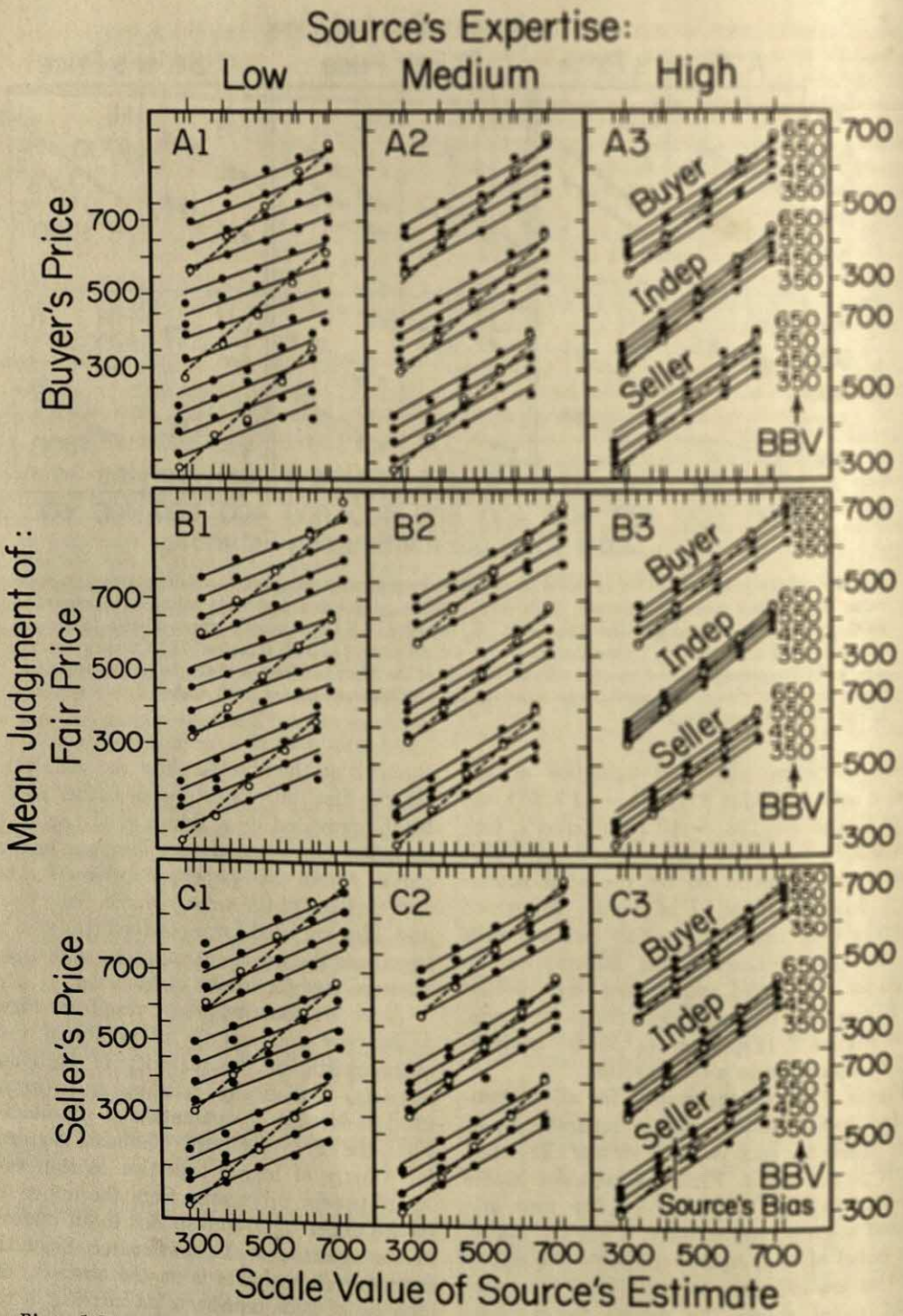


Figure 5. Mean judgment as a function of the source's estimate and blue book value. Panels A1–A3: Judgments of the highest price a buyer should pay (Experiment 3). Panels B1–B3: Judgments of "true" value (Experiments 1 & 2). Panels C1–C3: Judgments of the lowest acceptable selling price (Experiment 4). (Columns represent different levels of source expertise. Data for biased sources have been shifted vertically on the ordinate, $250 up for the friend of the buyer and $250 down for the friend

either \$300 or \$700. Dropping from an upper panel to a lower one corresponds to an increase in the expertise of the second source. Accordingly, the spread between the curves is greater in the lower panels. The tests of the Expertise × Estimate interactions for the first source yielded $F(1, 25) = 60.7$, $F(1, 69) = 239.2$, and $F(1, 24) = 45.0$ for the buyer's, fair, and seller's price judgments, respectively.

Second, the effect of either source's estimate is *inversely* related to the expertise of the *other* source. Thus, as the slopes increase from left to right, the spreads decrease. As the spreads increase from the upper to the lower panels, the slopes decrease. For the Expertise of the First Source × Estimate of the Second Source interaction, the *F*s were $F(1, 25) = 56.0$, $F(1, 69) = 329.3$, and $F(1, 24) = 59.6$ for buyer's, fair, and seller's price conditions, respectively.

Third, the effect of bias varies *directly* with the expertise of the same source. Note that the spread between open and filled circles (effect of the bias of the second source) is greater in the lower than the upper panels (when the expertise of the second source is greater). Fourth, the effect of bias varies *inversely* with the expertise of the other source. The spread between solid and dashed curves is *smaller* in the panels on the right (HL and HH, where the first source is high in expertise) than in the panels on the left. For the Expertise (of the first source) × Bias (of the second source) interaction, the *F* values were $F(1, 25) = 6.8$, $F(1, 69) = 26.2$, and $F(1, 24) = 21.3$, for buyer's, fair, and seller's price, respectively. Thus, bias interacts with expertise in the same fashion as estimate, consistent with the scale-adjustment model.

Figure 5 shows the results for the source-estimate and blue book value designs for all three points of view. The abscissa of each figure is spaced according to least squares estimates of the scale values of the source's estimate, based on Equation 4. The three leftmost notches on the abscissa, facing into each panel, show the positions of the scale values for an estimate of \$300 provided by a seller's friend, an independent, or a buyer's friend, respectively.

The solid points represent mean judgments based on a source's estimate and the blue book value (indicated next to curves). Open circles and dashed lines denote judgments based only on a source's estimate (no blue book value). Panels 1, 2, and 3 show the effects of increasing the source's expertise. It can be seen that proceeding from left to right, the slopes increase but the vertical spreads between the blue book value curves decrease. The curves for biased sources are shifted vertically (\$250 up for the friend of the buyer and \$250 down for the friend of the seller). The ordinate on the far right is labeled for these two; the left ordinate is labeled for the independent-source data.

A portion of the present experiment (independent sources, fair price condition) replicates Experiment 1 of Birnbaum et al. (1976). These 60 data points give virtually identical results to those of the earlier experiment.

The effects in Figure 5 for the source-estimate and blue book value design are highly reliable relative to the error terms. More importantly, similar effects occur in each experiment, providing evidence of replication. Statistical analyses confirm these graphic interpretations. For example, for the fair point of view, the $F(2, 138)$ for the main effect of bias was 102.3. The divergent interaction between bias and expertise, predicted by the scale-adjustment model, has an $F(4, 276) = 15.9$. The interaction between expertise and estimate yields $F(8, 552) = 122.5$. The theory that the weight of a source depends on expertise predicts the Expertise × Blue Book Value interaction, in which the effect of the blue book value is inversely related to expertise, $F(6, 414) = 114.7$. The effect of the blue book value also depends on the bias of the source, $F(6, 414) = 4.4$, and on the Expertise × Bias interaction, $F(12, 828) = 7.3$, consistent with the

---

of the seller [right-ordinate scales]. Separate curves are drawn for different levels of blue book value [BBV]. Open circles represent judgments based on one source's estimate only. Lines represent predictions based on scale-adjustment theory. Abscissa spacing represents fitted scale values, which are permitted to depend on source bias and estimate [Experiments 1–4].)

interpretation (Equation 4) that the source's weight depends on both bias and expertise.

The data for the blue book value and source-estimate design are consistent with the following: Scale value depends not only on the source's estimate but also on the source's bias. This theory (scale adjustment) explains the divergent Expertise × Bias interaction. Bias also affects the weight of a source, as evidenced by the differences in slope in Figure 5 for different bias–expertise combinations, in conjunction with the corresponding changes in the spread of the curves. Appendix A considers a complex alternative to all three models in Figure 2, which attempts to explain the effect of bias without scale adjustment. The complex model makes several incorrect predictions.

Confidence Intervals

The point size in Figures 4 and 5 is such that the solid and open circles contain a confidence interval of at least ±1 probable error in most cases. For Experiments 1 and 2, 42% of the means had standard errors less than \$6.00, 64% were less than \$8.00, 80% were less than \$10.00, and 93% were less than \$12.00. For Experiment 3, 35% of the means had standard errors less than \$8.00, 51% were less than \$10.00, and 63% were less than \$12.00. For Experiment 4, 50% of the standard errors were less than \$8.00, 58% were less than \$10.00, and 60% were less than \$12.00. These data seem quite neat, considering that adding another point size vertically on either side would include a 95% confidence interval in most cases.

Analyses for Individual Judges

Examination of data indicated that the group means are representative of the data for the vast majority of single judges. For example, four Expertise × Bias interactions were drawn separately for each of the 70 judges of Experiments 1 and 2. Model 2 predicts that the effect of bias should vary directly with the source's expertise and inversely with the expertise of the other source. The two-source design allows two examinations of each of these two predictions. Of the 70 subjects, 66% had either three or four interactions of the predicted form. Only 7 subjects had a greater number of inconsistent patterns than consistent patterns; 2 of these appear to have reversed the directions of bias for buyer and seller. Only 2 subjects had the pattern predicted by Model 3. Given that each point was the average of only 16 judgments and the fact that the Expertise × Bias interactions are predicted to be small, the small number of "deviant" judges doesn't provide evidence for the existence of large subgroups obeying different models. Of the 280 figures, 206 were of the predicted form, the same proportion for each prediction, suggesting that the group means are highly representative of individual data.

Model Analyses

Choice among models. Model 2 is vastly superior to Models 1 and 3, since it correctly anticipates the divergent Expertise × Bias interaction, in which the effect of bias is magnified by expertise. Accordingly, the 289 data points for each experiment were separately fit to Equation 4 by means of a computer program, which used Chandler's (1969) STEPIT subroutine to minimize the sum of squared model–data discrepancies.[1]

The data of Experiments 1 and 2 were analyzed separately with nearly identical results, showing excellent cross-validation of both the model and parameter values upon replication. The data of Experiments 1 and 2 are pooled in the analyses reported here.

Weights and scale values. The scale value for the blue book value of \$550 was set to its monetary value, and the weight of the blue book was arbitrarily set to 1.0. There were 29 parameters to estimate, 9 weights for the sources (Expertise × Bias), 15 scale values (Estimate × Bias), 3 scale values for the blue book value, and a weight and a scale value for the initial impression.

The weights of the initial impression were small: .08, .08, and .10 for the buyer's price, fair price, and seller's price, respectively. Scale

[1] We thank Ron Hinkle for checking our parameter estimates from STEPIT against those obtained by his subroutine, BLACKBOX, which uses an improved minimization algorithm.

Table 1
Estimated Weights and Scale Values for the Scale-Adjustment Model

Variable	Judge's point of view: Buyer's price — Source's bias: Buyer	Independent	Seller	Fair price — Source's bias: Buyer	Independent	Seller	Seller's price — Source's bias: Buyer	Independent	Seller
Estimated weights of sources[a]									
Source's expertise									
Low	.68	.59	.71	.76	.62	.84	.90	.76	.92
Medium	1.35	1.32	1.32	1.43	1.62	1.62	1.42	1.49	1.46
High	2.22	3.21	2.25	2.63	4.33	2.77	2.76	3.45	2.44
Estimated scale values[b]									
Source's estimate									
$300	307	291	271	330	301	267	330	297	276
$400	407	397	370	430	401	364	428	404	380
$500	515	491	464	538	502	459	534	504	474
$600	608	594	562	633	601	561	629	607	576
$700	692	686	647	729	691	651	723	702	671

[a] Each entry in the upper portion of the table is the estimated weight for each source as a function of the source's expertise and bias. A separate analysis was performed for each point of view. The weight of the blue book value was set to 1.0. For example, the weight of the high-expertise friend of the buyer was 2.22 for Experiment 3 (buyer's price).

[b] Each entry in the lower portion of the table is the estimated scale value as a function of the source's estimate and the source's bias. For example, the largest scale value for Experiment 3 (buyer's price), $692, was for an estimate of $700 by a friend of the buyer. For Experiment 4 (seller's price), the same estimate from the same source had an estimated scale value of $723.

values for the initial impression were related to the point of view: 243, 368, and 390 for buyer's, fair, and seller's price, respectively. Estimated scale values for the blue book were 339, 446, 550, and 648 for the fair point of view (values were very similar for the other points of view). The least squares estimates of weights and scale values are shown in Table 1.

Table 1 shows that the weights depend mostly on expertise, but tend to be larger for the independent source of high expertise and possibly smaller for the low-expertise, independent source. This pattern of weights appeared in all experiments. The weights for the 9 sources estimated from the least squares analysis are consistent with weights estimated graphically, from the effect of blue book value, using the method of Figure 1.

The scale values depend on three factors: estimate, source's bias, and judge's point of view. For example, a $500 estimate provided by an independent has a scale value of only 491 from the buyer's point of view, 502 from the fair price point of view, and 504 from the seller's point of view. If the $500 estimate in the fair price condition is provided by a friend of the buyer, the scale value jumps to 538, compared with only 459 if the estimate is provided by a friend of the seller.

Fit of the scale-adjustment model. Predictions of the model are shown in Figures 3, 4, and 5 and come very close to the data. The square roots of the mean squared model–data discrepancies were 12.15, 10.16, and 11.16 for buyer's, fair, and seller's price, respectively. Hence, the average discrepancy between the model and sample means is not much larger than the expected discrepancy (standard error) between sample and population means.

Figures 3, 4, and 5 show that Model 2 makes the following correct predictions: First, it correctly predicts the divergent Expertise × Bias interactions. These interactions are best seen in Figure 3, where they have been plotted for comparison with Figure 2. The

Expertise × Bias interactions can also be seen in Figure 4. Figure 5 does not permit one to see this prediction easily, but when data for the Blue Book Value × Source Estimate design were plotted as in Figure 3, a similar pattern was evident: The greater the expertise, the greater the effect of bias.

Second, Model 2 gives a good description of the slopes and spreads of the curves in Figures 4 and 5. The greater the expertise of a source, the greater the effect of that source's estimate and bias and the *less* the effect of the estimate and bias of another source (Figure 4) or of the blue book value (Figure 5).

Third, Model 2 gives a good account of the single-source data (open circles and dashed lines in Figure 5), which cross the curves for different levels of blue book value.

Deviations of fit. The deviations from the model should be taken seriously, since each circle in Figures 4 and 5 contains a fair-sized confidence interval. The model predicts that the curves in each set in Figure 5 should be parallel. Instead, the interaction between estimate and blue book value shows a divergence to the right for the buyer's and the independent's points of view, $F(12, 300) = 6.6$ and $F(12, 828) = 11.2$, respectively. Similar divergence was also obtained by Birnbaum et al. (1976) and can be described by a configural-weight model, which assigns greater weight to the lower estimate. A configural-weighting revision of the scale-adjustment model is tested in Experiment 5.

There is also a hint of a higher order configural effect. When the judge has the seller's point of view and the buyer provides a low estimate, the effect of blue book value is greater than predicted (points are spread wider than predictions), as though the buyer's low estimate receives reduced weight. When the buyer provides a higher estimate, the data points are compressed, as if the buyer receives higher relative weight for providing an unexpected estimate.

In summary, the results of Experiments 1–4 are generally consistent with the scale-adjustment model (Model 2) of Figure 2. Deviations are small but regular. Consequently, Experiment 5 explores these deviations, testing modifications of Model 2 that allow the weight of an estimate to depend on the stimulus configuration.

Experiment 5: Test of Configural Weighting Models

Experiment 5 addressed three issues raised by the first four experiments. First, although the between-subjects manipulation of point of view did affect scale values in a predictable fashion, the pattern of weights did not support a prior conjecture that the weight of a source would be greater when the source's bias matched the judge's point of view (Table 1). Interpretation of between-subjects results requires extreme caution, however, since the effects of a variable such as point of view may depend on the establishment of a context of comparison for the individual judge. Consequently, Experiment 5 allowed each judge to experience all three points of view.

Second, the fact that the effect of a source's estimate and bias varies inversely with the number and expertise of other sources would also be consistent with a linear-weighted model in which the effective weight of a source varies with the *difference* between the source's weight and some function of the total absolute weight instead of the *ratio;* that is, the effective weight would be $w - f(\Sigma w)$ instead of $w/\Sigma w$. In order to test this linear-weight model against relative-weight averaging models (such as Model 2), it is necessary to have at least two sources that each take on more than two levels of expertise. Therefore, Experiment 5 was designed to compare the success of the relative-weight model versus the linear-weight model.

Third, systematic deviations of fit appeared in Experiments 1–4. These deviations appear consistent with previous results obtained by Birnbaum (1972, 1973, 1974) and by Birnbaum et al. (1976). The deviations in Experiments 1–4, although convincing, were not large enough to warrant detailed theoretical analysis. In previous tests of averaging models, deviations from parallelism were found to be related to the range (maximum–minimum) of scale values within the set of stimuli to be integrated. Consequently, Experiment 5 contained greater variation in the range of scale values within sets in order to permit a better examination of the deviations from the averaging model.

Configural-Weight Theory

To account for stimulus interactions, simple configural-weight theories have been proposed and tested by Birnbaum, Parducci, and Gifford (1971), Birnbaum (1972, 1973, 1974), and Birnbaum and Veit (1974). The scale-adjustment model (Model 2 of Figure 2) can be modified to allow for configural weighting. The simplest configural-weight theory was termed the range model. According to this model, the relative weight of a stimulus depends in part on the rank of that stimulus in the configuration of stimuli to be integrated on a given trial.

For two stimuli, the range model may be written:

$$R = \frac{w_0 s_0 + w_1 s_1 + w_2 s_2}{w_0 + w_1 + w_2} + \omega_P |s_1 - s_2|, \quad (6)$$

where ω_P is the weight of the configural, or range ($|s_1 - s_2|$) effect, which, in the present case, would presumably depend on the judge's point of view. Note that when $s_1 > s_2$, the relative weight of s_1 can be written:

$$\frac{w_1}{w_0 + w_1 + w_2} + \omega_P. \quad (7)$$

However, when $s_1 < s_2$, the relative weight of s_1 can be written:

$$\frac{w_1}{w_0 + w_1 + w_2} - \omega_P. \quad (8)$$

The range model assumes that the effective relative weight of a stimulus depends on the rank of its scale value in the set of stimuli to be combined. As a limiting case, when ω_P equals the relative weight of a stimulus, the range model can become a maximum or minimum (conjunctive or disjunctive) model, depending on the sign of ω_P. The model implies that the response varies linearly with the range of scale values, holding mean scale value constant.

If the two stimuli are two estimates provided by different sources, if the weight depends on the expertise and bias of the source (e.g., $w_1 = w_{X_1B_1}$), and if the scale value depends on the source's estimate and bias (e.g., $s_1 = s_{E_1B_1}$), then Equation 6 becomes an extension of the scale-adjustment model (Model 2 of Figure 2). This configural-weight model adds a single parameter, ω_P, to account for the effects of the judge's point of view.

Note that ω_P in Equation 7 is the amount of relative weight taken from the lower valued stimulus and given to the higher. If ω_P is negative, weight is added to the lower valued stimulus, yielding a divergent stimulus interaction. Such an interaction might be expected in the buyer's point of view. A convergent interaction (positive ω_P) might be expected in the seller's point of view, since the larger of two estimates should receive greater weight.

Method

The instructions, stimuli, and procedure were similar to those of Experiments 1 through 4. The chief differences were as follows: (a) Each judge was exposed to all three points of view via instructions and then judged all of the 146 stimulus combinations separately under each point of view. (b) Descriptions of the levels of expertise were modified to provide five levels. (c) The experimental designs in Experiment 5 permitted assessment of previously untested predictions of the models.

Experimental Designs

The trials selected were a subset of a $(5 \times 3 \times 9) \times (5 \times 3 \times 9)$, First Source (Expertise $\times$ Bias $\times$ Estimate) $\times$ Second Source (Expertise $\times$ Bias $\times$ Estimate) factorial design. The complete design would have required 18,225 trials; therefore, six smaller factorial designs involving these six variables were selected to test particular implications of the models. These designs are shown in Table 2 and described below.

Estimate $\times$ Estimate. Two related $(1 \times 1 \times 4) \times (1 \times 1 \times 4)$ factorial designs investigated the hypothesis that the configural weight of the lower scale value depends on the point of view of the judge. In both designs, both sources were medium in expertise; the levels of estimate for Source 1 were \$350, \$450, \$550, and \$650; and the levels of estimate for Source 2 were \$300, \$400, \$600, and \$700. In one design, the first source was a friend of the buyer and the second was a friend of the seller; in the other design, the biases were reversed.

Bias $\times$ Bias. The design was a $(1 \times 3 \times 2) \times (1 \times 3 \times 2)$, in which both sources were medium in expertise; the three levels of bias for both sources were friend of the buyer, independent, and friend of the seller; the estimate of Source 1 was either \$350 or \$650; the estimate of Source 2 was either \$300 or \$700. The purpose of the Bias $\times$ Bias design was to investigate the hypothesis that when a source provides an estimate contrary to what one would expect from that source's bias, the weight of that estimate is increased.

Bias $\times$ Expertise. Two related designs reexamined the predictions of Model 2 that expertise amplifies the effect of bias (and estimate) in the same source and

Table 2
Two-Source Designs in Experiment 5

Design	First source			Second source		
	Expertise	Bias	Estimate	Expertise	Bias	Estimate
$Est_1 \times Est_2$ (BS)	1 (Med)	1 (B)	4	1 (Med)	1 (S)	4
$Est_1 \times Est_2$ (SB)	1 (Med)	1 (S)	4	1 (Med)	1 (B)	4
Bias × Bias	1 (Med)	3	2	1 (Med)	3	2
$Bias_1 \times Exp_2$	1 (Med)	3	2	5	1 (I)	1 ($500)
$Bias_1 \times Exp_1$	5	3	2	1 (Med)	1 (I)	1 ($500)
Exp × Exp	5	1 (I)	1 ($300)	5	1 (I)	1 ($650)

Note. Each entry represents the number of levels of the factor listed above the column. The product of the entries in each row gives the number of cells in each design. For example, the first row indicates that an Estimate (Est) × Estimate design consisted of 16 trials in which the first source was a medium-(Med) expertise, friend of the buyer (B) who provided one of four estimates, and the second source was a medium-expertise friend of the seller (S) who provided one of four estimates. The last row shows that the Expertise (Exp) × Expertise design contains 25 cells in which the expertise of each source could attain one of five levels, the first source was an independent (I) who gave an estimate of $300, and the second source was also an independent with a $650 estimate.

diminishes the effect of bias (and estimate) in a second source. A (1 × 3 × 2) × (5 × 1 × 1) factorial design used five levels of expertise (very low, low, medium, high, or very high) for the second source, an independent who gave an estimate of $500, combined with a medium-expertise first source, with three levels of bias of Source 1 (S, I, B) and two estimates, $350 or $650. A second factorial design, a (5 × 3 × 2) × (1 × 1 × 1), was identical, except that it shifted the five levels of expertise to the first source and held the expertise of the second source at medium. The purposes of these two designs were to constrain the estimates of weights and to replicate the Bias × Expertise × Estimate interactions of Experiments 1–4.

Expertise × Expertise. To test the averaging model against the linear-weighted model requires more than two levels of expertise for two sources. Hence, a (5 × 1 × 1) × (5 × 1 × 1) factorial design was employed in which the five levels of expertise for each source were very low, low, medium, high, and very high. The bias of each source was "independent." The estimate of the first source was $300, whereas the second source gave an estimate of $650.

The designs employing two sources generate a total of 153 stimulus sets, 15 of which are shared by two designs, leaving 138 distinct two-source test trials. These designs constrain all of the two-way interactions and many of the three- and four-way interactions among the six variables. In addition, 8 trials were produced by a single-source design, in which only one source provided an estimate. The design was a 2 × 2 × 2, Expertise × Bias × Estimate, factorial design, in which the two levels of expertise were low and high, the two levels of bias were friend of buyer and friend of seller, and the estimate was either $300 or $700.

Procedure

The 146 trials were printed in random orders in booklets in three sets, with additional instructions and 10 warm-up trials reminding the judge of the point of view for each set. Judges were permitted to work at their own paces, completing the 438 experimental trials in about 2 hours.

Research Participants

The judges were 60 undergraduates at the University of Illinois who received extra credit in introductory psychology for participating. Ten students served in each of the six possible orders of three points of view.

Results

The upper panels of Figure 6 show mean judgments as a function of the estimate provided by a medium-expertise seller with a separate curve for each level of estimate provided by a medium-expertise buyer. The lower panels plot judgments as a function of the buyer's estimate, with a separate curve for each level of seller's estimate. Panels A to C represent the judge's point of view. Lines in Figures 6–9 are predictions of the range model (Equation 6), an extension of Model 2 of Figure 2, using a different value of w_p for each point of view. This model is discussed in a later section.

Figure 6 shows a striking result. Whereas the models considered in Experiments 1–4 predict parallel curves in each panel, the curves in Figure 6 are systematically non-parallel. Furthermore, the interactions are related to the judge's point of view. The curves

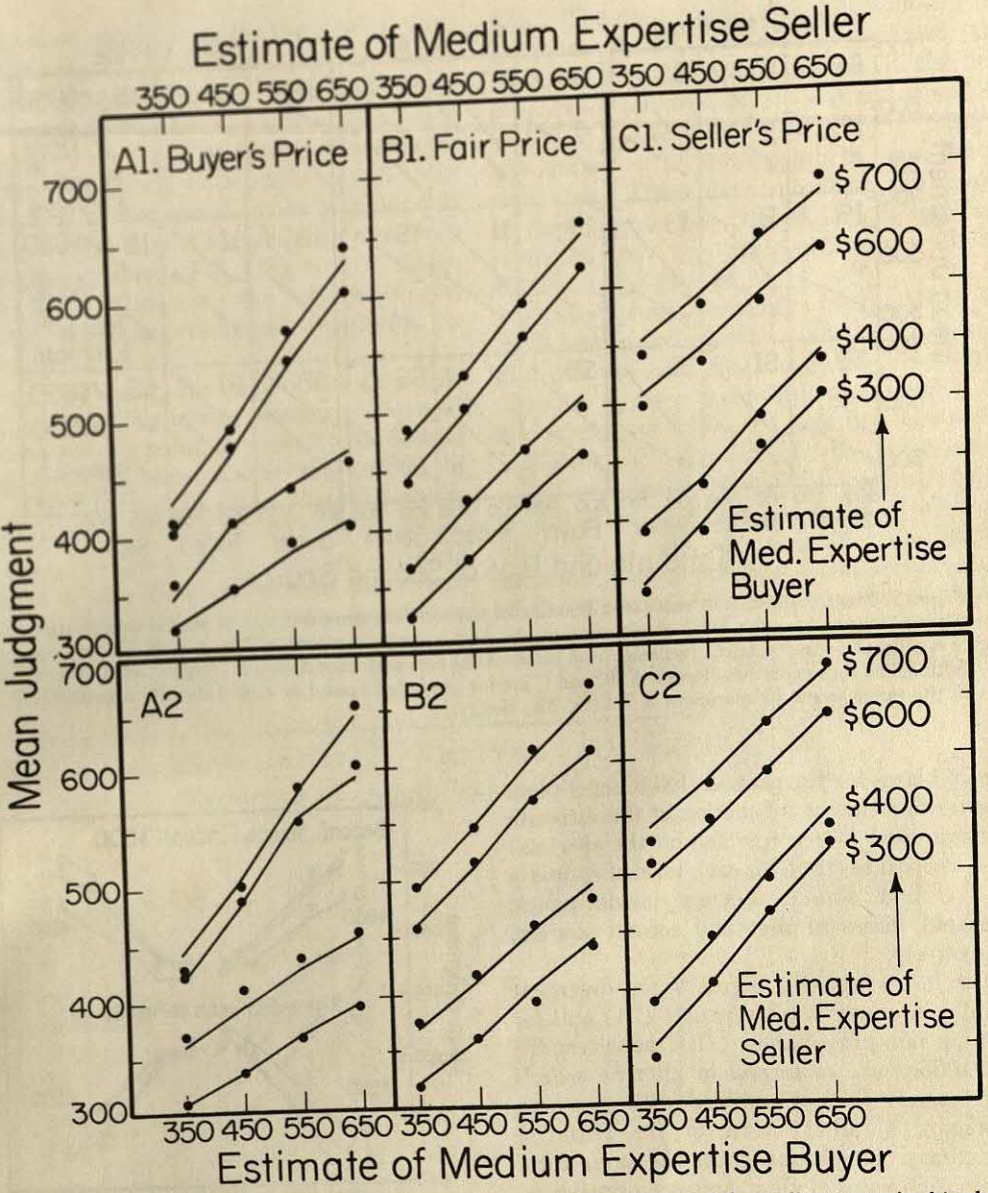


Figure 6. Upper panels: Judgment of value as a function of the estimate of the medium-expertise friend of the seller with a separate curve for each estimate of the medium-expertise friend of the buyer. Lower panels: Judgment of value as a function of the estimate of the buyer's friend (on the abscissa) with separate curves for the estimates of the seller's friend. (Panels A, B, and C are for the judge's point of view. Lines represent predictions of the range model [Experiment 5, Estimate × Estimate designs].)

diverge for the buyer's point of view and for the fair price point of view; however, they converge for the seller's point of view. It is as if the judge places greater weight on the lower estimate when identifying with the buyer and places greater weight on the higher estimate when identifying with the seller. The three-way interaction of Point of View × First Estimate × Second Estimate yielded Fs(18, 1062) = 11.8 and 11.3 for the two panels of Figure 6.

Figure 7 shows Estimate × Estimate interactions with a separate panel for each combina-

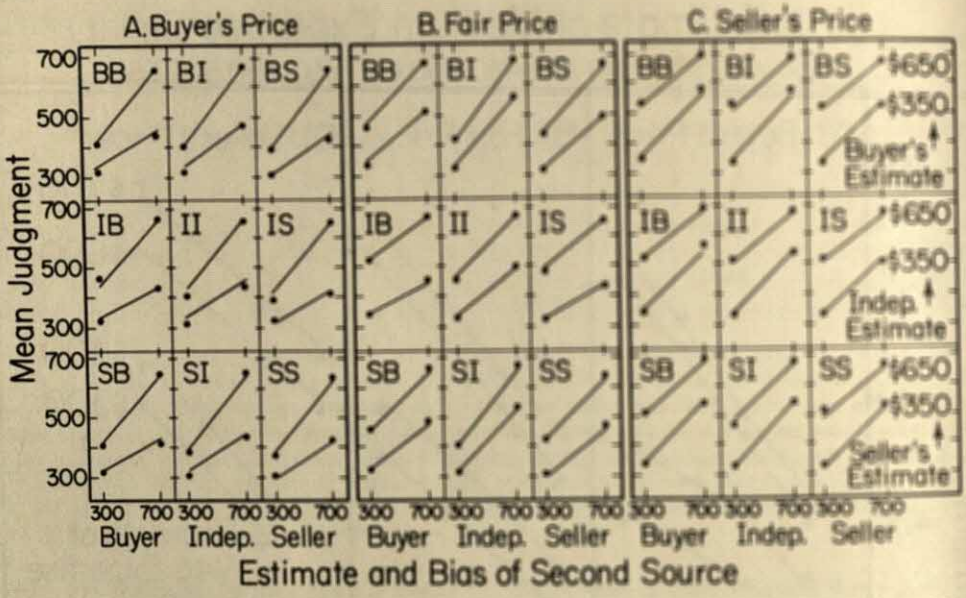


Figure 7. Mean judgment of value as a function of the estimate provided by the second source, with a separate curve for each level of estimate of the first source. (Letters within panels [B = buyer; S = seller; I = independent] represent the biases of the first and second sources, respectively, who were both medium in expertise. Panels A, B, and C are for the judge's point of view. Lines show predictions of the range model [Experiment 5, Bias × Bias design].)

tion of biases for the sources. Each panel plots mean judgments as a function of the estimate of the second source (plotted on the abscissa) with a separate curve for each level of estimate of the first source. Letters inside panels represent biases of first and second sources, respectively.

The interactions in Figure 7 are divergent for all nine buyer's price panels (7A) and for all nine fair price panels (7B); however, the interactions are *convergent* in all nine seller's price panels (7C). Combined with the data of Figure 6, all 11 tests of the Estimate × Estimate interaction show the same form within each point of view. Assuming a relative-weight averaging model (all of the models in Figure 2 predict parallelism), it would be extremely unlikely to obtain 11 divergent or convergent interactions, all of the same type, within each point of view.

The group means are highly representative of data for individual judges. For example, 58 of 60 judges show the divergent Estimate × Estimate interaction of Figure 7 for the buyer's point of view; 48 judges show divergence for the fair price of view, and 40

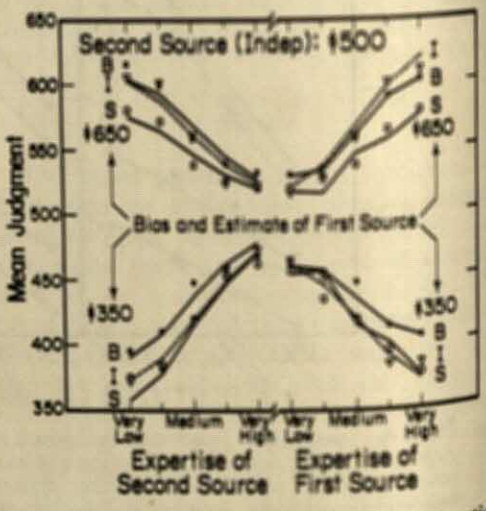


Figure 8. Left: Mean judgment of value as a function of the expertise of the independent second source (who said $500), with a separate curve for each level of bias (B = buyer; S = seller; I = independent) and estimate of the first source. Right: Mean judgment of value as a function of the expertise of the first source. (The second source was a medium-expertise independent who said $500. Lines show predictions of the range model [Experiment 5, Bias × Expertise designs].)

judges show convergence for the seller's point of view. Hence, these data provide strong evidence that the Estimate × Estimate interactions of Experiments 1–4 and of Birnbaum et al. (1976) are "real" and, most importantly, that the Estimate × Estimate interaction can be reversed by manipulating the judge's point of view. For the data of Figure 7, this three-way interaction yielded $F(2, 118) = 39.7$. Thus, some modification of the relative-weight models, such as configural weighting, is necessary.

Figure 8 shows the results for the Bias × Expertise × Estimate designs, averaged over the judge's point of view. These designs retested the implications of the models of Figure 2. The left panel of Figure 8 shows that as the expertise of the second source increases, the effects of the bias and estimate of Source 1 (the distances between the curves) decrease. The Expertise × Bias interaction yielded $F(8, 472) = 6.1$, and the Expertise × Estimate interaction had an $F(4, 236)$ of 226.1. The right-hand panel shows that as the expertise of the first source increases, the effects of bias and the estimate of the first source increase, $F(8, 472) = 4.8$ and $F(4, 236) = 158.2$, respectively. The general pattern was similar for all three points of view, drawn separately. These data reconfirm the results of Experiments 1–4; this pattern of results is predicted by the scale-adjustment model (Model 2 of Figure 2).

Figure 9 shows the results for the Expertise × Expertise design, with a separate panel for each point of view. The abscissa represents the expertise of an independent source who gave an estimate of \$650; the separate curves are for levels of expertise of the independent source who said \$300. The positive slopes indicate that judged value increases with the expertise of the source giving the higher estimate. Judged value decreases as a function of the expertise of the source giving the lower estimate.

Averaging models predict a nonparallel set

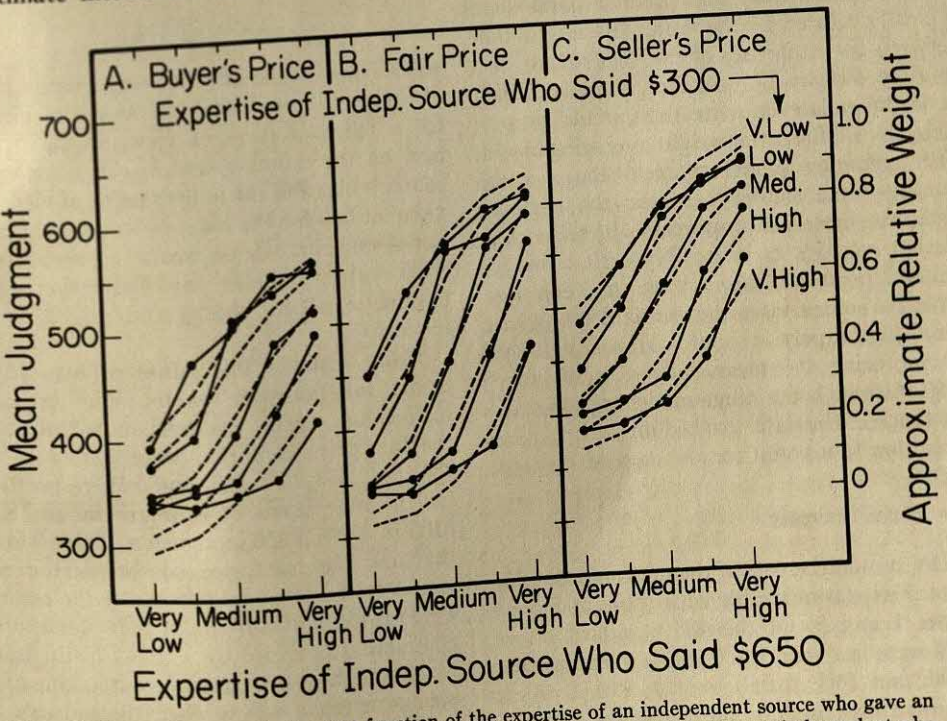


Figure 9. Mean judgment of value as a function of the expertise of an independent source who gave an estimate of \$650, with a separate curve for each level of expertise of the first source, an independent who said \$300 (Panels A, B, and C are for buyer's, fair, and seller's points of view, respectively. Dashed curves show predictions of the range model. Right ordinate represents an approximate scale of relative weight [Experiment 5, Expertise × Expertise design].)

Table 3
Estimated Bias Adjustments for the Range Model

Point of view	Source's bias		
	Buyer	Independent	Seller
Buyer's price	8.79	6.39	−33.50
Fair price	20.22	0	−34.76
Seller's price	24.26	1.26	−14.79

Note. Each entry is the estimated value of b_{BP} in Equation 9: $s_{EBP} = s_E + b_{BP}$. The estimated values for s_E are \$322, \$352, \$410, \$451, \$511, \$563, \$620, \$662, and \$703 for the nine levels of estimate ranging from \$300 to \$700. For example, the largest value in the table (24.26) means that the scale value for an estimate provided by a friend of the buyer raises the scale value \$24.26 when the judge takes the seller's point of view. Thus, an estimate of \$500 has a scale value of \$511 + \$24 = \$535 in this condition. The negative numbers show that when a friend of the seller provides the estimate, the scale value is reduced. The value for the independent source in the fair price condition was set to 0.

of curves that bulge in the middle. An opposing linear-weights theory that predicts parallelism is clearly refuted by the data. The interaction between the expertises of the two sources has an $F(16, 944) = 29.7$. In fact, the data pinch in at the ends even more than would be predicted by a differential-weight averaging model. The averaging model predicts that as the expertise of a source increases, the response should asymptotically approach the scale value of that source's estimate. Note that for the buyer's point of view, when a very low-expertise source rates the car at \$300, even a very high-expertise source reporting \$650 cannot bring the mean judged value above \$550. Although the range model approximates the shape of the data (dashed lines), it requires a revision to account for the data of Figure 9.

Confidence Intervals

The standard errors of the means for Experiment 5 were comparable with those of Experiments 1 and 2. Of the 483 standard errors, 37% were less than \$6.00, 64% were less than \$8.00, and 76% were less than \$10.00. As in the other experiments, the size of the standard error correlates with the range of estimates. The point size in Figure 7 contains a fair-sized confidence interval; however, note that the ordinate scales in Figures 6, 8, and 9 are expanded, so that a point size does not necessarily contain a large confidence interval.

Model Analyses

Range model. The range model (Equation 6) was fit to the data by means of a computer program utilizing the STEPIT subroutine, designed to minimize the sum of squared data-model discrepancies over all 483 (161 × 3) cells in the experiment. Except for the configural-weight parameter, w_P, the model is a form of scale-adjustment model (Model 2 of Figure 2).

The number of parameters required to estimate scale values was reduced by the simplifying assumption that the scale value of a biased source's estimate is an additive function of a scale value depending on the estimate alone and a bias parameter depending on the source's bias and the judge's point of view:

$$s_{EBP} = s_E + b_{BP}, \tag{9}$$

where s_{EBP} is the effective scale value for an estimate, E, provided by a source of bias, B, for a judge of point of view, P; s_E depends only on the estimate, and b_{BP} depends on the source's bias and the judge's point of view. The value of b_{BP} for the independent source in the fair price point of view was set at zero. Hence, scale values require the estimation of 17 parameters, 9 for the s_E and 8 ($3 \times 3 - 1$) for b_{BP}.

Table 3 shows the values of bias adjustments for the scale values. The estimated scale values, s_E, for the independent source for the fair price condition were 322, 352, 410, 451, 511, 563, 620, 662, and 703, respectively, for the nine levels of estimate ranging from \$300 to \$700 in \$50 increments. Table 3 shows that for the fair price condition, the scale values would be about \$35 less if the estimate were provided by a friend of the seller or about \$20 more if provided by a friend of the buyer. As one might expect, the absolute value of the bias adjustment is less when the source's bias matches the judge's point of view, as if the judge views a source of his or her own point of view as less biased.

To permit examination of how weights for

different sources might depend on point of view, a different weight was estimated for each type of source in each point of view. Thus, 45 weights, w_{XBP}, were estimated for the $5 \times 3 \times 3$, Expertise $\times$ Bias $\times$ Point of View combinations. The weight of the low-expertise independent in the fair price condition was set to 1.0, leaving 44 parameters to be estimated.

The least squares estimates of weights are shown in Table 4. The general pattern of weights in Experiment 5 is similar to that of Experiments 1–4. The ratio of the weight of high- to low-expertise independents for the fair price condition is 6.78, not far from the corresponding value of 6.98 for Experiments 1 and 2. The weights depend mostly on expertise, with the greatest weights being for independents of very high expertise. For the buyer's point of view, the buyer's weight exceeds that of the seller for all levels of expertise. For the seller's point of view, the friend of the seller receives greater weight than the corresponding buyer for the very high-expertise source. There is thus a very slight hint of support for the notion that biased sources receive relatively higher weight when the judge shares the source's bias, but the evidence does not seem decisive.

In addition, separate weights and scale values for the initial impression and separate values of ω_P, the configural weight parameter, were permitted for each point of view. These 9 parameters, plus 44 source's weights, plus 17 for scale values make a total of 70 parameters to be estimated to describe the data in Figures 6–9.

The estimated scale values of the initial impressions were 281, 308, and 413, with weights of .30, .15, and .24, for buyer's, fair, and seller's price, respectively. The values of the configural-weight parameter, ω_P, were $-.194$, $-.066$, and .055 for the buyer's price, fair price, and seller's price, respectively.

The predictions of this model are shown in Figures 6–9. The model is an extension of Model 2 of Figure 2, with only one elaboration—the configural-weight effect to allow for Estimate $\times$ Estimate interactions. Considering that only one additional parameter is used for each point of view, the range model gives a very good account of the data in Figure 6 and a good account of the data of Figure 7. The root mean squared errors were 6.83 and 10.40 for Figures 6 and 7, respectively. The overall sum of squared deviations was 61,497 over 483 points, yielding a root mean squared deviation of 11.28. This value was 13.08, 10.23, and 10.30 for the buyer's, fair, and seller's price conditions, respectively.

Table 4
Estimated Weights of Sources for the Range Model

Source's bias	Expertise of source: Very low	Low	Medium	High	Very high
Buyer's point of view					
Buyer	.91	.80	2.42	4.52	6.42
Independent	1.00	1.53	3.05	5.70	9.98
Seller	.76	.45	2.39	2.63	4.90
Fair price point of view					
Buyer	.66	.77	1.40	3.22	4.89
Independent	.47	1.00	2.68	6.78	17.57
Seller	.43	.38	1.50	2.34	4.13
Seller's point of view					
Buyer	.96	1.43	2.37	4.10	4.95
Independent	.75	1.22	2.62	5.44	10.20
Seller	.85	1.08	2.25	3.47	6.35

Note. Each entry is the least squares estimate of weight for each type of source under each condition. The weight of the low-expertise independent source in the fair price condition was set to 1.0. For example, the table shows that when the judge takes the seller's point of view, the very high-expertise friend of the seller has greater weight (6.35) than the very high-expertise friend of the buyer (4.95).

The range model would have been a great success had the data of Figure 9 not been collected. When a very high-expertise, independent source gives an estimate of \$300 and a very low-expertise independent source gives an estimate of \$650, the mean judgment for the buyer's point of view is \$342, compared with a prediction of \$292, a deviation of about \$50. The range model makes this extremely low prediction because the relative weight of the \$650 estimate is so low in this case (.09) that when the negative value of $\omega_P(-.194)$ is added to it, the effective relative weight of the \$650 estimate becomes negative. In other words, the best-fit value of ω_P for the entire

experiment (including the data of Figures 6 and 7) is a poor value for the data of Figure 9.

To examine the effective relative weights, the right ordinate of Figure 9 has been labeled from 0 to 1. Since the scale values are approximately equal for the independent sources from different points of view, and since the weight of the initial impression is so small, the right-ordinate values can be read off to give the approximate relative weight of the $650 estimate for each point in Figure 9.

Revised configural-weight theory. The assumption of the range model that the configural change of weight is independent of relative weight appears inconsistent with the data of Figure 9. The model was revised to incorporate the principle that the amount of *decrease* in absolute weight is directly proportional to the absolute weight apart from the configural effect. Weight taken from one stimulus is given to another. Hence, if a source's estimate loses weight, it does so in proportion to its original weight; if a source's estimate gains weight, it does so in proportion to the weight of the other source's estimate. Thus, there is a conservation of absolute weight: The loss of one source is another source's gain.

The revised configural-weight model is a scale-adjustment model (Equations 4 & 9). It is distinguished from the range model (Equation 8) in that the configural change of weight operates on absolute rather than relative weight. The weights are given by the following equation:

$$w_{XDFE} = w_{XDF} + w_F \sigma_{FFE}, \qquad (10)$$

where w_{XDFE} is the effective absolute weight of an estimate; w_{XDF} is defined as in the range model; w_F is the configural-weight parameter; $\sigma_E = 1$, if the estimate is the largest in the set, $\sigma_E = -1$ if it is the smallest, and $\sigma_E = 0$ otherwise. By definition, $\rho_{FE} = w_{XDF}$ if $w_F \leq 0$; otherwise, $\rho_{FE} = w_{X\cdot D\cdot F}$, where $w_{X\cdot D\cdot F}$ is the absolute weight of the other source. Thus, a source loses weight in proportion to its original value but increases by taking the weight lost by the other source.

This modified configural-weight model (Equations 4, 9, & 10) provides a great im-

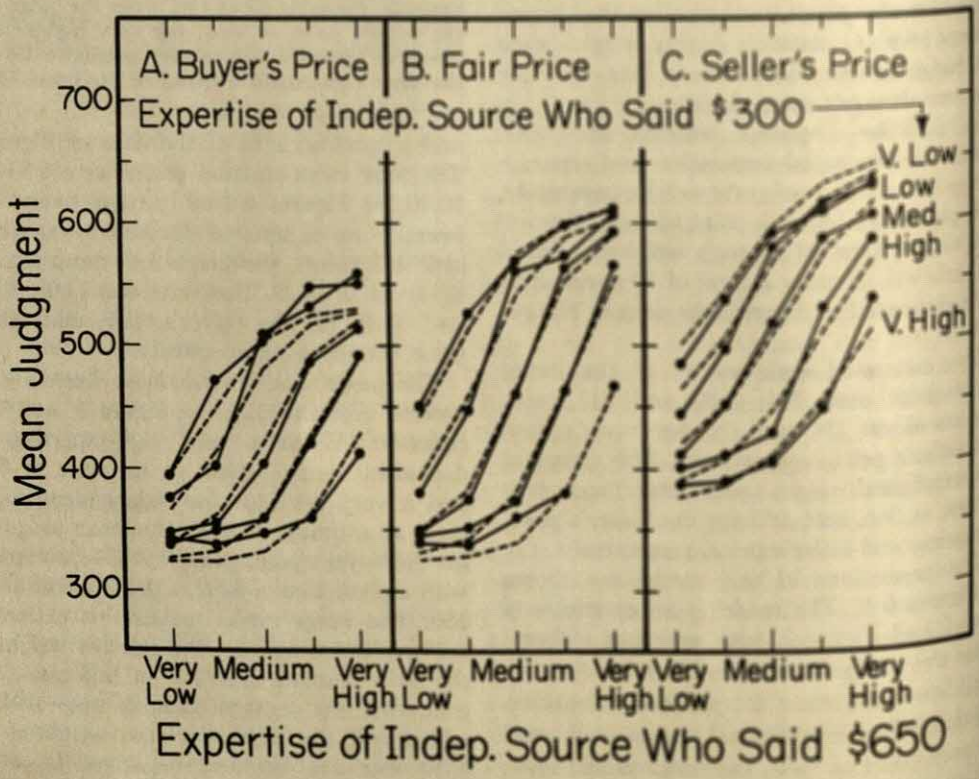


Figure 10. Data of Figure 9 with predictions (dashed curves) of revised configural-weight theory.

provement over the simple range model (Equations 4, 6, & 9), though it uses the same number of parameters. The total sum of squared deviations over all 483 cells was reduced from 61,497 to 48,333, yielding a root mean squared error of 10.00. The improvement in fit was greatest for the Expertise × Expertise design (Figure 9), where the root mean squared mistake was reduced from 17.24 to 13.09. The fit was also noticeably better for the Expertise × Bias design (Figure 8). Fit was about the same or slightly better in the other designs.

Figure 10 shows the fit of the improved configural-weight model to the Expertise × Expertise design. Solid points are empirical means; dashed curves represent predictions. As can be seen by comparing Figures 9 and 10, the revised configural-weight theory gives a superior description of these data, especially for the fair price point of view.

The estimated scale values and bias parameters of this configural model were similar to those of the range model. The pattern of estimated weights was also similar, except that the weights always increased with expertise for this model (an improvement), and the range of weights increased (especially for the independent sources). The values of ω_P were −.385, −.155, and .130 for the buyer's, fair, and seller's points of view, respectively. An extension of the configural-weight model is discussed in Appendix B.

Differential-weight theory. In differential-weight theory, the weight of a stimulus is permitted to depend on the estimate. For the present data, the weight would be required to depend on point of view as well. The most complex, scale-adjustment, differential-weight model allows an increment in weight to depend on three factors: bias, point of view, and estimate. There are 81 values of the weight increment, 9 of which (for the independent, fair price conditions) were set to 0, leaving 72 weight parameters to be estimated (69 more than the configural-weight model). Two versions were tried: In one version the added increment in weight was independent of w_{XBP}; in the other it was proportional to w_{XBP}. Scale adjustment was allowed, in accordance with Equation 9. Neither of these differential-weight models described the data of Figure 9 as well as the configural-weight model (Equation 10). The best-fitting differential-weight model can be written as follows:

$$w_{XBPE} = w_{XBP}(1 + \theta_{BPE}), \qquad (11)$$

where θ_{BPE} is the weight increment and w_{XBP} is defined as in Equation 10. The root mean squared deviations for the Expertise × Expertise design were 16.8, 16.9, and 14.3 for the buyer's, fair, and seller's price judgments, respectively, compared with 15.3, 10.6, and 12.9, respectively, for the modified configural-weight theory (Equation 10).

The differential-weight theory requires a relationship between the shape of the curves in Figure 9 and their approach to the scale values. The curves should "bow over" only as they asymptotically approach the scale values (Birnbaum, 1973; Riskey & Birnbaum, 1974). As can be seen from Figure 9, however, the curves bow over long before reaching the estimated scale values. For example, the highest point in Figure 9B was predicted to be $641 by the differential-weight averaging model, compared with $610 for the data and $608 for the prediction of the configural-weight model (Equation 10). This deviation from the averaging model is similar to that obtained by Birnbaum (1973) and by Riskey and Birnbaum (1974). Since the differential-weight model of Equation 11 uses nearly twice as many parameters yet fits the data of Figure 9 worse, configural-weight theory seems preferable.

Discussion

The purpose of this research is to examine theories of how judges (who may be biased) combine information from sources who vary in their ability to report the truth (expertise) and their motivation to distort it (bias). By representing these processes with mathematical models, it becomes possible to test experimentally among explicit theories of social judgment.

Experiments show that certain algebraic representations cannot account for the judgments. Additive and constant-weight averaging formulations have been tested and rejected in previous studies of information integration (T. Anderson & Birnbaum, 1976; Birnbaum,

1972, 1973, 1974, 1976; Birnbaum et al., 1976). The present data are in agreement with these previous studies in refuting additive and constant-weight averaging models in favor of some form of configural, relative-weight averaging model.

The present research goes beyond previous work to test three distinct forms of the relative-weight averaging model that attempt to explain the effects of the source's bias. In one, the effect of the source's expertise does not alter the effect of his or her bias; in another, expertise magnifies the effect of bias; in the third, expertise diminishes the effect of the source's bias. The data give a clear indication that the response-revision and weighted-bias models can be rejected. The retained model, scale-adjustment, which predicts that expertise will magnify the effects of bias, makes additional, independent predictions that are fulfilled by the data. By making these correct predictions, the status of the model as a representation of source effects is enhanced. The present research also extends the investigation to examine configural-weight theories of how the judge's point of view affects the weights and scale values of estimates from biased sources.

Source Bias and Expertise

The scale-adjustment model of Figure 2 (Model 2) correctly predicts that the source's expertise will amplify the effect of the source's bias, a prediction that was confirmed in all five experiments. The model does not imply that experts are necessarily *judged* to be more biased, only that the *effect* of bias will be greater for sources of greater expertise. It is as if the judge, hearing that the seller's friend provided an estimate of $500, thinks, "That means $470," before averaging the information, giving greater weight to the estimate provided by the source of greater expertise.

The weight of an estimate depends mostly on the source's expertise; however, weight also varies with bias. A consistent trend across all experiments is that the unbiased source (the independent) of high expertise tends to have greater weight than either biased source of the same expertise.

It is important to note that the effect of bias on weight and its effect on scale value can work in tandem or in opposition in any given instance. Perhaps this would explain the seeming contradictions in research on attitude change for the effects of source's trustworthiness, or bias (see McGuire, 1968). Bias is both a signed variable (plus or minus) in its effect on scale value and an *absolute* variable in its effect on weight. For example, a biased source (e.g., a scientist employed by a utility company using nuclear reactors) might produce larger attitude change in favor of his message than an unbiased source if he said, "Nuclear reactors are unsafe." On the other hand, the effect of this message on the impact of other messages (weight) might be less than that of an unbiased source of the same expertise who gave the same message.

The data are suggestive, but by no means conclusive, on two other weighting effects. First, a source may receive extra weight for making an estimate that would not be expected on the basis of his or her bias. Second, the weight of a biased source may be greater if the judge is of the same point of view as the source.

Judge's Point of View

The judge's point of view consistently affects the scale values and the value of the initial impression: The values tend to be lower for the buyer's point of view and higher for seller's point of view. Perhaps the buyer's and seller's price estimations reflect persuasive judgments, meant as the opening round for bargaining. Interestingly, the absolute value of the bias adjustments in the scale values are less when the source's bias matches the judge's point of view, as if the judge perceives a compatible source to be less biased.

The effects of point of view are best represented in Experiment 5, which used a within-subjects design. This experiment showed that the weight of the lower estimate is greater for the buyer's point of view and that the weight of the higher estimate is greater for the seller's point of view. The fair price judgments were intermediate, but showed the divergent interaction characteristic of the buyer's point of view. Similar divergence was also obtained for fair price judgments in Experiments 1 and

2 and by Birnbaum et al. (1976), who obtained fair price judgments. Perhaps undergraduates tend to identify with the consumer under the fair price condition.

A modified configural-weight theory (a simple extension of the range model) gives a good account of this change of weight across points of view. The configural-weight theory uses just one additional parameter, besides the basic structure of Model 2 of Figure 2, to represent the configural effect of point of view.

Comparison of Configural- and Differential-Weight Theories

In configural theories, stimulus parameters depend on the stimulus pattern. Hence, the *higher* or *lower* estimate receives extra weight in the seller's or buyer's point of view, respectively. On the other hand, differential-weight theories assume that weight depends on stimulus magnitude, a seemingly context-free theory. In actuality, differential-weight theory requires (implicitly) a contextual theory that assigns the weights as a function of the context of the entire distribution of weights and scale values presented. For the present study, $400 would be a low estimate, deserving of a large weight from the buyer's point of view.

In contrast, configural-weight theory would not declare a $400 estimate to be low, *except in comparison with the other estimate of the same car*—that is, except with respect to the *within-set context* (Birnbaum et al., 1971). Thus, $400 would be a low estimate if the other stimuli are greater, but it would be high if the other estimates were lower. It is the *relationships* among estimates of the same car that define the configural weights. Thus, configural-weight theory takes the implicit context theory of differential weighting and extends it explicitly to the immediate (within-set) context of stimuli presented on each trial.

Configural-weight theory can be represented by means of a fulcrum and balance analogue, as in Figure 2. The configural parameter, ω, represents the proportion of weight that, depending on the point of view, is taken from either the higher or lower stimulus and given to the other. Differential-weight theory can also be represented by the lever and fulcrum; however, each location on the lever has a different weight associated with it. In addition, differential weighting requires a complete remapping of weights to locations for each point of view, using many more parameters than the configural-weight models.

Differential weighting may be required for certain situations in which multidimensional social stimuli vary in both weight and scale value simultaneously (e.g., T. S. Anderson & Birnbaum, 1976). Differential weighting is an extremely powerful curve-fitting device, one that requires special experimental designs to test. When appropriate designs have been employed, differential-weight models have failed to account for the data (Birnbaum, 1973; Riskey & Birnbaum, 1974). Judgments of morality fail to show the compensatory effects predicted by nonconfigural averaging theories: Given that a person has committed a very bad deed, there appears to be no number of good deeds that will make the person's judged morality approach the same high asymptote as if the person had done only good deeds. Similarly, the data of Figure 9 suggest that from the buyer's point of view, if a low-expertise source says $300, no level of expertise of the $650 estimate will compensate to bring the judged value above $550. For the present interactions, the simple configural-weight models provide a more elegant, accurate, and theoretically appealing description of the interactions.

Concluding Comments

This research investigates information integration under conditions in which the relevant variables can be manipulated, and it attempts to discover principles that explain the results. The hope is that principles that apply in controlled experiments are characteristic of basic psychological processes that have applicability to a wide range of judgmental phenomena. Results of experiments in a large number of domains encourage the hope that a reasonably small set of premises may account for a large array of data. Previous research has shown that the same principles of source expertise can be applied to intuitive numerical predictions, ratings of likableness, and judgments of the value of used cars. We are currently investigating the generality of

the principles of bias discovered here to see if they account for judgments of probable guilt in simulated court cases.

The present research suggests that a simple algebra can account for the complex effects of bias in information integration. Judgments can be represented by the center of gravity of a lever. Information can be represented by weights placed at various locations along the lever. The location, or scale value, of the information depends on the source's communication. This location is adjusted to account for the source's bias and the judge's point of view. The weight of a source's communication depends mostly on the source's expertise, but diminishes if the source is biased. In addition, judges appear to increment the weight of the communication most consistent with their own bias, or point of view.

Given these rules of operation, the lever is an analog computer, or model, that will reproduce the judgments obtained in these experiments. When a model can give a qualitative account of a variety of predictions, its plausibility is increased. It may then be used as a measuring device to study the effects of other variables on its parameters. With this Archimedian lever, a fulcrum, and a place to stand, we hope to raise our understanding of human judgment.

References

Anderson, N. H. Integration theory and attitude change. *Psychological Review*, 1971, *78*, 171–206.

Anderson, T. S., & Birnbaum, M. H. Test of an additive model of social inference. *Journal of Personality and Social Psychology*, 1976, *33*, 655–662.

Birnbaum, M. H. Morality judgments: Tests of an averaging model. *Journal of Experimental Psychology*, 1972, *93*, 35–42.

Birnbaum, M. H. Morality judgment: Test of an averaging model with differential weights. *Journal of Experimental Psychology*, 1973, *99*, 395–399.

Birnbaum, M. H. The nonadditivity of personality impressions. *Journal of Experimental Psychology* 1974, *102*, 543–561. (Monograph)

Birnbaum, M. H. Intuitive numerical prediction. *American Journal of Psychology*, 1976, *89*, 417–429.

Birnbaum, M. H., Parducci, A., & Gifford, R. K. Contextual effects in information integration. *Journal of Experimental Psychology*, 1971, *88*, 149–157.

Birnbaum, M. H., & Veit, C. T. Scale convergence as a criterion for rescaling: Information integration with difference, ratio, and averaging tasks. *Perception & Psychophysics*, 1974, *15*, 7–15.

Birnbaum, M. H., Wong, R., & Wong, L. Combining information from sources that vary in credibility. *Memory & Cognition*, 1976, *4*, 330–336.

Chandler, J. P. STEPIT: Finds local minima of a smooth function of several parameters. *Behavioral Science*, 1969, *14*, 81–82.

Hovland, C. I., Janis, I. L., & Kelley, H. H. *Communication and persuasion*. New Haven, Conn.: Yale University Press, 1953.

Lichtenstein, S., Earle, T. C., & Slovic, P. Cue utilization in a numerical prediction task. *Journal of Experimental Psychology: Human Perception and Performance*, 1975, *1*, 77–85.

McGuire, W. J. The nature of attitudes and attitude change. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (Vol. 3). Reading, Mass.: Addison-Wesley, 1968.

Riskey, D. R., & Birnbaum, M. H. Compensatory effects in moral judgment: Two rights don't make up for a wrong. *Journal of Experimental Psychology*, 1974, *103*, 171–173.

Rosenbaum, M. E., & Levin, I. P. Impression formation as a function of source credibility and order of presentation of contradictory information. *Journal of Personality and Social Psychology*, 1968, *10*, 167–174.

Rosenbaum, M. E., & Levin, I. P. Impression formation as a function of source credibility and the polarity of information. *Journal of Personality and Social Psychology*, 1969, *12*, 34–37.

Wyer, R. S. *Cognitive organization and change: An information-processing approach*. Potomac, Md.: Erlbaum, 1974.

Appendix A

An alternative to all three of the models of source bias presented in Figure 2 is a differential-weight model:

$$R = \frac{w_0 s_0 + ws + w_{XBE} s_E}{w_0 + w + w_{XBE}},$$

in which the weight of an estimate, w_{XBE}, depends on expertise, bias, and estimate, but the scale value is independent of bias. This model assumes that a high estimate provided by the buyer's friend receives greater weight then if the seller's friend had made the same estimate. Consequently, the judged worth would be greater for the buyer's high estimate than for the seller's. Similarly, a low estimate is assumed to receive greater weight if it is provided by the seller's friend than if provided by the buyer's friend; consequently, judged worth would still be greater for the buyer's than the seller's estimate.

This differential-weight model cannot describe the present data. It makes three incorrect predictions for effects that are correctly predicted by the scale-adjustment model. First, it predicts that if w_0/w_{XBE} is near zero, the effects of *both* bias and expertise should be very small in the source-estimate designs, where blue book value is not presented, since $w_{XBE}/(w_0 + w_{XBE}) \approx 1$. In contrast, Equation 3 predicts that the Expertise × Estimate interaction should be small but the effect of bias should be maximal, since if $w_0/w_X \approx 0$ for all w_X, then

$$\frac{w_X(s_E + b_B)}{w_0 + w_X} \approx s_E + b_B.$$

The data for the source-estimate designs (open circles in Figure 5) show a small Expertise × Estimate interaction and a large effect of bias. The average effect of bias is \$76.6 in the source-estimate design of the fair price condition, compared with \$46.6 when blue book value is also presented. The differential-weight model incorrectly predicts that the effect of bias should have been larger when blue book value is presented.

Second, the differential-weight model predicts that there exists some level of estimate for which bias has no effect, contrary to the data. Third, the differential-weight model predicts a complex four-way interaction between expertise, bias, estimate, and blue book value that did not materialize in the predicted form. Consequently, differential weighting alone cannot explain the effect of bias—scale adjustment appears to be necessary.

Appendix B

It seems intuitively reasonable that if a source gives an estimate that would not be expected on the basis of his bias and the other estimate, he might receive greater weight. For example, if a *friend of the seller* gave an estimate below the blue book value, his estimate might receive greater weight.

To investigate this hypothesis, the data of Experiment 5 were fit using the following equation:

$$w_{XBPE} = w_{XBP} + \omega_P \sigma_E \rho_{PE} + \alpha_{PB} \beta_B \sigma_E \delta_{PBE},$$

where w_{XBPE} is the absolute weight of an estimate, E, by a source of expertise, X, and bias, B, from a point of view, P; w_{XBP} is the configural free weight, as in Equation 10; α_{PB} is the estimated configural-weight parameter that expresses the magnitude of the expectancy-contrast effect; ω_P, σ_E, and ρ_{PE} are defined as in Equation 10; $\beta_B = -1$ if the source is a seller, $\beta_B = 0$ if the source is an independent, and $\beta_B = 1$ if the source is a buyer; $\delta_{PBE} = w_{XBE}$ if $\alpha_{PB}\beta_B\sigma_E \leq 0$, and otherwise δ_{PBE} equals the configural free weight of the other source.

The product $\beta_B \sigma_E$ will be positive when either (a) a buyer provides the higher estimate or (b) a seller provides the lower estimate—instances in which the source seems to deserve extra weight for doing the unexpected. It will be negative when either a buyer provides the lower estimate or a seller provides the higher (i.e., when the source's estimate is expected). Notice that δ_{PBE} and ρ_{PE} make a decrease in weight proportional to the weight to be decreased.

Figure 7A, Panel IB shows that when the buyer gives an estimate of \$300, the effect of the independent's estimate is greater than predicted, as if the buyer lost weight. Figure 7A, Panel IS, shows that when the seller says \$300, the effect of the independent is decreased relative to the predictions of the range model.

For the buyer's point of view, the α_{PB} weights for the buyer and seller are −.08 and .20, respectively; for the fair price point of view, the values are close to zero, −.04 and −.01, respectively; and for the seller's point of view, they are .05 and .11, respectively. Thus, it appears that for either biased point

of view, the seller receives extra weight for making the lower estimate. The friend of the buyer receives additional weight for the seller's point of view when he provides the higher (unexpected) estimate; however, for the buyer's point of view, the buyer receives greater weight for providing the lower estimate. The near-zero values for the fair price condition may reflect the fact that the fit of Equation 10 was already quite good in that condition (see Figure 7B).

The expectancy weight modification reduces the overall sum-of-squares discrepancies from 48,333 to 44,267, an improvement of 8.4% over Equation 10. The root mean squared deviation was 9.57. Improvements were greatest for the Expertise × Expertise, Bias × Bias, and Expertise × Bias designs for the buyer's point of view.

Received November 7, 1977

Journal of Personality and Social Psychology
1979, Vol. 37, No. 1, 75-86

Self-Consciousness, Self-Attention, and Social Interaction

Allan Fenigstein
Kenyon College

Two experiments were conducted to test the effects of self-focused attention on positive and negative social interactions. In the first study, the behavior of dispositionally high and low publicly self-conscious women was examined in an interpersonal situation involving rejection by a group. It was hypothesized that persons high in self-consciousness, being more aware of how they are perceived by others, would be more sensitive and react more negatively to the rejection than those low in self-consciousness. The predictions were confirmed. In Experiment 2, female subjects were presented with favorable or unfavorable feedback in the context of an interview, and self-attention was experimentally manipulated by exposing half the subjects to their images in a mirror. Self-awareness increased the negative response to the negative evaluation and tended to increase the positivity of the positive evaluation. The implications of self-awareness theory for the social self and social interaction are discussed.

Everyday observations as well as theoretical approaches to social behavior suggest that in the presence of others, one is apt to become self-conscious, that is, aware of the self as a social object that can be observed and evaluated by others. Goffman (1959) has argued quite persuasively that when one is attending to and involved in an ongoing interaction, that interaction can proceed smoothly and naturally; but if one is engaged in self-focused thought during that interaction, then concern is shifted away from what is being said toward whether what one says will be received favorably or unfavorably.

Argyle (1969) also proposes that self-consciousness, conceived of as the activation of the "self-system," produces a decreased concern with evaluating the behavior of others and an increased concern with the personal and public assessment of one's own behavior. It follows that when the self-system remains dormant, relatively less thought is given to one's own behavior or its effects on others. Similarly, Duval and Wicklund (1972) have experimentally demonstrated that a state of self-focused attention causes one to engage in self-examination and self-evaluation.

Underlying all of these approaches to self-consciousness are several common, but crucial, assumptions: (a) During a social encounter, attention may be directed either toward or away from the self. (b) When attention is self-directed, the person becomes *conscious* of the *self* as an object of attention to others; conversely, when attention is directed away from the self toward external stimuli, there is little consciousness of the self as a social object. And (c) a major consequence of self-consciousness is an increased concern with the presentation of self and the reactions of others to that presentation.

Determinants of Self-Attention

Self-focused attention has several sources, both situational and dispositional. It can be brought about through the presence of others, whereby the person becomes cognizant of the

This article is partially based on a doctoral dissertation submitted to the Graduate School of the University of Texas at Austin. The manuscript was written while the author was on leave at the University of Virginia. The author is grateful to Arnold Buss for his support and advice through all phases of the research and to the reviewers for their helpful comments on an earlier draft. Portions of this article were presented at the annual convention of the American Psychological Association, San Francisco, August 1977.

Requests for reprints should be sent to Allan Fenigstein, Department of Psychology, Kenyon College, Gambier, Ohio 43022.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3701-0075$00.75

fact that others perceive him or her as an object (Carver & Scheier, 1978); or it may be the result of a reflective stimulus, such as a mirror or a tape recording of one's own voice, which is a reminder of one's status as an object (Argyle, 1969; Duval & Wicklund, 1972).

In addition to situationally induced self-attention, there may also be individual differences in the degree to which self-consciousness occurs. Argyle (1969) has theorized that persons differ in the extent to which the self-system becomes salient in the presence of others, and more recently, Fenigstein, Scheier, and Buss (1975) constructed and extensively administered a scale to measure *self-consciousness*, defined as the enduring tendency of persons to direct attention toward themselves. Factor analyses of the scale consistently yield two stable self-consciousness dimensions: *public* and *private*. The private factor is defined by an awareness of one's personal thoughts and feelings (e.g., "I'm always trying to figure myself out"). The public subscale involves an awareness of the self as a social object, that is, an awareness that others are aware of the self (e.g., "I'm concerned about what other people think of me"). In addition, a social anxiety factor emerges. Social anxiety, which is defined as discomfort in the presence of others (e.g., "I get embarrassed very easily") may be seen as a reaction to the process of self-focused attention. Previous research involving several independent samples has established that the two self-consciousness factors are only weakly correlated (Fenigstein et al., 1975; Scheier, 1976) and that the scale as a whole has considerable discriminant validity (Carver & Glass, 1976; Scheier & Carver, 1977).

Public Self-Consciousness and Social Behavior

The first study reported here had two purposes: One was to examine the effects of public self-consciousness; a second aim was to extend the study of self-focused attention into the area of social interaction. Although self-awareness has been shown to have consequences for many realms of behavior, virtually all of these behaviors have been nonsocial, involving little interaction with others. For example, heightened self-awareness has been found to facilitate task performance, increase the consistency between attitudes and behavior, decrease self-esteem, increase self-attribution (see Wicklund, 1975), and heighten the experience of emotion (Scheier & Carver, 1977)—none of these behaviors is interpersonal. Even in those studies which do involve social behavior, there is no sense of any ongoing, face-to-face interaction. In research concerning the effects of self-awareness on conformity, the reference group's opinion is simply represented on paper (e.g., Duval, 1976); and in the application of self-awareness theory to aggression, the aggressor remains essentially isolated from the target during the aggressive encounter (e.g., Scheier, Fenigstein, & Buss, 1974; Carver, 1974). It may be concluded that little is known about the consequences of self-focused attention on interpersonal behavior.

The present research focused on the role of public self-consciousness in social interaction. Publicly self-conscious persons are presumed to be susceptible to feelings of being "observed" when in the company of others (Argyle & Williams, 1969; Fenigstein et al., 1975). Believing that others are preoccupied with their appearance and behavior, these persons are always in position to infer how others perceive them and are highly sensitive to the reactions of others in a social situation.[1] Extreme examples of public self-consciousness

---

[1] The present construct of public self-consciousness should be differentiated from a seemingly related dimension called self-monitoring (Snyder, 1974). The self-monitor is also sensitive to the social behavior of others, but primarily for the purposes of monitoring and manipulating his or her own behavior so as to present an appropriate image. Although the self-monitor must obviously attend to his or her own behavior, the orientation is primarily outward toward situational requirements, as opposed to the self-conscious person, whose attention is primarily self-directed.

It may be argued that self-focused attention is related to self-monitoring: A person must observe his or her own behavior in order to create a desired impression (Goffman, 1959; Mead, 1934). But self-attention may also lead to feelings of "self-consciousness" in the popular sense of the term, that is, anxiety, discomfort, and embarrassment, which effectively disrupt, rather than facilitate, any attempt at impression management. The point is that the publicly self-conscious person may be a self-monitor or may not be. Recent investigations by the

are easily recognized. At the high end is the recently stigmatized person who, almost by definition, is an object of attention and is sensitive to the concern, disgust, or pity that is elicited from others. At the opposite end is the totally unself-conscious social boor who not only lacks any conception of how he or she appears to others but could not care less.

These speculations about the behavior of the publicly self-conscious person were tested using a relatively ubiquitous situation in social interaction: the experience of being shunned by or excluded from a group. Rejection by others may be extremely aversive. As an implicit negative evaluation, rejection has been shown to produce feelings of anxiety, embarrassment, and worthlessness (Geller, Goodstein, Silver, & Sternberg, 1974) and to decrease attraction toward the group (Dittes & Kelley, 1956; Mettee, Fisher, & Taylor, 1971). However, other studies have found that rejection does not alter liking for the group (Snoek, 1962) and may even increase one's evaluation of the others (Jackson & Saltzstein, 1959). Apparently, some persons are capable of shrugging off social rejection and seem oblivious to its effects, whereas others are very hurt by such experiences.

One factor that may help to account for these differences in reactions to rejection is the degree of awareness of the public or social self (i.e., the degree to which persons recognize and are concerned about the way they are perceived by others). For persons low in public self-consciousness, there is little awareness of how others regard them. Attention is not directed toward the self—presumably, it is consumed by the group activity or other external events—and the group's behavior in relation to the self is not relevant. Thus, the experience of being excluded from the interaction is not very salient. But highly self-conscious persons are sensitive to the issue of how they are regarded by others. For them, the group's behavior is perceived as having personal relevance, and the experience of rejection should have a clear impact on subsequent behavior.

Public self-consciousness should also affect causal explanations of the rejection. A person may make either internal attributions and assume personal responsibility for the rejection or external attributions and assume that the cause of the rejection lies with the other persons or within the situation (Pepitone & Wilpizeski, 1960). Much of the recent attribution literature has shown that causality for an event is likely to be attributed to those salient entities that attract attention and can reasonably be seen as a cause (e.g., Duval & Wicklund, 1973; Pryor & Kriss, 1977; Storms, 1973; Taylor & Fiske, 1975). Given that the social self is salient to publicly self-conscious persons and a primary target of their attention, it follows that high publicly self-conscious persons are more likely than low publicly self-conscious persons to perceive themselves as the cause of the rejection.

In a recent study, Buss and Scheier (1976) found that private self-consciousness increased the amount of causality attributed to the self, but public self-consciousness had no effect on self-attributions. Those authors reasoned that attention to one's own private thoughts and feelings is an important determinant of self-attributions, whereas focus on the self as a social object is irrelevant to such attributions. However, their research involved the use of hypothetical situations into which subjects imaginally projected themselves, and under these circumstances, it is reasonable that only private self-consciousness affected causal attributions. The present experiment involves a real social interaction, with attributions being made about behaviors that have significant social consequences, and in this case, public self-consciousness should play an important role.

In summary, the first study hypothesized that persons high in public self-consciousness would be more sensitive and react more negatively to rejection by a group, and would hold themselves responsible for the rejection to a greater degree, than those low in public self-consciousness. The major dependent variables were attraction toward and desire to affiliate

---

author have found a correlation of .24 ($N = 380$) between public self-consciousness and self-monitoring; similar results were obtained by Scheier and Carver (1977). Thus, it is unlikely that the effects of public self-consciousness can be accounted for by self-monitoring.

with the group, and causal attributions for the group's behavior.

Experiment 1

Method

The study was a 2 × 2 factorial design with one manipulated variable (rejection vs. acceptance) and one subject variable (high vs. low public self-consciousness).

Subjects

Several weeks prior to the experimental manipulation, during a period designated for this purpose, all students from courses in introductory psychology filled out a number of questionnaires. The crucial items consisted of the Self-Consciousness Scale (Fenigstein et al., 1975) and three generalized self-esteem items (e.g., "In general, I feel very good about myself").[2] Measures of self-esteem were included because previous research (Dittes, 1959; Jacobs, Berscheid, & Walster, 1971; Jones, Knurek, & Regan, 1973) suggested that it may affect reactions to rejection.

Subjects scoring in the upper and lower thirds of the distribution of public self-consciousness scores were selected. A total of 92 women participated in the experiment. Women were used because pilot studies indicated that as a group, they were more sensitive to rejection than men. Twelve subjects were dropped: eight because of suspicion expressed during the postexperimental interview and four because of failure to follow instructions. Dropouts were unrelated to experimental conditions. This left 80 subjects, 20 in each condition.

Procedure

All the experimental manipulations occurred while the subject was waiting to participate in the supposed "real" experiment. The subject arrived first at the experimental room, which was bare except for a table and chairs, and was told that several other subjects were expected and that the experiment would begin when they arrived. All subsequent participants were female confederates of the experimenter.

Before entering the "waiting room," all subjects were instructed to leave their belongings outside the room so that the experiment could begin promptly as soon as everyone had arrived. (This was actually done to prevent subjects from directing their attention and behavior toward easily available and irrelevant objects such as books or purses.)

Upon entering the room, the subject was subtly directed toward one end of a rectangular table that had one side against the wall and a chair at each end. About one minute later, the first confederate arrived and was given the same introduction as the subject. The experimenter brought another chair into the experimental room, placed it halfway between the two ends of the table, and then left. The first confederate took this seat and remained quiet. If the subject attempted to initiate a conversation, the confederate made as brief and neutral a response as possible (e.g., "yes," "no," or "I don't know").

Rejection. Approximately 30 sec later, the second confederate arrived and was told that one more subject was expected. After she took the remaining chair opposite the subject and the experimenter had left, both confederates remained quiet for about 20 sec (to establish the lack of any acquaintanceship between them). The second confederate then began a casual conversation with the first confederate, who was reticent at first but slowly became more talkative.

The first confederate played the role of a naive freshman who was living in one of the campus dorms, and the second confederate appeared to be a somewhat more experienced upper-division student who was living off campus. The conversation eventually turned to the restrictiveness of dormitory living. A recent on-campus incident concerning a resident assistant who was actually an undercover narcotics informer served as the focal point of the discussion. The younger confederate took the more conservative position, defending the school's obligation and responsibility to protect its students from the potential dangers of drugs. The other student attacked the school's intrusiveness into the private lives of students and the general authoritarian leanings of the university administration.

During the conversation, both confederates directed their comments only toward each other and ignored the subject. If the subject tried to speak, the older confederate made a brief, unenthusiastic remark such as "I see." Eye contact with the subject was minimal. The impression created by the confederates was not dislike or hostility but simply lack of interest.

Acceptance. Subjects in this group were encouraged to participate in the conversation. If the subject did not do so spontaneously, her opinions were solicited by the second confederate. Any remark uttered by the subject was attended to with interest and responded to appropriately.

Questionnaire. Five minutes after the second confederate began the conversation, the experimenter reentered the room, remarking that the "fourth subject" apparently was not coming. He went on to explain that the experiment was one of a series of studies, and all subjects would be reassigned to another part of the experiment. To facilitate the procedure, subjects were asked to provide some preliminary information based on their brief experience with each other. Questionnaires containing manipulation checks and the dependent variables were then distributed. After subjects had completed the questionnaire, they were asked a series of questions designed to determine suspicion, and then were debriefed.

Results

Covariation

Self-esteem did not correlate with the public self-consciousness variable, nor did it af-

[2] All items are available from the author upon request.

fect any of the dependent variables; thus, it was not included in the analyses. Modest correlations did exist between public self-consciousness and measures of private self-consciousness ($r = .17$) and social anxiety ($r = .24$). To measure the independent effects of public self-consciousness, 2(high vs. low public self-consciousness) × 2 (rejection vs. acceptance) analyses of covariance were used, with private self-consciousness and social anxiety as the covariates.

Manipulation Checks

Group's behavior. Subjects were asked to rate on a 7-point scale how favorably the others, as a group, behaved toward them. Rejected subjects rated the group's behavior as less favorable ($M = 3.8$) than subjects who were not rejected ($M = 5.4$). This difference was significant, $F(1, 74) = 66.4$, $p < .001$, thus establishing the adequacy of the rejection manipulation. No other effects were significant. Additional support for the success of the rejection manipulation is provided below in other analyses.

Self-Consciousness. The crucial conceptual difference between high and low publicly self-conscious persons lies in the amount of attention they focus on themselves during a social interaction. An attempt was made to assess this difference directly by asking subjects to indicate the relative percentage of time that they spent thinking about themselves during the experiment. Results showed that high publicly self-conscious subjects reported spending more time attending to themselves than low publicly self-conscious subjects: 24% vs. 9%, on the average—a highly significant difference, $F(1, 74) = 24.9$, $p < .001$. The data also suggest that the experience of rejection by others may focus attention on oneself: There was a slight tendency for rejected subjects to engage in more self-attention ($M = 18.5\%$) than accepted subjects ($M = 15\%$), but this difference was marginally significant, $F(1, 74) = 2.6$, $p < .1$.

Table 1
Number of Subjects Choosing to Affiliate with the Same Group as a Function of Treatment by the Group and Subjects' Level of Public Self-Consciousness

Self-Consciousness	Treatment	
	Acceptance	Rejection
Low	14	10
High	15	3

Note. $n = 20$ subjects per cell.

Table 2
Mean Liking for the Group as a Function of Group Treatment and Subjects' Level of Public Self-Consciousness

Self-Consciousness	Treatment	
	Acceptance	Rejection
Low	5.5	4.4
High	5.6	3.4

Note. Ratings were on a 7-point scale on which 1 corresponded to "did not like the group at all" and 7 to "liked the group very much."

Dependent Variables

Affiliation. Subjects were asked whether they wanted to continue to work with the same group or with a different group in subsequent parts of the study. Table 1 provides a frequency analysis of subjects' choices and shows that rejected subjects were less willing to continue to affiliate with the same group than nonrejected subjects, $\chi^2(1) = 12.8$, $p < .001$. Within the acceptance condition, public self-consciousness had no effect on affiliative behavior. But following rejection, high publicly self-conscious subjects chose to affiliate with the same group far less frequently than low self-conscious subjects, $\chi^2(1) = 5.6$, $p < .02$.

Attraction. Liking for the group, indicated on a 7-point scale, was considerably less in the rejection condition than in the acceptance condition (see Table 2). Within the rejection condition, public self-consciousness had a clear effect on liking: High self-conscious subjects were less attracted toward the group than low self-conscious subjects. The significance of these effects was revealed in a 2 × 2 covariance analysis that yielded a main effect for rejection, $F(1, 74) = 43.2$, $p < .001$, and a main effect for public self-consciousness, $F(1, 74) = 4.0$, $p < .05$, due primarily to the differences between the rejected subjects. This result was confirmed by a significant Public

Table 3
Mean Percentage of Self-Attributions for the Group's Behavior as a Function of Treatment by the Group and Subjects' Level of Public Self-Consciousness

Self-Consciousness	Treatment	
	Acceptance	Rejection
Low	38.7	39.2
High	55.7	49.7

Self-Consciousness × Rejection interaction, $F(1, 74) = 5.1$, $p < .05$; planned contrast indicated that public self-consciousness had no effect in the acceptance condition, but in the rejection conditions, high self-conscious subjects reacted more negatively toward the group than low self-conscious subjects $t(38) = 2.97$, $p < .004$.

Causal attributions. Subjects were asked to distribute responsibility for the way the group behaved toward them to either themselves, the others as a group, or the situation; the amount of responsibility to be attributed totalled 100%. Table 3 shows that the high, compared to the low, publicly self-conscious subjects made more causal attributions to themselves for the group's behavior across both rejection and control conditions. This finding was confirmed by a 2 × 2 covariance analysis that showed a main effect for self-consciousness, $F(1, 74) = 9.6$, $p < .01$. Causal attributions were also affected by social anxiety, $F(1, 74) = 8.3$, $p < .01$; those who became anxious in the presence of the group were more likely to feel responsible for the group's behavior than low anxious subjects. No other effects were significant.

Discussion

Experimental Findings

Rejection. Rejection by others, not surprisingly, had a strong effect on subsequent behavior. Rejected subjects perceived the group as less cooperative, liked the group less, and were less likely to continue affiliation with the group than accepted subjects. In addition, rejected subjects were somewhat more likely to focus attention on themselves during the interaction than nonrejected subjects.

For ethical reasons, rejection was manipulated in a deliberately passive manner: Other group members failed to show an interest in the subject. The obtained results suggest that interpersonal rejection can be effectively manipulated and studied in the laboratory without the use of any explicit derogation or overt hostility toward the target.

Public self-consciousness. The study found that individual differences in public self-consciousness had strong and consistent effects on reactions to an interpersonal rejection. After being shunned by a peer group, high publicly self-conscious women were less attracted to the group and were less likely to continue affiliating with that group than women low in public self-consciousness. Presumably, persons high in public self-consciousness were more aware of how they were seen by others. Thus, they were more sensitive to the shunning and were more disposed to avoid the people who ignored them.

It was assumed that private self-consciousness would have no effect on reactions to rejection. The focus of the experimental situation was on how subjects were perceived by others, that is, on their social self, and private, covert aspects of the self were considerably less salient. Covariance analyses confirmed that the role of private self-consciousness was negligible in this situation. Thus, the present study provides further evidence of the discriminant validity of the two self-consciousness dimensions (see also Scheier & Carver, 1977).

It is important to note that public self-consciousness had no effect on perceptions of the group's behavior. The "rejection group" was seen as equally unfriendly by both low and high self-conscious persons. Their different reactions were presumably a result of differing sensitivities to or differing interpretations of the group's behavior. That self-consciousness affects one's interpretation of an event is suggested by the finding that publicly self-conscious persons tended to see themselves as being largely responsible for the group's behavior. Both Duval and Wicklund (1973) and Taylor and Fiske (1975) postulate that attentional mechanisms are major determinants of causal attributions, that is, causes are of-

ten attributed to the most salient source of information, and salience is highly dependent on attention. Thus, the high publicly self-conscious person, because she or he directs attention to herself or himself (even while observing others), is likely to perceive a causal relationship between the self and the behavior of others. Events perceived as unrelated to the person have little impact on that person. However, if those same events are believed to be personally relevant, they should have a strong effect on behavior (Jones & Davis, 1965). The publicly self-conscious person presumably sees the rejection as being uniquely focused on or caused by her- or himself, and this self-blame for the aversive event may add to the negativity of the reaction (e.g., Abramson & Sackheim, 1977; Kuiper, 1978).

Acceptance

It may be argued that just as self-focused attention during an unpleasant interaction increases one's negative response, self-focused attention should also heighten the positivity of a positive interaction. Scheier and Carver (1977), for example, found that self-awareness increased responsivity to both positive and negative affective states in a nonsocial context. Although the acceptance condition of the present study may clearly be considered a pleasant encounter, reactions to this event were not altered by public self-consciousness. One possible explanation may be that persons high in public self-consciousness are especially sensitive to rejection but are not responsive to acceptance.

A related explanation for the failure of public self-consciousness to affect reactions to acceptance lies in the nature of the acceptance manipulation. Acceptance is neither an unusual nor an impactful occurrence, and it is unlikely that acceptance generates as much affect as rejection. As William James (1890) noted in his discussion of the social self, we all have the need to be noticed favorably, and there is "no more fiendish punishment" then to be ignored and have others act as if we were nonexistent. It may simply be that rejection and acceptance are not comparably balanced instances of positive and negative social interactions. Pilot observations and postexperimental interviews offer some support for this position. Rejected subjects displayed and reported a great deal of emotional discomfort, but very little emotion was shown by accepted subjects. In addition, accepted subjects were actively engaged in conversation: listening to and processing information, formulating and expressing opinions. To the extent that these behaviors were occurring, self-attention would be reduced and its effects mitigated (Duval & Wicklund, 1972). Rejected subjects, on the other hand, were much less involved in the ongoing interaction, and under these conditions, the effects of self-consciousness may have been considerably stronger.

Experiment 2

Experiment 2 was designed to provide a more unequivocal test of the effects of self-focused attention on *both* positive and negative interpersonal feedback. It was hypothesized that under conditions of increased self-attention, when persons become more aware of and concerned with self-evaluation, they also become more closely attuned to the self-relevant evaluations of others. Under these conditions, interpersonal feedback should be more salient than when attention is directed away from the self and self-evaluation is of little concern.[3] An attempt was made in this study to control (a) the involvement of subjects in the interaction and (b) the relative valences of the positive and negative feedback. Also, in order to enhance the plausibility and generalizability of a self-attention interpretation, self-awareness was experimentally manipulated in this study through the presence or absence of a mirror. Mirrors have been used extensively to induce self-awareness (e.g., Duval & Wicklund, 1973; Scheier et al., 1974), and there is now strong evidence that mirrors provide a valid means of increas-

[3] Theories of *deindividuation* lead to similar predictions. For example, Zimbardo (1969) argues that under conditions of deindividuation, a state in which one's sense of identity is lost, there is a decreased concern with self-evaluation or evaluation by others. Although it is tempting to argue that deindividuation represents the extreme low end of the self-awareness dimension, there is little research on the issue, and the question remains unresolved.

ing the occurrence of self-relevant thoughts (Carver & Scheier, 1978; Geller & Shaver, 1976).

Evaluative descriptions of the subject were provided in the context of an interview during which active participation by the subject was minimal. Subjects were presented with a generalized, impersonal interview that was either favorable or unfavorable.[4] Content was controlled so as to insure that positive and negative evaluations were of equal magnitude. Half the subjects were confronted with an unavoidable mirror image throughout the interview. This generated a 2 × 2 factorial design: favorable and unfavorable interview content crossed with mirror or no mirror. The major dependent variables were affective responses to the interview and the interviewer.

Method

Subjects

A total of 52 female undergraduates, all having siblings, were recruited from courses in introductory psychology. Four subjects were dropped from the experiment because they expressed suspicion during the postexperimental interview. The remaining 48 subjects were randomly assigned to the four experimental conditions: 10 each in the two positive feedback conditions and 14 in each of the two negative feedback conditions. (As explained below, these ns were unequal because the valence of the feedback could not be determined until after the experiment was over.)

Procedure

Subjects were brought into the experimental room and were told that they were going to take part in an interview. The purported purpose of the interview was to get the subject's "intuitive reactions to recent experimental findings concerning the effects of birth order on personality." These findings were described as either favorable or unfavorable to persons of the birth order of the subject. For example, if the interview was favorable to firstborns, then the subject would be in the positive feedback condition if she were a firstborn, but would be in the negative feedback condition if she were a later-born. Thus, the *same* interview content could serve as *either* pleasant or unpleasant feedback, with the subject's experimental condition determined by her birth order. The interview lasted about 10 min, during which time the female interviewer established brief eye contact with the subject only four times—once in the introduction, twice during the interview, and once at the conclusion. This was intended to minimize the induction of self-awareness through eye contact (e.g., Scheier et al., 1974). For half the subjects, there was a mirror directly behind the interviewer; for the other subjects, the mirror was turned around so that its reflective side faced the wall. Following the interview, subjects filled out a questionnaire evaluating the interviewer and the situation. Finally, the subject was debriefed.

Favorability manipulation. Essentially the same interview was given to all subjects, except that the terms *firstborn* and *later-born* were interchanged. After asking the subject some preliminary background information, the interviewer reviewed the "findings" on the relation of birth order to 11 different personality characteristics including warmth, assertiveness, independence, achievement, and intelligence. For half the subjects, firstborns were credited with the favorable aspects of these traits, and later-borns were given the unfavorable end. For the remaining subjects, this was reversed. Subjects were instructed not to reveal their birth-order position; thus, the experimenters were successfully blinded as to the subject's favorability condition.

Mirror manipulation. The mirror measured 3 feet × 2 feet (.91 m × .61 m) and was located on the wall directly behind and above the interviewer. The tilt of the mirror insured that the subject was exposed to her own image.

Dependent variables. At the conclusion of the interview, the subject filled out a questionnaire in another room, where a second experimenter was waiting. The questionnaire's ostensible purpose was to get information concerning the interviewer's strengths and weaknesses as seen by the subject. The questionnaire consisted of five 7-point scales concerning the interviewer and three 7-point scales about the interview. The major dependent variables were the subject's ratings of how pleasant and likable the interviewer was and how interesting and pleasant the interview was. Other scales dealt with the sincerity, competence, and poise of the interviewer and with how "personal" the situation was. In addition, subjects were asked to indicate on a 7-point scale how anxious they felt during the interview.

Results and Discussion

Affective Responses

The affective responses toward the interviewer and the interview were highly consistent (with correlations ranging from .52 to .73), and so they were combined into one overall affective rating (see Table 4). Again, as expected, positive feedback led to more

[4] The procedure and interview were originally developed by Ellsworth and Carlsmith (1968), and the reader is directed to that study for further details.

[5] Measures of attribution concerning the "cause" of the evaluative feedback were not considered relevant here because the feedback was presumably derived from "scientific findings" rather than from the principals in the experiment.

positive feelings then negative feedback, $F(1, 44) = 6.10$, $p < .02$, and there was no main effect for the mirror. The interaction was highly significant, $F(1, 44) = 7.73$, $p < .001$, indicating that the effects of the verbal evaluations were strongly influenced by the presence or absence of a mirror. Planned contrasts revealed that when the interview was critical, subjects responded more negatively in the presence of the mirror than in its absence, $t(26) = 2.98$, $p < .005$; following positive feedback, there was a nonsignificant tendency toward greater positive affect in the mirror condition.

Other Interview Measures

Concerning the ratings for the interviewer's competence, sincerity, and poise, and for how personal the situation was, there were no significant effects. From these results, it can be inferred that the interviewer's presentation was consistent across conditions, and the experimental manipulations had no effect on the subject's objective appraisal of the interviewer's capabilities. Finally, there were no significant differences between the ratings of firstborns and those of later-borns.

Anxiety

Subjects in the negative feedback group ($M = 3.0$) were more anxious than the positive feedback subjects ($M = 2.0$), $F(1, 44) = 6.40$, $p < .02$. The mirror by itself did not evoke anxiety, and the interaction was not significant.

Table 4
Mean Overall Affective Response Toward the Interview and Interviewer for Each Treatment Group of Experiment 2

Stimulus	Positive evaluation	Negative evaluation
No mirror	5.1	5.2
Mirror	5.6	4.1

Note. Ratings were on a 7-point scale on which 1 corresponded to a highly unfavorable response and 7 to a highly favorable one.

Eye Contact

The interviewer reported that subjects reciprocated over 92% of her gazes across all conditions. It could be argued that the mirror merely provided an easily available distraction to avoid a gaze-averting interview. However, these data suggest that the interviewer's eye contact was consistently met and cast serious doubt over any attempt to explain the effects of the mirror as due to its "distraction" qualities.

Discussion

It was hypothesized that evaluations from others would have a greater effect on subjects in the presence of a mirror than in its absence. These predictions were partially confirmed: The mirror substantially enhanced negative reactions to negative interpersonal feedback; the effect for positive feedback was in the expected direction but was not significant. Overall, the results suggest that self-attention intensifies the process of self-evaluation, and when self-awareness is low, the importance of self-relevant information from others is reduced.

It may be argued that the mere absence of a mirror is not sufficient reason to assume a low state of self-awareness. But, as in the present experiment, if the mirror's absence is combined not only with an interviewer who establishes a minimal amount of eye contact but also with a highly impersonal interview about birth order (in general, not specifically the subject's own), then there may be sufficient justification to argue for considerably diminished self-awareness in the no-mirror condition.

General Discussion

Empirical Findings

It has been widely assumed by self-consciousness theorists (e.g., Argyle, 1969; Goffman, 1959; James, 1890) that when attention is directed toward the self, there is an increased concern with how one is perceived by others; and when self-attention is low, feedback from others is of little importance.

The present research provides confirmation of these hypotheses. The results of both studies together demonstrate that when persons become more aware of the self during social interactions, because of either chronic dispositions toward self-consciousness or the reflection of their images in a mirror, there is an increased responsiveness to the evaluations of others (especially in the case of unpleasant feedback).

Although the present research has been interpreted in terms of self-attention processes, it may be argued that both public self-consciousness and the mirror produce an increase in *arousal* and a facilitation of dominant responses (e.g., Liebling, Seiling, & Shaver, 1974). In the present context, dislike and liking could be construed as the dominant responses to disparaging and complimentary feedback, respectively. This alternative explanation of the results is weakened by several considerations. The mirror in the second study had no consistent effect on subjects' self-reported anxiety. The lack of an association between the mirror and arousal is corroborated by physiological research showing that mirrors do not increase palmar sweating (Paulus, Annis, & Risner, Note 1). In the first study, covariance analyses established that the effects of public self-consciousness were independent of the effects of social anxiety. Thus, although high publicly self-conscious subjects may be more anxious in the presence of others than subjects low in public self-consciousness, this anxiety apparently does not account for their reactions to rejection. In addition, Carver and Glass (1976) found public self-consciousness to be uncorrelated with test anxiety and only weakly correlated with emotionality. There appears to be little evidence that arousal played a major role in either study. On the other hand, the fact that both a stable self-consciousness attribute (in Experiment 1) and a self-reflective stimulus (in Experiment 2) produced similar results lends support to an attentional explanation.

Direction of Causality

The convergence between the findings of the two studies also allows a question of causal interpretation to be resolved. Since public self-consciousness in Experiment 1 was a nonmanipulated individual difference, its causal relationship to reactions to rejection cannot be clearly established. It had been assumed that being aware of oneself as an object of attention to others causes rejection to be especially painful. However, it may be that vulnerability to rejection both causes persons to become chronically self-conscious and increases their sensitivity to rejection.[4]

Although this latter interpretation is a plausible account of the results of the first study and cannot be ruled out completely, it seems unlikely that a specific concern, such as fear of rejection, would induce a generalized state of public self-consciousness. Moreover, fear of rejection should be reflected in a measure of social anxiety, and as already indicated, there is only a modest correlation between social anxiety and public self-consciousness. Most importantly, the results of Experiment 2, in which self-attention was experimentally induced, clearly demonstrate that heightened sensitivity to social evaluation is a function of self-attention. Given the similarities in paradigm for the two studies, it seems reasonable to infer that an analogous causal sequence was operating in Experiment 1.

Positive and Negative Social Interactions

Experiment 1 found that publicly self-conscious women were highly vulnerable to rejection but were not responsive to acceptance. However, because it was difficult to determine the relative negativity of rejection compared to the positivity of acceptance in that study, a second experiment was conducted in which the valences of positive and negative social interactions were more comparable. Again, the effects of self-attention were stronger for unfavorable feedback than for favorable feedback. Although further research is warranted, the consistency of the present findings suggests a negativity bias (see Kanouse & Hanson, 1972) such that concern over evaluations is particularly focused on the possibility of negative rather than positive feedback. It

[4] The author is indebted to an anonymous reviewer who raised this argument.

would then follow that the salience of negative events such as rejection or criticism is heightened more by self-awareness than the salience of positive events such as acceptance or praise.

Self-Awareness Theory

Self-awareness has been construed primarily as a process that regulates the direction and intensity of thoughts, feelings, and actions (Duval & Wicklund, 1972). Much of the research in this area has been interested either in the relationship between behavior and standards of behavior (e.g., Carver, 1974; Scheier et al., 1974) or in the effects of self-focused attention on internal states (e.g., Scheier & Carver, 1977) and cognitive phenomena (e.g., Carver & Scheier, 1978). In the context of this previous research, it has been assumed that self-focusing stimuli such as mirrors direct attention toward covert, internal aspects of the self. Similarly, interest in personality dimensions of self-consciousness has concentrated primarily on the private component (e.g., Scheier, 1976).

The study of self-awareness, then, seems to have been largely concerned with what James (1890) called the "spiritual" self, which refers to our inner being, our cognitive faculties, our emotional states, and our behavioral tendencies. But the self involves many facets, and a particularly important constituent is the "social" self (James, 1890), which is concerned with the recognition or regard we get from others. The present research suggests that when we attend to ourselves in social situations, it is the social self that is focused on; that is, we become more aware of ourselves as objects of attention to others. In this respect, it has been shown that public self-consciousness (awareness of ourselves as social objects), rather than private self-consciousness (awareness of our internal selves), is the crucial determinant of how much the evaluations of others affect us. In addition, the parallel between the effects of public self-consciousness and the mirror manipulation suggests that in social situations, mirrors direct attention toward the public, external aspect of the self. Thus, the consequences of self-attention are not limited to a greater awareness of our own standards, emotions, and cognitions; self-attention also affects our awareness of the self-relevant actions, thoughts, and feelings of others. Self-awareness theory needs to acknowledge the existence of both the private, internal self and the public, social self if it is to develop a fuller understanding of our unique ability to be conscious of ourselves.

Conclusion

The present findings indicate that self-attention is an important mediator of social behavior. By raising or lowering attention to oneself, one can heighten or diminish the effect of others' evaluations, perhaps through the facilitation of selective attention processes (Duval & Wicklund, 1972). Indeed, our own personal exerience would suggest that the reduction of self-attention through, for example, avoiding eye contact (Fenigstein, Note 2) or engaging in random physical activity (Duval & Wicklund, 1973) may be an effective strategy for warding off the impact of criticism, derogation, or embarrassment. Further research confirming these inferences would suggest that the control and modification of self-attention may have implications for the way in which persons deal with the social feedback of everyday life.

Reference Notes

1. Paulus, P. B., Annis, A. B., & Risner, H. T. *An analysis of the mirror-induced self-awareness effect.* Unpublished manuscript, 1976. (Available from P. B. Paulus, University of Texas, Arlington, Texas 76010.)
2. Fenigstein, A. *Eye contact, self-attention, and the processing of self-relevant feedback.* Unpublished manuscript, 1978. (Available from the author.)

References

Abramson, L. Y., & Sackheim, H. A. A paradox in depression: Uncontrollability and self-blame. *Psychological Bulletin,* 1977, *84,* 839–851.

Argyle, M. *Social interaction.* Chicago: Aldine, 1969.

Argyle, M., & Williams, M. Observer or observed? A reversible perspective in person perception. *Sociometry,* 1969, *32,* 396–412.

Buss, D. M., & Scheier, M. F. Self-consciousness, self-awareness, and self-attribution. *Journal of Research in Personality,* 1976, *10,* 463–468.

Carver, C. S. The facilitation of aggression as a function of objective self-awareness and attitudes toward punishment. *Journal of Experimental Social Psychology*, 1974, *10*, 365–370.

Carver, C. S., & Glass, D. C. The self-consciousness scale: A discriminant validity study. *Journal of Personality Assessment*, 1976, *40*, 169–172.

Carver, C. S., & Scheier, M. F. Self-focusing effects of dispositional self-consciousness, mirror presence, and audience presence. *Journal of Personality and Social Psychology*, 1978, *36*, 324–332.

Dittes, J. E. Attractiveness of group as a function of self-esteem and acceptance by group. *Journal of Abnormal and Social Psychology*, 1959, *59*, 77–82.

Dittes, J. E., & Kelley, H. H. Effects of different conditions of acceptance upon conformity to group norms. *Journal of Abnormal and Social Psychology*, 1956, *53*, 100–107.

Duval, S. Conformity on a visual task as a function of personal novelty on attitudinal dimensions and being reminded of the object status of self. *Journal of Experimental Social Psychology*, 1976, *12*, 87–98.

Duval, S., & Wicklund, R. A. *A theory of objective self-awareness*. New York: Academic Press, 1972.

Duval, S., & Wicklund, R. A. Effects of objective self-awareness on attribution of causality. *Journal of Experimental Social Psychology*, 1973, *9*, 17–31.

Ellsworth, P. C., & Carlsmith, J. M. Effects of eye contact and verbal content on affective responses to a dyadic interaction. *Journal of Personality and Social Psychology*, 1968, *10*, 15–20.

Fenigstein, A., Scheier, M. F., & Buss, A. H. Public and private self-consciousness: Assessment and theory. *Journal of Consulting and Clinical Psychology*, 1975, *43*, 522–527.

Geller, D. M., Goodstein, L., Silver, M., & Sternberg, W. C. On being ignored: The effects of the violation of implicit rules of social interaction. *Sociometry*, 1974, *37*, 541–556.

Geller, V., & Shaver, P. Cognitive consequences of self-awareness. *Journal of Experimental Social Psychology*, 1976, *12*, 99–108.

Goffman, E. *The presentation of self in everyday life*. Garden City, N.Y.: Doubleday, 1959.

Jackson, J. M., & Saltzstein, H. D. The effect of person–group relationships on conformity processes. *Journal of Abnormal and Social Psychology*, 1959, *57*, 17–24.

Jacobs, L., Berscheid, E., & Walster, E. Self-esteem and attraction. *Journal of Personality and Social Psychology*, 1971, *17*, 84–91.

James, W. *The principles of psychology* (Vol. 1). New York: Holt, 1890.

Jones, E. E., & Davis, K. E. From acts to dispositions. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2). New York: Academic Press, 1965.

Jones, S. C., Knurek, D. A., & Regan, D. T. Variables affecting reactions to social acceptance and rejection. *The Journal of Social Psychology*, 1[illegible] *90*, 269–284.

Kanouse, D. E., & Hanson, L. R. Negativity in evaluations. In E. E. Jones et al., (Eds.), *Attribution: Perceiving the causes of behavior*. Morristown, N.J.: General Learning Press, 1972.

Kuiper, N. A. Depression and causal attributions for success and failure. *Journal of Personality and Social Psychology*, 1978, *36*, 236–246.

Liebling, B. A., Seiler, M., & Shaver, P. Self-awareness and cigarette-smoking behavior. *Journal of Experimental Social Psychology*, 1974, *10*, 325–332.

Mead, G. H. *Mind, self and society*. Chicago: University of Chicago Press, 1934.

Mettee, D. R., Fisher, S., & Taylor, S. E. The effect of being shunned upon the desire to affiliate. *Psychonomic Science*, 1971, *23*, 429–431.

Pepitone, A., & Wilpinski, C. Some consequences of experimental rejection. *Journal of Abnormal and Social Psychology*, 1960, *60*, 359–364.

Pryor, J. B., & Kriss, M. The cognitive dynamics of salience in the attribution process. *Journal of Personality and Social Psychology*, 1977, *35*, 49–55.

Scheier, M. F. Self-awareness, self-consciousness, and angry aggression. *Journal of Personality*, 1976, *44*, 627–644.

Scheier, M. F., & Carver, C. S. Self-focused attention and the experience of emotion: Attraction, repulsion, elation, and depression. *Journal of Personality and Social Psychology*, 1977, *35*, 625–636.

Scheier, M. F., Fenigstein, A., & Buss, A. H. Self-awareness and physical aggression. *Journal of Experimental Social Psychology*, 1974, *10*, 264–273.

Snoek, J. D. Some effects of rejection upon attraction to a group. *Journal of Abnormal and Social Psychology*, 1962, *64*, 175–182.

Snyder, M. The self-monitoring of expressive behavior. *Journal of Personality and Social Psychology*, 1974, *30*, 526–537.

Storms, M. D. Videotape and the attribution process: Reversing actors' and observer's points of view. *Journal of Personality and Social Psychology*, 1973, *27*, 165–175.

Taylor, S., & Fiske, S. Point of view and perceptions of causality. *Journal of Personality and Social Psychology*, 1975, *32*, 439–445.

Wicklund, R. A. Objective self-awareness. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 8). New York: Academic Press, 1975.

Zimbardo, P. G. The human choice: Individuation, reason and order versus deindividuation, impulse and chaos. In W. J. Arnold & D. Levine (Eds.), *Nebraska Symposium on Motivation* (Vol. 17). Lincoln: University of Nebraska Press, 1969.

Received November 15, 1977 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 1, 87-96

Multidimensional Reaction to Therapeutic Touch in a Hospital Setting

Sheryle J. Whitcher and Jeffrey D. Fisher
University of Connecticut

A field experiment focused on some implications of interpersonal touch not explored in earlier research. Conceptually, the research included measuring the effects of touch over a relatively long time frame, for a broad range of response dimensions, and in a nonreactive setting characterized by dependency. On an applied level, the research studied the value of touch as a concomitant of nurse–patient interactions. Specifically, a 2 (touch vs. no touch) × 2 (male vs. female) between-subjects design assessed the effects of nurses touching patients, during preoperative teaching, on patient affective, evaluative, behavioral, and physiological responses. Results indicated that female subjects in the touch condition experienced more favorable affective, behavioral, and physiological reactions than a no-touch control group. In contrast, males in the touch condition reacted more negatively than control subjects on these dimensions.

Background

The various modes of nonverbal communication (e.g., proxemics, eye contact) have received increased attention from social psychologists in recent years. There have been several theoretical (e.g., Argyle & Dean, 1965; Patterson, 1976) and numerous empirical (e.g., Buck, Miller, & Caul, 1974) attempts to understand the way nonverbal messages are interpreted and to predict the quality of subsequent reactions.

Surprisingly, although touch is considered by many to be the most powerful of the nonverbal modalities, it has traditionally received the least research attention (Duncan, 1969). Further, what data are available on tactile stimulation have generally focused on the responses of animals and human infants to touch. Within the animal literature, studies have revealed that early tactile contact can influence later emotional and physiological behavior (Denenberg, 1963; Harlow, 1971; Levine, 1960). With human infants, the classic work of Spitz (1946) and more recent research by Montagu (1971) highlight the importance of touch as a stimulus essential for normal intellectual, emotional, and social development.

Only in the last few years have researchers begun to explore the parameters of touch with human adults. Further, the limited work with this population suggests that the effects of touch may be either positive or negative. For example, in research demonstrating positive consequences, Kleinke (1977) reported that touch led to greater compliance than in a control group, and Fisher, Rytting, and Heslin (1976) observed that females who were touched experienced more positive affective and evaluative reactions than no-touch controls. In research suggesting negative effects, Walker (1971) reported that communication by means of touch made subjects feel anxious and uncomfortable, and Henley (1973) observed that touch may be perceived as ex-

This research was supported in part by Research Grant 5.171-000-22-0215-35-335 from the University of Connecticut Research Foundation to Jeffrey D. Fisher. Portions of this article were presented at the meeting of the Psychonomic Society, Washington, D.C., November 1977. The contribution of the nursing staff at the University of Connecticut Health Center, who were extremely helpful to the authors throughout all phases of this research, is gratefully acknowledged.

Requests for reprints should be sent to Jeffrey D. Fisher, Department of Psychology, U-20, University of Connecticut, Storrs, Connecticut 06268.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3701-0087$00.75

ploitative and/or as highlighting the lower status of the recipient. Finally, it has been shown that the same touch may be experienced positively by one sex and negatively by the other (Fisher et al., 1976; Nguyen, Heslin, & Nguyen, 1975).

A review of past research with adult populations suggests that whether touch is experienced positively or negatively depends on the meaning and evaluation inferred by the recipient (Fisher et al., 1976). Clearly, touching a person can mean many things, for example, a desire for intimacy, sexual attraction, or dominance. Further, any of these messages could be expected to elicit negative reactions if it oversteps the boundaries a person has defined as appropriate. In line with these observations, and supported by conceptual work on nonverbal communication (Altman, 1975; Argyle & Dean, 1965; Patterson, 1976), Fisher et al. (1976) suggested that a touch will be experienced as positive to the extent that it (a) is appropriate to the situation, (b) does not impose a greater level of intimacy than the recipient desires, or (c) does not communicate a negative message (e.g., is not perceived as condescending). These assertions have also been supported in empirical work (e.g., Hall, 1966; Henley, 1973; Nguyen et al., 1975).

Problems With the Extant Literature on Touch

Although the present literature on touch in human adults affords some organizing notions for predicting responses and suggests that reactions may vary widely, it is limited in important ways. For example, the range of situations examined has been decidedly narrow (e.g., touch has rarely been studied in situations characterized by dependency, anger, or sexual intimacy; Henley, 1977). Also, very few studies have been conducted in a nonreactive context (Fisher et al., 1976), which is a serious limitation, since touch is associated with strong socialization-related prohibitions (Heslin, Note 1). In addition, past research has measured only reactions that occur very soon after tactile stimulation; the longer-term consequences of touch have yet to be explored. Further, the responses assessed have generally been limited to the affective and evaluative domains; behavioral and physiological reactions to touch have been measured infrequently, and no research has provided a simultaneous assessment of all four response dimensions. Finally, most studies on the effects of touch have been empirical demonstrations. Little research has been conducted either to test a particular conceptual perspective or in a context where findings could have direct applied significance.

The Present Experiment and its Implications

The present experiment was designed to yield initial data pertinent to some of the weakness of past research described above. Specifically, the study explored the effects of nurse–patient touch in a hospital setting. On a conceptual level, this research contributes to the extant literature in the following ways. First, nurse–patient interactions are highly characterized by dependency (Barnett, 1972), a context variable ignored in past research on touch (Henley, 1977). In addition, the effects of touch were simultaneously assessed on affective, evaluative, behavioral, and physiological dimensions, and over a relatively long period of time. Finally, the nature of the touch stimulus and of the dependent measures rendered the study relatively unobtrusive, so that it would be affected less by problems associated with reactivity.

In addition to its conceptual implications, the experimental study of touch in a hospital setting may have practical value. Although no past research on nurse–patient interaction has directly measured the applied consequences of tactile stimulation, the potential of touch is suggested by work in related areas. Specifically, it has been found that hospitalization is frequently an anxiety-laden, depersonalized situation (Barnett, 1972), that appropriate supportive care (of which touch may constitute an element) leads to reduced anxiety (Johnson & Leventhal, 1974), and, most importantly, that the level of patient anxiety may be related to postoperative recovery (Johnson, Leventhal, & Dabbs, 197

Hence, by addressing the effects of touch in the present setting, this study may disclose some initial answers to an important applied problem.

Predicting the Effects of Touch in a Hospital Context

In predicting the effects of touch in a context characterized by dependency, some important factors should be taken into account. Clearly, in such a setting touch may convey a mixture of positive (e.g., caring) and negative (e.g., power, dominance) elements (Henley, 1977; Heslin, Note 1). This interpretive ambiguity is exacerbated, because in a physical sense, touches that express negative aspects of dependency (e.g., inferiority) and those that express positive aspects of caring and concern are quite similar (Henley, 1977; Heslin, Note 1). Thus, it is not surprising that competing rationales exist for predicting which of these two messages will be most salient.

One prediction stems from research on sex differences in socialization. Specifically, it is assumed that males are socialized to be more uncomfortable with dependency than females (Hoffman, 1972; Maccoby, 1966; Stein & Bailey, 1973). Further, since touches signaling dependency and those expressing caring may be physically similar, it is believed that socialization may importantly affect their interpretation. Thus, females may interpret nurse–patient touch as a primarily positive message of caring and warmth, which is situationally appropriate and of a reasonable level of intimacy (cf. Fisher et al., 1976). On the other hand, males may react negatively, because they interpret the nurse's touch as conveying a message of relative inferiority and dependency, which is neither situationally appropriate nor of a comfortable level of intimacy. In effect, it is suggested that touch in a dependency context is more sex role-appropriate for females than for males and should therefore lead to more positive female than male responses across measurement indexes (e.g., affect, evaluation). While sex role-inappropriate touch has been shown to lead to negative effects in other contexts (e.g., romantic encounters; Nguyen et al., 1975), it has not yet been explored in dependency situations.

An opposing prediction centers on the special status that may be accorded to dependency that results from illness. Specifically, when one is a patient, being dependent may be socially acceptable for both sexes. In line with this interpretation, it would be expected that both males and females would interpret nurse–patient touch as an expression of caring that is situationally appropriate and of a reasonable level of intimacy (cf. Fisher et al., 1976). If this hypothesis were supported, the positive reaction of males and females to touch should generalize across the various response dimensions.

Method

Subjects

Subjects were patients at a major university hospital complex. Forty-eight individuals (19 males, 29 females) who entered the hospital for elective surgery participated in the study; anonymity was maintained through the use of code numbers. The prerequisites for inclusion were (a) an adequate reading and speaking knowledge of English, (b) signing a document upon admission to the hospital that permitted participation in research,[1] (c) assignment to the seventh floor of the hospital, and (d) assignment to a primary nurse who had agreed to participate in the study.[2] All individuals who met these criteria were included in the experiment.

Design

A 2 (touch vs. no touch) × 2 (male vs. female) between-subjects design was employed, in which patients were randomly assigned to conditions. For ethical reasons, and to permit a comparison with "routine procedures," individuals in control groups were not deprived of any aspect of normal hospital care and were exposed to customary professional/functional tactile contact with nurses. In experi-

---

[1] The number of patients who refused to sign the informed-consent document averaged 10%, with no differences between the experimental cells.

[2] All nursing personnel scheduled to work on the seventh floor volunteered to take part in the study and were blind to the experimental hypotheses. Further, it should be noted that primary nursing care, a system in which a patient is assigned to one nurse who coordinates care throughout the hospital stay, is used at the facility where the study was run.

mental conditions, the nurses administered an additional touch manipulation (described in detail below) during the preoperative teaching session. The same nurses conducted preoperative instruction with patients in the experimental and control groups.

Experimental Procedures

If a patient met the prerequisites for participation, the floor clerk attached a specially prepared "work sheet" to his or her medical chart when it arrived at the seventh floor. The work sheet specified the patient's treatment condition and had been randomly sequenced by the experimenter prior to the beginning of the experiment. When the nurse examined the patient's chart, the work sheet informed her if the patient was participating in the study and what condition he or she was in.

The experimental manipulation took place shortly after admission. Initially, the nurse closed a curtain that separated the patient from any other individuals in the room. Next, she touched the patient's hand for a few seconds while introducing herself and explaining that she was going to inform the patient about his or her surgery. Toward the end of the teaching session, the nurse gave the patient a booklet that further detailed the procedures for surgery, to be read at his or her leisure. At that time the nurse put one hand on the patient's arm and maintained this touch for approximately 1 minute, while she and the patient examined the booklet.[3] Patients in the control condition experienced the same interaction without physical contact. Subsequent to this, in both the experimental and control groups, the nurse extended her hand in a gesture suggesting contact and noted the patient's response as an unobtrusive measure of the subject's desire to reciprocate contact.

Upon completing the teaching session, the primary nurse left the room. Several hours later the patient was approached by a male experimenter, blind to experimental hypotheses, who explained that the university was conducting a study of patient reactions to hospitalization. The experimenter requested that the patient fill out a questionnaire, emphasizing that all responses were anonymous and confidential. The patient was asked to seal the completed form in an envelope, which was later collected by the experimenter.

Dependent Measures

The dependent measures that were used to assess the effects of touch represented a set of affective, evaluative, behavioral, and physiological indexes expected to discriminate between patients who react favorably and those who react unfavorably to the hospital experience.

Measure of Affect

The measure of affect was contained in the questionnaire administered by the experimenter. It consisted of three items with intercorrelations above .40 that assessed the nature of the patient's feelings concerning hospitalization and surgery. Specifically, the items measured how anxious patients felt about having to undergo surgery, how unpleasant it was for them to be in the hospital, and how worried they were concerning the possibility of postoperative complications. Each item was presented as a 7-point semantic differential rating scale, ranging from "not at all" to "extremely."

Evaluative Measures

The evaluative measures were also contained in the questionnaire administered by the experimenter. These consisted of a measure of satisfaction with preoperative instruction and one reflecting attraction toward the nurse.

Measure of satisfaction with preoperative instruction. This index consisted of two items, which asked patients how satisfied they were with the information they received concerning surgery and how satisfied they were with the nursing care provided up to that point. The items showed an intercorrelation of .36 and were arranged as 7-point semantic differential scales with endpoints of "not at all" and "extremely."

Measure of attraction. Three items comprised a measure of the degree of patient liking for the nurse. Specifically, they asked patients to indicate how warm they perceived the nurse to be, how interested they felt the nurse was in their questions and feelings about surgery, and how friendly they felt toward the nurse. The items showed intercorrelations above .50, and each was presented as a 7-point semantic differential scale with endpoints of "not at all" and "extremely."

Behavioral Measures

Two unobtrusive measures were included that assessed responses made by subjects during their hospital stay.

Measure of preoperative material read. It was assumed that touch might lead to differential patient reading of the booklet left by the nurse during preoperative teaching. Therefore, a 4-point Likert-type scale assessed how much of the booklet the patient had read. This measure ranged from "none at all" to "the entire booklet thoroughly."

---

[3] Prior to running subjects, the researchers observed the setting and the procedure in which the touch stimulus was to be introduced. This was done to verify that the touch manipulation could be performed naturally and could fit comfortably in the preoperative instruction. Further, the researchers trained the staff in the execution of the touch. All nurses role played the manipulation with each other before the experiment began to insure that it would be administered in a uniform fashion. In addition, they were provided with a typewritten set of instructions, which they reviewed periodically.

Measure of reciprocation of touch. At the end of the preoperative instruction, the nurse extended her hand in a gesture that allowed for reciprocal intimacy and then observed the patient's reaction. The response was coded in terms of whether the patient (a) grasped or touched the nurse's hand, (b) reached toward but did not touch the nurse's hand, (c) looked at but did not reach toward the nurse's hand, or (d) ignored the nurse's gesture.

Physiological Measures

The physiological data recorded for patients included measures taken directly after surgery while the patient was in the recovery room and periodic vital signs readings collected by the staff during the first several days of hospitalization. Specifically, the recovery room data consisted of five repeated measures of patient pulse and systolic and diastolic blood pressure, recorded at 15-minute intervals. The periodic vital signs data consisted of seven repeated measures of pulse, temperature, and systolic and diastolic blood pressure, collected approximately every 4 hours after the preoperative teaching. Past research has shown that these indexes can be considered to discriminate between anxious and nonanxious patients (e.g., Martin, 1961). All readings of blood pressure were recorded by a Baum portable mercury-meter sphygmomanometer with a branchial artery cuff. Measures of heart rate were assessed by a radial pulse, and temperature data were recorded with a standard mercury thermometer.[4]

Results

Equivalence of Groups

Preliminary 2 (touch vs. no touch) × 2 (male vs. female) analyses of variance (ANOVAs) and chi-square tests were run to insure that randomization had resulted in equivalence of patients in all cells. Since the ANOVAs revealed no differences between groups in terms of physiological indexes or age at hospital entry ($Fs < 1$), and the chi-square tests indicated no differences in type (e.g., orthopedic, eye, thoracic) or severity (major vs. minor) of surgery ($\chi^2 < 1$), each of the groups was assumed to be equivalent. Thus, the dependent variables were subject to further analysis.

Measure of Affect

A 2 (touch vs. no touch) × 2 (male vs. female) multivariate ANOVA was performed on the three questionnaire items pertaining to affect concerning surgery.[5] This procedure revealed a significant main effect for touch, $\lambda(1, 3) = 3.09$, $p < .037$, which was qualified by a significant Touch × Sex interaction, $\lambda(1, 3) = 4.84$, $p < .006$. Parallel univariate analyses run on the three affect items indicated a similar pattern of findings.

Specifically, the univariate analyses revealed a significant main effect for touch on the item reflecting worry about complications, $F(1, 43) = 9.27$, $p < .004$, and a marginally significant main effect for touch on the item that measured unpleasantness of hospitalization, $F(1, 43) = 2.82$, $p < .10$. These findings indicated that patients who were touched during preoperative instruction experienced more positive reactions than individuals not touched. Further, a significant univariate Touch × Sex interaction was observed for ratings of anxiety concerning surgery, $F(1, 43) = 15.01$, $p < .001$, and trends toward significant Touch × Sex interactions were observed for the measures of unpleasantness of hospitalization ($p < .13$) and worry about complications ($p < .16$); see Table 1. For the measure of anxiety concerning surgery, Newman-Keuls comparisons indicated that females who were touched reported less anxiety than control females ($p < .05$), whereas males in the touch condition reported more anxiety than control males ($p < .05$). Also, females were more anxious than males in control conditions ($p < .05$), whereas males were more anxious than females in touch conditions ($p < .05$). The overall pattern of the nonsignificant interactions paralleled that observed for the measure of anxiety concerning surgery.

---

[4] All experimental procedures were approved by the hospital ethics committee composed of doctors, lay people, and clergy.

[5] F_{max} tests indicated that homogeneity of variance was maintained for all of the dependent measures. Since there were minor discrepancies in the n in various cells, a least squares ANOVA was used (cf. Overall & Spiegel, 1969; Winer, 1971). Finally, the Wilks's Lambda criterion (Finn, 1974) was used as a significance criterion for all the multivariate tests.

Table 1
Cell Means for Touch × Sex Interactions

Measure	Touch		No touch	
	Male	Female	Male	Female
Anxiety over surgery	6.2$_a$	3.8$_b$	3.9$_b$	5.9$_a$
Unpleasantness of hospitalization	3.2$_a$	3.5$_a$	3.0$_a$	5.3$_a$
Worry concerning complications	3.0$_a$	2.9$_a$	3.6$_{a,b}$	5.1$_b$
Recovery room systolic pressure	144.6$_a$	121.8$_a$	126.9$_a$	135.0$_a$
Recovery room diastolic pressure	96.2$_a$	78.3$_a$	82.4$_a$	83.1$_a$

Note. Means with common subscripts do not differ at the .05 level, as indicated by the Newman–Keuls procedure. For the measures of anxiety over surgery, unpleasantness of hospitalization, and worry about complications, scores can range from 1 (low) to 7 (high).

Evaluative Measures

Measure of Satisfaction With Preoperative Instruction

A 2 × 2 multivariate ANOVA and parallel univariate ANOVAs performed on the two questionnaire items assessing patient satisfaction with preoperative teaching were not significant. However, responses on these measures indicated a ceiling effect, that is, patients in all conditions felt extremely satisfied with the information and nursing care they received.

Measure of Attraction

A 2 × 2 multivariate ANOVA and parallel univariate ANOVAs performed on the three questionnaire items that assessed patient liking for the nurse revealed no significant effects. However, when the data for males and females were analyzed separately, some suggestive findings emerged. For females, the multivariate ANOVA indicated a trend toward a main effect for touch, $\lambda(1, 3) = 1.64$, $p < .20$. This finding was paralleled by a univariate effect for touch on the item measuring perceived nurse interest, $F(1, 26) = 5.10$, $p < .03$, which suggests that females perceived nurses who touched to be more interested in them than nurses who did not touch. For the measures of perceived nurse warmth and friendliness, the separate male–female univariate analyses revealed no significant effects; however, responses on these items indicated a ceiling effect.

Behavioral Measures

Measure of Preoperative Material Read

Patient reading of the preoperative teaching booklet was measured on a 4-point scale. The overall 2 × 4 chi-square analysis was significant, $\chi^2(3) = 9.10$, $p < .03$, and suggested that patients in the touch condition read more of the booklet than controls. Separate 2 × 4 chi-square tests revealed that for females, the effect of touch paralleled the overall chi-square analysis, $\chi^2(3) = 10.63$, $p < .01$. However, for males the separate analysis was nonsignificant, $\chi^2(3) = 1.78$, $p < .5$, indicating that it was the female data that were primarily responsible for the overall effect.

Measure of Reciprocation of Touch

Patient responses to the nurse's outstretched hand were coded on a 4-point scale reflecting the subject's desire for reciprocal intimacy. The overall 2 × 4 chi-square analysis was marginally significant, with patients in the touch condition more frequently reaching out and touching the nurse's hand relative to control patients, $\chi^2(3) = 6.97$, $p < .07$. Separate 2 × 4 chi-squares for males and females revealed that while the pattern approached significance for females, $\chi^2(3) = 7.39$, $p < .06$, it was not significant for males ($\chi^2 < 1$). Here again, it was the female response that determined the overall effect.

Physiological Measures

Recovery Room Measures

A 2 (touch vs. no touch) × 2 (male vs. female subject) × 5 (time) ANOVA was performed for each of the physiological measures taken in the recovery room.

Systolic blood pressure. The ANOVA performed on the recovery room index of systolic blood pressure revealed a significant Touch × Sex interaction, $F(1, 31) = 4.15$, $p < .05$ (see Table 1). The pattern of the means indicates that at all five measurement intervals, touch females tended to have lower systolic pressure than control females, whereas touch males tended to have higher readings than male controls. However, Newman-Keuls analyses revealed that none of these differences were significant. The overall ANOVA further revealed a significant effect for time, $F(1, 128) = 2.85$, $p < .03$, which indicated that patients in both conditions demonstrated a significant decline in systolic pressure over the period of measurement.

Diastolic blood pressure. The ANOVA performed on the five measures of diastolic blood pressure showed a significant Touch × Time interaction, $F(1, 128) = 2.47$, $p < .05$, as well as a marginally significant Touch × Sex interaction, $F(1, 31) = 3.26$, $p < .08$ (see Table 1). Taken together, the interactions suggest that touch males tended to increase in diastolic blood pressure relative to control males over the five intervals. In contrast, touch females tended to decrease in diastolic pressure over time relative to control females. However, Newman-Keuls comparisons of the means did not reach a conventional level of significance for either interaction.

Pulse. The repeated measures ANOVA on the five pulse readings indicated that only the effect for time was significant, $F(1, 124) = 4.51$, $p < .003$, with patients demonstrating a decrease in pulse over the period of measurement.

Periodic Vital Signs Measures

In addition to the physiological data collected in the recovery room, periodic measures of pulse, temperature, and systolic and diastolic blood pressure were recorded for patients approximately every 4 hours after preoperative teaching. A 2 (touch vs. no touch) × 2 (male vs. female subject) × 7 (time intervals following preoperative teaching) ANOVA was performed on these measures.

Systolic blood pressure. The ANOVA on the periodic index of systolic pressure revealed a significant effect for sex of subject, $F(1, 30) = 13.46$, $p < .002$, with males demonstrating higher systolic pressure than females. Although the analysis failed to indicate a main effect for touch ($p < .26$), the direction of cell means suggested that patients in the touch condition ($M = 122.73$) tended to have lower systolic pressure over the seven measurement intervals than controls ($M = 127.37$).

Diastolic blood pressure. The ANOVA on the periodic diastolic blood pressure readings revealed a significant main effect for sex of subject, $F(1, 30) = 14.25$, $p < .001$, with males demonstrating higher diastolic pressure at all measurement intervals than females. The data further indicated a trend toward significance for the effect of touch, $F(1, 30) = 2.40$, $p < .13$, with patients in the touch condition ($M = 76.03$) tending to have lower diastolic pressure over the seven measures than controls ($M = 86.38$).

Pulse. The repeated measures ANOVA performed on the seven pulse readings revealed no significant effects ($Fs < 1$).

Temperature. The repeated measures ANOVA performed on the periodic vital signs temperature data revealed a significant effect for time, $F(1, 30) = 6.03$, $p < .001$, with patients showing an increase in temperature over the seven measurement intervals. This main effect was qualified by a significant Sex × Time interaction, $F(1, 186) = 2.26$, $p < .04$, which indicated that males' temperatures tended to increase at a more rapid rate than females'.

Discussion

In an attempt to provide some conceptual and applied data not available from earlier work, the present study explored the effects of touch in a hospital setting. The findings

on a variety of measures suggested that in this context, touch led to positive effects primarily for females. Specifically, on affective dimensions touch precipitated positive reactions for females, whereas the reverse was true for males. Further, the behavioral data indicated that touch led to reciprocal touching and to more extensive reading of the preoperative material for female but not for male subjects. Finally, several physiological indexes also tended to show this pattern of positive reactions only for females.

Evaluation of the Support for the Sex Difference Hypothesis

While there was a degree of inconsistency across the various dependent measures, the overall pattern of results supports the sex difference hypothesis, which predicts that when dependency cues are pervasive, males and females will react differently to touch. The rationale for this hypothesis is that males are socialized to be relatively uncomfortable with dependency (Hoffman, 1972; Maccoby, 1966; Stein & Bailey, 1973) and thus experience nurse–patient touch as a threatening gesture communicating inferiority and asserting dominance, which leads to negative reactions. Conversely, females are socialized to be relatively affiliative and comfortable with dependency (Hoffman, 1972; Maccoby, 1966; Stein & Bailey, 1973) and thus experience nurse–patient touch as a supportive gesture of caring and warmth, which leads to favorable reactions. Our support for the sex difference hypothesis complements earlier research on tactile stimulation. In contexts other than dependency, it has been found that when touch conveys a sex role-inappropriate message, negative effects occur (Nguyen et al., 1975; Heslin & Boss, Note 2).

Although the present data were explained in terms of the sex difference hypothesis detailed above, several other possibilities are tenable. One centers on females' richer and more varied tactile history. Specifically, females are touched more often by a variety of others (mother, father, same/opposite-sex friends) than males (Jourard, 1966; Jourard & Rubin, 1968), and in other nonverbal modalities (e.g., personal space) they display more permeable boundaries. Further, females probably have more exposure to tactile stimulation in medical contexts than males (e.g., yearly gynecological exams, giving birth). Hence, touch during preoperative teaching may have been reassuring for females but disruptive for males in part because males have less experience with touch in general and, more particularly, in medical contexts.

A second possibility stems from the fact that female patients interacted with a same-sex other, whereas males interacted with an opposite-sex individual. Since there was no way to separate out the possible confounding influence of sex of toucher from the effects of touch in the present setting, it is possible that being touched during preoperative teaching was a qualitatively different experience for males and females. For example, males could be more likely than females to interpret the present touch as sexual in nature. However, past research tends to argue against this possibility. Specifically, a study by Nguyen et al. (1975) indicates that touches to the hand and/or shoulder, whether applied by a same- or an opposite-sex other, are interpreted as nonsexual in nature. Further, an earlier field experiment that varied sex of toucher and employed a similar touch manipulation found no differential effects for sex of toucher (Fisher et al., 1976). Nevertheless, it is not possible to rule out the above explanation solely on the basis of this study.

Consistency of the Data Across Dependent Measures

While the present findings generally showed a consistent pattern of effects for the affective and behavioral measures, the physiological measures were less consistent. For example, the measures taken in the recovery room indicated that touch females had lower systolic and diastolic blood pressure than controls, whereas similar measures taken in the patient's room showed only a trend toward a effect for touch. Also, indexes of pulse and temperature did not discriminate between experimental and control conditions. Inter

estingly, on some of these measures sizable mean differences were obtained, yet effects were not significant. If the high degree of variability associated with physiological measurement is taken into account (e.g., Martin, 1961), these nonsignificant patterns become more impressive. Further, it should be noted that where sources of variability were minimized (i.e., in the recovery room), stronger effects were observed than under less controlled conditions (i.e., in the patient's room).

Although the affective, behavioral, and physiological measures were generally sensitive to the effects of touch, the evaluative items were not. We think that this result can be explained in terms of a ceiling effect, since the means show almost no differences and indicate that all patients felt very positively toward the nurse and were extremely satisfied with their preoperative teaching. Considering the very high level of health care at the hospital where the present study was run, a ceiling effect on these indexes should not have been unexpected. However, it would not be unreasonable to expect stronger effects on evaluative measures in settings where health care is much less personalized (e.g., emergency rooms).

Conceptual Implications

Moving to conceptual implications, the present data may be seen as adding to the extant literature. For example, past research tended to look at single rather than multiple responses to tactile stimulation. Based on this study, it appears that touch may have consequences for affective state, behavior, and physiological reactions and that these response modalities may be related (i.e., parallel effects may occur along all three dimensions). Further, the data provide a first look at some long-term consequences of touch, suggesting that it may have powerful multidimensional effects over time, in addition to the short-term effects demonstrated in earlier research (e.g., Fisher et al., 1976). It should be noted, however, that touch in the present study occurred in a context where it was particularly salient and meaningful. While touch in other situations of this type might be expected to have long-term consequences, the effects of incidental tactile stimulation (cf. Fisher et al., 1976) or touch that is commonplace or role defined should be relatively short-term.

It is of further conceptual interest that the present findings are in accord with Patterson's (1976) arousal model of interpersonal intimacy and can be interpreted in terms of that formulation. According to Patterson, changes in intimacy produce arousal change, which may be a signal to evaluate and interpret the environment. The change in arousal may be labeled as positive or negative, depending on the context in which it occurs (e.g., socialization-related beliefs concerning touch and dependency). To the extent that a positive label is engendered, it is assumed that favorable reactions (e.g., positive affect, reciprocal intimacy) will occur. In the present situation for females, contextual factors may have permitted a positive label for decreased arousal, which resulted in favorable affective, physiological, and behavioral reactions. For males, the context may have produced a negative labeling of increased arousal, which resulted in unfavorable responses on these dimensions. It is suggested that the fit of Patterson's model to the present data indicates that it may be useful as an interpretive framework for past research on touch and as a predictive framework for future work.

Practical Implications

Finally, the results of this study may have practical implications for patient care. Earlier research (e.g., Johnson et al., 1971) found that various conditions that facilitate low preoperative anxiety improve patient recovery. In the present study, touch for females resulted in lower anxiety and more positive behavior preoperatively, and these results were associated with more favorable postoperative physiological responses. This finding suggests that touch during preoperative teaching is a form of communication that may be beneficial to the well-being of certain patients. Clearly, however, the negative reactions of males in the present context prohibit any unqualified statements concerning the practi-

cal implications of the data. Before application will be possible, future research must assess the consequences of touch across a variety of medical settings and individual difference variables, in order to very precisely delineate the domain in which touch has beneficial effects.

Conclusions

In conclusion, touch in a hospital setting tended to elicit positive reactions in females and more negative reactions in males. It was suggested, based on these findings, that male and female subjects attributed differential meaning to touch in a dependency context. Further, the findings demonstrated that touch may have long-term effects, that reactions to touch may be related across various response dimensions, and that tactile stimulation in a hospital setting may sometimes be of therapeutic value.

Reference Notes

1. Heslin, R. *Attraction in the dyad.* Unpublished manuscript, Purdue University, 1976.
2. Heslin, R., & Boss, D. *Nonverbal boundary behavior at the airport.* Unpublished manuscript, Purdue University, 1975.

References

Altman, I. *The environment and social behavior.* Monterey, Calif.: Brooks/Cole, 1975.

Argyle, M., & Dean, J. Eye contact, distance, and affiliation. *Sociometry,* 1965, *28,* 289–304.

Barnett, K. A theoretical construct of the concepts of touch as they relate to nursing. *Nursing Research,* 1972, *21,* 102–110.

Buck, R., Miller, R. E., & Caul, W. F. Sex, personality, and physiological variables in the communication of affect via facial expression. *Journal of Personality and Social Psychology,* 1974, *30,* 587–596.

Denenberg, V. Early experience and emotional development. *Scientific American,* 1963, *208,* 138–146.

Duncan, S. Nonverbal communication. *Psychological Bulletin,* 1969, *72,* 118–137.

Finn, J. *A general model for multivariate analysis.* New York: Holt, Rinehart & Winston, 1974.

Fisher, J. D., Rytting, M., & Heslin, R. Affective and evaluative effects of an interpersonal touch. *Sociometry,* 1976, *39,* 416–421.

Hall, E. T. *The hidden dimension.* New York: Doubleday, 1966.

Harlow, H. *Learning to love.* New York: Albion, 1971.

Henley, N. M. The politics of touch. In P. Brown (Ed.), *Radical psychology.* New York: Harper & Row, 1973.

Henley, N. M. *Body politics.* Englewood Cliffs, N.J.: Prentice-Hall, 1977.

Hoffman, L. W. Early childhood experiences and women's achievement motives. *Journal of Social Issues,* 1972, *28,* 157–176.

Johnson, J. E., & Leventhal, H. Effects of accurate expectations and behavioral instructions on reactions during a noxious medical examination. *Journal of Personality and Social Psychology,* 1974, *29,* 710–718.

Johnson, J. E., Leventhal, H., & Dabbs, J. M., Jr. Contribution of emotional and instrumental processes in adaptation to surgery. *Journal of Personality and Social Psychology,* 1971, *20,* 55–64.

Jourard, S. M. An exploratory study of body accessibility. *British Journal of Social and Clinical Psychology,* 1966, *5,* 221–231.

Jourard, S. M., & Rubin, J. E. Self-disclosure and touching: A study of two modes of interpersonal encounter and their interrelation. *Journal of Humanistic Psychology,* 1968, *8,* 39–48.

Kleinke, C. L. Compliance to requests made by gazing and touching experimenters in field settings. *Journal of Experimental Social Psychology,* 1977, *13,* 218–223.

Levine, S. Stimulation in infancy. *Scientific American,* 1960, *202,* 80–86.

Maccoby, E. (Ed.). *The development of sex differences.* Stanford, Calif.: Stanford University Press, 1966.

Martin, B. The assessment of anxiety by physiological and behavioral measures. *Psychological Bulletin,* 1961, *58,* 234–255.

Montagu, A. *Touching: The significance of the human skin.* New York: Columbia University Press, 1971.

Nguyen, T., Heslin, R., & Nguyen, M. L. The meaning of touch: Sex differences. *Journal of Communication,* 1975, *25,* 92–103.

Overall, J. E., & Spiegel, D. K. Concerning least squares analysis of experimental data. *Psychological Bulletin,* 1969, *72,* 311–322.

Patterson, M. An arousal model of interpersonal intimacy. *Psychological Review,* 1976, *83,* 235–245.

Spitz, R. A. Hospitalism. *The Psychoanalytic Study of the Child,* 1946, *2,* 113–117.

Stein, A. H., & Bailey, M. M. The socialization of achievement orientation in females. *Psychological Bulletin,* 1973, *80,* 345–366.

Walker, D. N. Openness to touching: A study of strangers in nonverbal interaction. (Doctoral dissertation, University of Connecticut, 1971). *Dissertation Abstracts International,* 1971, *32,* 574. (University Microfilms No. 71-18, 454)

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1971.

Received January 6, 1978

Journal of Personality and Social Psychology
1979, Vol. 37, No. 1, 97-109

Effects of Message Repetition and Position on Cognitive Response, Recall, and Persuasion

John T. Cacioppo
University of Notre Dame

Richard E. Petty
University of Missouri—Columbia

Although the mere exposure effect has been researched widely, surprisingly little is known about the attitudinal and cognitive effects of message repetition. It was hypothesized that the sequence of topic-relevant thoughts generated in response to a (repeated) persuasive message would parallel attitude change. To test this prediction, two experiments were conducted. In Experiment 1, individuals heard a communication either zero (control), one, three, or five times in succession, rated their agreement with the advocated position, and listed the message arguments they could recall. In Experiment 2, individuals heard a communication either one, three, or five times, rated their agreement, listed their thoughts, and listed the message arguments they could recall. In both experiments, agreement first increased, then decreased as exposure frequency increased (regardless of the position advocated), but agreement was unrelated to the recall of the message arguments. In Experiment 2, analyses of the listed thoughts revealed that counterargumentation decreased, then increased, whereas topic-irrelevant thinking increased as exposure frequency increased; as expected, only topic-relevant thoughts were related to agreement. These results are interpreted in terms of an attitude-modification model in which repetition and content of a persuasive advocacy affect the type and number of thoughts generated; these thoughts, in turn, affect the attitudinal reaction to the advocacy.

In this article, we will consider the attitudinal effects of repeated exposure to persuasive communications, an area that has generated surprisingly little research by social psychologists despite its frequent occurrence in and importance to everyday life. This area is not well understood (cf. Harrison, 1977), in part because most research has focused on repetition in contexts that do not involve communication. In Zajonc's (1968) original statement on the effects of repeated exposure, evidence was provided for relationships between (a) the frequency of usage and evaluation of words, (b) the frequency of interpersonal contact and attraction, and (c) the familiarity of aesthetic stimuli (e.g., musical selections) and liking. Since then, immediate effects of mere exposure have been observed using children as well as adults (Heingartner & Hall, 1974), employing between-subjects (Moreland & Zajonc, 1976) as well as within-subjects (Crandall, 1972) designs, in field (Zajonc & Rajecki, 1969) as well as laboratory (Matlin, 1970) settings, employing pictorial magazine advertisements (McCullough & Ostrom, 1974) as well as nonsense syllables (Harrison & Hines, 1970) as stimuli, and using stimuli evaluated initially either positively or negatively (Hamm, Baum, & Nikels, 1975; Zajonc, Markus, & Wilson, 1974).

However, conditions in which the exposure–liking relationship breaks down have also been identified. When the stimulus is simple, repetition leads to either a decrease in liking (Skaife, cited in Berlyne, 1971, pp. 191–194) or an initial increase, then decrease in liking (Saegert & Jellison, 1970; Smith & Dorfman, 1975). The homogeneity of the stimulus sequence presented seems also to be an important factor. A stimulus that is presented

Requests for reprints should be sent to John T. Cacioppo, Department of Psychology, University of Notre Dame, Notre Dame, Indiana 46556.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3701-0097$00.75

repeatedly within a homogeneous sequence of stimuli results typically in first increasing, then decreasing liking of the stimulus (Harrison & Crandall, 1972). Moreover, when subjects are able to expose themselves to stimuli, they first display exploratory behavior and then expose themselves repeatedly to a self-selected subset of best-liked stimuli (Brickman & D'Amato, 1975). And finally, when measurement is administered immediately after the presentations of a stimulus, exposure effects may be attenuated or obliterated (Johnson & Watkins, 1971; Stang, 1974; Stang & O'Connell, 1974).

Models of Exposure Frequency and Affect

Several theoretical accounts have been offered for these results, including (a) response competition, (b) arousal theories, (c) classical conditioning, (d) intuition as artifact, (e) expectancy arousal, (f) satiation/generation, and (g) two-factor theories (see Harrison, 1977, and Stang, 1973, for reviews). Of these interpretations, the two-factor theories provide the most flexible explanation because they "can account for any pattern of results (by drawing differentially on each factor)" (Harrison, 1977, p. 74). However, two-factor theories are not cogent unless the contribution of each factor can be described a priori.

Berlyne (1970) proposed an inverted-U relationship between familiarity and liking. The notions are that (a) two separate and opposing psychological processes, positive habituation (that is, a reduction in uncertainty or conflict) and tedium, operate simultaneously; (b) the relative strengths of each vary as a function of exposure to the stimulus; and (c) initially, the process of habituation has greater impact on liking than does tedium. Thus, repeated exposure leads initially to liking, but ultimately leads to disliking. Accordingly, stimulus complexity and sequence heterogeneity slow the positive habituation process and extend the inflection point of the inverted-U curve to higher levels of exposure.

Stang (1973, 1975) proposed an extension of Berlyne's (1970) two-factor account: Repeated exposure provides more opportunity to learn about the stimulus; this learning is presumably rewarding and leads to increased liking for the stimulus. With continued repetition of the stimulus, however, boredom or satiation develops; hence, repeated exposure leads ultimately to negative affect toward the stimulus.

Stang (1975) presented three experimental demonstrations that affect toward and learning of Turkish words and trigrams were affected similarly by exposure frequency. Although these results may indicate that learning mediated liking, there are three other possibilities. First, liking may have mediated learning rather than vice versa. Second, both learning and affect may have been mediated by a third variable. For instance, the cognitive elaboration of stimuli results in both enhanced recall (Craik & Lockhart, 1972; Craik & Tulving, 1975; Rogers, Kuiper, & Kirker, 1977; Cacioppo & Petty, Note 1) and polarized affect (Petty & Cacioppo, 1977; Petty, Wells, & Brock, 1976; Tesser & Conlee, 1975; Tesser & Leone, 1977). Stimulus learning, however, has not been related consistently to affective reactions (cf. Greenwald, 1968; Petty, 1977). Nevertheless, repeated exposure does provide the opportunity for more elaborate processing of the stimuli (Wyer, 1974, pp. 15–16). Thus, there is the distinct possibility that cognitive elaboration (or extent of semantic processing) mediates both learning and affect.

This latter interpretation may appear at odds with some of the existing research on mere exposure. Specifically, preexposure liking for the stimulus does not necessarily alter the direction or form of the exposure effect (Zajonc et al., 1974). But in the domain of persuasive communication, Petty et al. (1976) demonstrated that it is not the preexposure affect toward an advocacy but the nature of the associates (i.e., cognitive responses) elicited by the message that determines the attitudinal effect of a communication. Indeed, a slightly different form of their hypothesis has been employed in past studies of mere exposure as well (Brickman, Redfield, Harrison, & Crandall, 1972; Mitchell & Olson, 1977; Pearlman & Oskamp, 1971). For example, repeated presentations of words that elicited

either positive or negative associations resulted in a polarization of affect toward the stimuli (Grush, 1976). Finally, the third possibility is that although learning and affect show similar patterns, they are mediated *independently*.

Cognitive and Attitudinal Effects of Message Repetition: Two Studies

The current experiments focus on the relationships among attitudinal, associative, and learning effects of message repetition. These experiments differ in several important respects from previous research on message repetition and attitudes. McCullough and Ostrom (1974) found an exposure effect when they presented messages repeatedly in a magazine advertisment format to subjects. However, each presentation differed in the phrasing and ordering of the message arguments, and a different photograph and headline accompanied each presentation. Weiss (1971) presented the same argument repeatedly and found that exposure led to quicker ratings of agreement with the argument. However, to the extent that practiced responses are also quicker, Weiss's results may be irrelevant to the study of message repetition and attitudes. Finally, Wilson and Miller (1968) and Johnson and Watkins (1971) found an attitudinal effect for message repetition only on a delayed posttest. The absence of the immediate attitudinal effect is due presumably to satiation (cf. Harrison, 1977; Sawyer, in press; Stang, 1974). Thus, if a wider variety of levels of message repetition, more complex stimuli, or a more heterogeneous stimulus sequence had been studied, immediate attitudinal effects might have been observed. In the present studies, Experiment 1 was primarily methodological and exploratory in nature, serving as a test of the experimental materials and of the effects of the repetition of persuasive communications on liking and learning. Experiment 2 was designed to assess the viability of the various possible processes mediating the liking and learning effects in repetition experiments using such stimuli.

To explore more fully the predictive value of two-factor theories, experimental conditions were designed to elicit first increasing, then decreasing favorability toward the advocacy. Mere exposure research indicates that the inflection of the inverted-U function between exposure and affect occurs more quickly when stimuli are homogeneous (Harrison & Crandall, 1972) and simple (e.g., Smith & Dorfman, 1975) and ratings are made immediately after their presentation (Stang, 1974). In the present experiments, one of two persuasive communications was presented either one, three, or five times in succession, and the dependent measures were administered immediately afterwards. We expected to find that agreement with the advocacy would increase, then decrease as exposure frequency increased.

Experiment 1

Method

Subjects and procedure. One hundred thirty-three introductory psychology students participated in an experiment employing a 2 × 4 factorial design in which the position advocated (proattitudinal versus counterattitudinal) and the number of presentations (0, 1, 3, or 5) served as between-subjects factors. Subjects were tested in groups of 12 to 29 in language-laboratory cubicles constructed so that no subject could have visual or verbal contact with any other subject. During any one session, in which one level of repetition was presented, half of the subjects heard the proattitudinal message over the headphones, while half of the subjects heard the counterattitudinal message over the headphones. Following a procedure employed previously by Petty and Brock (1976), the same highly persuasive arguments, which were equally applicable to each advocacy, were used in each communication. Thus, the affective qualities of the elicited associates (Grush, 1976) or cognitive responses (Greenwald, 1968; Petty & Cacioppo, 1977) were presumably equated for the two communications.

Two groups (the zero-exposure pro- and counterattitudinal conditions) rated their agreement with a recommendation that university expenditures be increased. One group was told that the expenditures were to be financed by instituting a 25% service tax on visitor luxuries (proattitudinal position), and one group was told that a $70-per-quarter increase in student tuition would be instituted to finance the expenditures (counterattitudinal position). Ratings were made on a 15-point Likert-type scale where 15 indicated "agree completely" and 1 indicated "disagree completely."

Subjects in each of the remaining cells heard a taped message through headphones either one, three, or five consecutive times. The message contained a

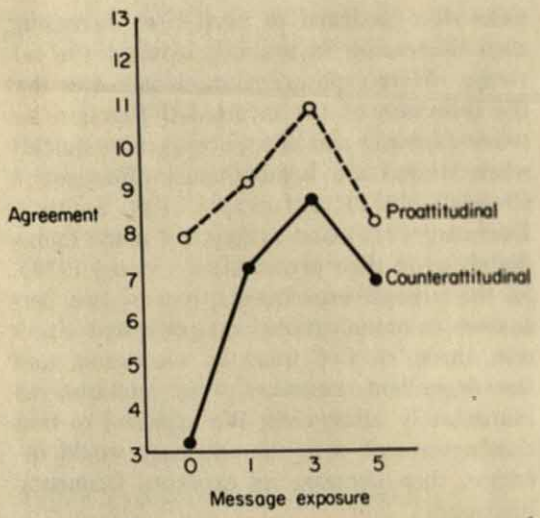


Figure 1. The effects of preexposure position and message repetition on agreement: Experiment 1.

statement that the university currently provided a substandard level of education to the undergraduates and that this problem could be remedied by increasing university expenditures. Half the subjects heard a message that contained a paragraph about instituting a 25% service tax on visitor luxuries to finance the expenditures, and half heard a message that contained a paragraph about instituting a student tuition increase of $70 per quarter to finance the expenditures. All subjects then heard the following eight arguments in favor of increasing expenditures: (a) Teaching and educational materials for students would be improved; (b) classroom sizes would be reduced; (c) handouts containing the important points in lectures could be provided; (d–f) the library facilities, job placement, and placement in graduate and professional schools could be improved; (g) better teachers could be hired; and (h) improved counseling to students, on both academic and personal matters, could be provided.

A postexperimental questionnaire containing 15-point Likert-type scales was used to assess the effects of the advocated position and message repetition on agreement with increasing university expenditures and to assess the perceived amount of distraction, effort, and involvement in the task. A recall measure was obtained after completion of the questionnaire by asking subjects to list on the last page of their booklets all of the message arguments that were presented in the communication. A judge, blind to the experimental conditions, counted an item as recalled if it correctly summarized one of the eight message arguments listed above. After completing the dependent-variable booklets, subjects were debriefed, thanked, and dismissed.

Results and Discussion

Attitude measure. The means for each cell on the attitude measure are displayed in Figure 1. It was expected that (a) subjects would agree more with the pro- than with the counterattitudinal advocacy, and (b) repetition would lead to increasing, then decreasing agreement with the advocacy. The analyses supported both hypotheses: Subjects agreed more with the pro- ($M = 9.05$) than the counterattitudinal message ($M = 6.62$), $F(1, 125) = 14.87$, $p < .001$; the number of presentations of the message affected agreement ($M_0 = 5.58$, $M_1 = 8.23$, $M_3 = 9.83$, $M_5 = 7.70$), $F(3, 125) = 7.74$, $p < .001$; and the interaction between position advocated and number of presentations did not approach significance ($p > .25$). Because the lack of an interaction indicated that the position of the advocacy (pro or counter) did not alter the effect of message repetition on agreement, trend tests were conducted on the data collapsed across position. The trend analyses were conducted according to Gaito's (1965) procedure for determining trend coefficients for unequal spacing and unequal n. A significant quadratic trend on the agreement measure, $F(1, 125) = 25.45$, $p < .001$, provided evidence of a curvilinear effect of message repetition on agreement (see Figure 1). Pairwise comparisons using the Newman-Keuls procedure for unequal n (Winer, 1971, pp. 215–218) revealed that presentations of the same persuasive message one, three, or five times led to increasing, then decreasing agreement ($ps < .05$).

Recall and ancillary measures. An analysis of the number of message arguments recalled revealed that message repetition, $F(2, 79) = 12.04$, $p < .001$, and message position, $F(1, 79) = 14.27$, $p < .001$, altered the incidental learning of the message arguments. The first main effect indicated that learning increased with repetition. The mean number of arguments recalled at one, three, and five repetitions, respectively, was 4.08, 4.67, and 6.00. The second main effect indicated that subjects recalled more message arguments when the position advocated was counterattitudinal ($M = 5.54$) than when the position advocated was proattitudinal ($M = 4.29$). Compared to previous research on selective learning (cf. Greenwald & Sakumara, 1967), Experiment 1 had the following methodological advantages:

(a) Subjects with initially similar attitudes were assigned randomly to pro- and counterattitudinal conditions rather than selecting subjects for participation on the basis of their initially different attitudes; (b) the message arguments to be learned by the different groups of subjects were identical, though they were used to support either a pro- or a counterattitudinal advocacy. Thus, the differential learning could not be attributed to the ease with which the arguments could be learned.

In addition, a Message Position × Message Repetition interaction was found, $F(2, 79) = 5.10$, $p < .01$. Specifically, the increase in recall was greatest between three and five repetitions of the proattitudinal advocacy ($p < .05$) and between one and three repetitions of the counterattitudinal advocacy (*ns*—all comparisons made using the Newman-Keuls procedure for unequal *n*).

The first experiment provided strong support for the notion that a persuasive appeal presented repeatedly in close temporal proximity leads to increasing, then decreasing acceptance of the advocacy. However, the finding that more message arguments were recalled when the advocacy was disliked initially (i.e., counterattitudinal) than when it was liked initially (i.e., proattitudinal) suggests that liking does not necessarily result in greater learning. The results of Experiment 1 are also at odds with the learning-leads-to-liking hypothesis. This is especially dramatic for the proattitudinal advocacy: The greatest increase in affect (between one and three repetitions) was associated with no change in learning, whereas the greatest learning (between three and five repetitions) was associated with a decrease in affect. A within-cell correlation (Insko, Lind, & La Tour, 1976, p. 69) between agreement and recall indicated that these variables were not related significantly ($r = -.02$).

Analyses of the ancillary measures (effort, involvement, and distraction) revealed that they were not affected by any of the manipulations.

Experiment 2

A second experiment was conducted to investigate other possible mediators of the exposure effect observed in the preceding experiment. Of particular interest in Experiment 2 was the possibility that a recipient's cognitive responses would be influenced by message repetition and would predict more accurately the subsequent attitudinal reactions than would learning. Indeed, evidence already exists for just such a hypothesis. Grush (1976) investigated the effects of repeatedly presenting infrequently used words that elicited either positive or negative associations. He found that words that elicited positive associations were evaluated more positively after repeated exposures, but that the opposite effect occurred for words that elicited negative associations. Grush proposed an attitude-formation explanation of his results. The notion is that a stimulus initially elicits only a few cognitive responses (associations). With repeated exposure, subjects generate increased (or increasingly consistent) cognitive responses to the stimulus (thus, it is similar in some respects to the response competition interpretation). The final evaluation of an exposed stimulus is assumed to be a function of the summed evaluations of the cognitive responses it elicits. If the responses are generally favorable, increased exposure should lead to a more positive evaluation of the stimulus, but if the responses are generally negative, increased exposure should lead to a less favorable evaluation of the stimulus.

Miller (1976) investigated the effects of mere exposure using communicative stimuli (posters containing a political message) and found that moderate but not high levels of exposure led to attitude change (increased agreement). Miller argued that the high exposure loads caused individuals to feel that their personal freedom was being restricted, and reactance (Brehm, 1966) led to resistance to persuasion. It is equally possible, on the other hand, that an unpleasant state of satiation (e.g., boredom) developed at the high exposure levels. These models suggest that the cognitive responses (i.e., counterarguments, favorable thoughts, neutral/irrelevant thoughts) elicited in a repetition experiment may mediate the affective reactions to the stimulus. However, no study to date that has employed repetitions of the same persuasive

communications has investigated the nature or the temporal sequence of the cognitive responses elicited. Thus, in Experiment 2 subjects heard the previously tested pro- or counterattitudinal advocacy either one, three, or five times in succession. Afterwards, subjects (a) were instructed to list everything about which they had thought during the preceding minutes, (b) rated their attitude toward the advocacy and completed ancillary measures, and (c) responded to a measure of the incidental learning of message arguments.

Method

One hundred ninety-three introductory psychology students participated in an experiment employing a 2 × 3 factorial design in which the position advocated (pro- versus counterattitudinal) and the number of presentations (1, 3, or 5) served as between-subjects factors.[1] As in Experiment 1, subjects were tested in groups of 12 to 29 in cubicles in a language laboratory. During any one session, in which one level of repetition was presented, half the subjects heard the proattitudinal and half the subjects heard the counterattitudinal message through headphones. Subjects rated their agreement with increasing university expenditures on a 15-point Likert-type scale and completed the postexperimental questionnaire. In a procedure adapted from Brock (1967) and Greenwald (1968) and employed previously by Petty and Cacioppo (1977), subjects were given 3 minutes to list the actual thoughts that occurred to them during the presentations of the communication they had heard. Subjects were then asked to go back through the thoughts they had listed and to place a plus (+) next to each thought that supported the advocacy, a minus (−) next to each thought that attacked the advocacy, and a zero (0) next to each thought that was neutral or irrelevant to the advocacy. Subjects were then asked to list all of the message arguments they could recall. (Recall was scored in the same manner as in Experiment 1.)

Two judges who were blind to the experimental conditions rated the cognitive responses. Counted as unfavorable thoughts (i.e., counterarguments) were statements directed against the advocated position that mentioned specific unfavorable consequences, statements of alternative methods of raising money, challenges to the validity of arguments in the message, and statements of affect opposing the advocated position. Counted as favorable thoughts were statements in favor of the advocated position that mentioned specific favorable consequences, statements ruling out alternatives, statements that supported the validity of the message arguments, and statements of affect supporting the advocated position. All other listed items were rated as neutral/irrelevant thoughts. Similar items were counted as one thought. Judges agreed on over 95% of the ratings; disagreements were resolved through discussion.

Results

Attitude measure. The mean for each cell on the attitude measure is graphed in Figure 2. As in Experiment 1, subjects agreed more with the pro- ($M = 9.61$) than the counterattitudinal advocacy ($M = 8.19$), $F(1, 187) = 6.48$, $p < .02$, and subjects agreed differentially as a function of the number of message repetitions ($M_1 = 8.09$, $M_3 = 9.77$, $M_5 = 8.86$), $F(2, 187) = 3.04$, $p < .05$. The interaction again did not approach significance ($p > .20$). Trend tests were conducted using Gaito's (1965) procedure for unequal n and revealed that message repetition resulted in a significant quadratic effect, $F(1, 187) = 9.86$, $p < .001$. Pairwise comparisons of cell means using Newman-Keul's procedure for unequal n provided evidence that agreement with the advocacy increased ($p < .05$), then decreased only marginally ($p < .10$) as exposure frequency increased.

Cognitive-response measures. The means for each cell on the cognitive-response measures are graphed in Figure 2. Overall analyses of variance revealed that subjects generated more counterarguments, $F(1, 187) = 10.71$, $p < .001$, and fewer neutral/irrelevant thoughts, $F(1, 187) = 4.35$, $p < .05$, in response to the counterattitudinal than in response to the proattitudinal advocacy. The number of message repetitions marginally affected the production of favorable, $F(2, 187) = 2.36$, $p < .10$, and neutral/irrelevant, $F(2, 187) = 2.61$, $p < .08$, thoughts. There were

[1] An additional factor included in the design was whether subjects were informed as to how many times they would hear the message prior to exposure. That is, half of the subjects were told that they were to hear a taped message one, three, or five times (the number being determined by the number of times they actually heard the message) so that they might have time to consider completely the taped communication. The other half of the subjects were told nothing about the number of times the communication would be presented. Analyses revealed that this factor made no difference on any of the measures obtained, and it will not be discussed further.

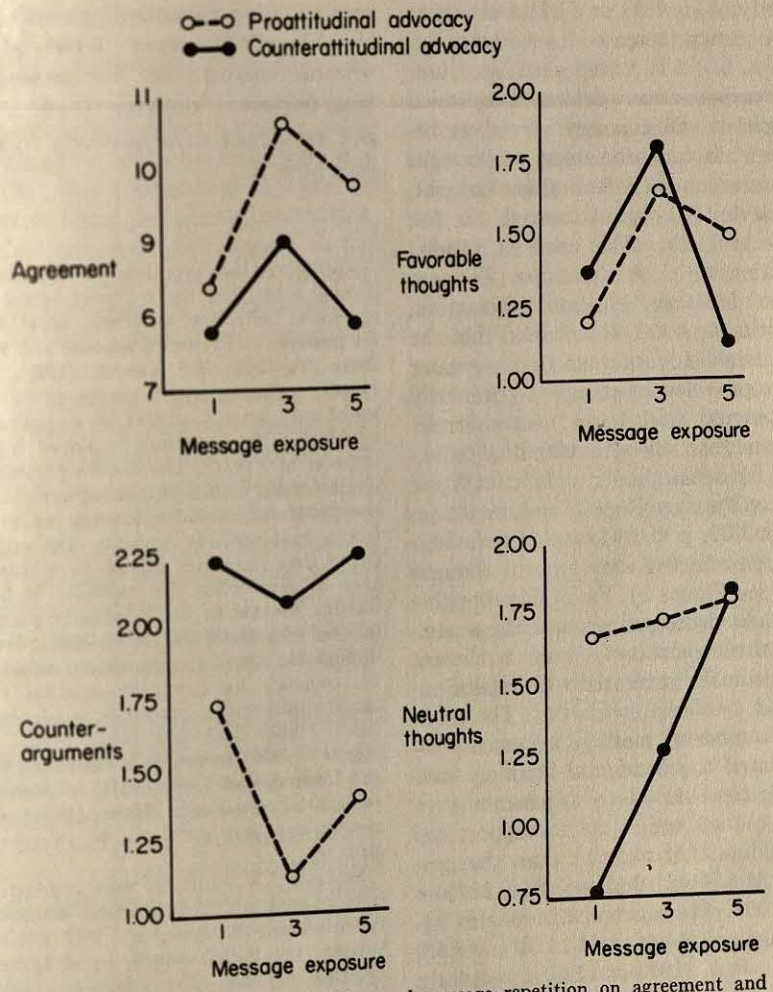


Figure 2. The effects of preexposure position and message repetition on agreement and cognitive response: Experiment 2.

no significant interactions; thus, trend contrasts were computed on the data collapsed across message position. The results of trend contrasts on each cognitive-response measure (employing Gaito's procedure for unequal n) indicated that counterargumentation was affected quadratically, $F(1, 187) = 15.37$, $p < .001$, and neutral/irrelevant thinking was af-

Table 1
Within-Cell Correlations: Experiment 2

Variable	Agreement	Counter-arguments	Favorable thoughts	Neutral thoughts
Counterarguments	−.56			
Favorable thoughts	.45	−.45		
Neutral thoughts	.03	−.28	−.16	
Recall	−.01	.17	−.07	.02

fected linearly, $F(1, 187) = 5.87$, $p < .02$, as exposure frequency increased (see Figure 2).

Finally, a $2 \times 3 \times 3$ repeated measures analysis of variance was performed in which message position and exposure served as between-subjects factors and type of thought (i.e., counterargument, favorable thought, neutral/irrelevant thought) served as the within-subjects factor. The analysis yielded two interactions: (a) A significant Type of Thought $\times$ Message Position interaction, $F(2, 370) = 6.14$, $p < .003$, revealed that the counterattitudinal advocacy elicited a greater number of topic-relevant thoughts (primarily counterarguments) and fewer neutral/irrelevant thoughts than the proattitudinal advocacy; and (b) a marginally significant Type of Thought $\times$ Message Repetition interaction, $F(4, 370) = 2.02$, $p < .09$, showed that message repetition affected each type of thought differently (see Figure 2). Favorable thoughts increased, then decreased; counterarguments decreased, then increased; and irrelevant thoughts continually increased with repetition.

Recall and ancillary measures. The analysis of the number of message arguments recalled replicated the incidental learning finding in Experiment 1: More arguments were recalled when they were used to support the counterattitudinal ($M = 5.51$) than the proattitudinal ($M = 4.48$) position, $F(1, 187) = 14.63$, $p < .001$. Message repetition also affected message recall ($M_1 = 4.14$, $M_3 = 5.50$, $M_5 = 5.36$), $F(2, 187) = 11.49$, $p < .001$; the interaction, however, was not significant this time. Trend analyses (collapsed across message position) yielded significant Fs for both the linear, $F(1, 187) = 14.14$, $p < .001$, and quadratic, $F(1, 187) = 15.01$, $p < .001$, trends.

The analyses of the ancillary measures revealed again that neither the position advocated nor the number of presentations affected any of these ratings.

Within-cell correlations. A within-cell correlation between each pair of dependent measures was calculated to provide measures of association between the variables with the effects of the treatments held constant. The results of these analyses are presented in Table 1. Note that agreement and topic-relevant thinking (counterarguments and favorable thoughts) were interrelated highly, whereas learning of the message (recall), topic-irrelevant thinking (neutral thoughts), and agreement were relatively independent of each other.[2]

---

[2] Although the hypothesized attitude-modification process holds that the cognitive responses mediated agreement with the advocacy, other causal chains are possible. Following Osterhouse and Brock (1970), Insko, Turnbull, and Yandell (1974), Petty et al. (1976), and Petty and Cacioppo (1977), four analyses of covariance (ANCOVA) were conducted. The rationale for the procedure employed is discussed by Insko et al. (1974). The ANCOVA procedure compares specific causal models by holding constant statistically the postulated mediator between an initial variable and a final criterion variable. The relationship between these measures should be within error variance of zero when the mediator is held constant through the use of the covariance procedure; a reduction of a significant F for the criterion measure through the use of this procedure would suggest that the covariate mediated the criterion variable. The causal-model-testing procedure is adapted from Blalock's (1964) technique using partial correlations, Scheffé's (1959) technique of examining partial slopes, and Cochran and Cox's (1957) technique employing analysis of covariance. Heise (1969) discusses the assumptions and problems of causal-model-testing analyses.

Separate 2×3 ANCOVAs were conducted using (a) counterarguments, (b) favorable thoughts, (c) neutral/irrelevant thoughts, and (d) recall as the covariate (i.e., initial variable) and agreement as the criterion variable. The originally significant F of 3.02 for the repetition main effect on agreement was reduced to Fs < 1 when counterarguments and favorable thoughts served as the covariate. When neutral/irrelevant thoughts served as the covariate, the F of 3.02 was reduced to a marginally significant F of 2.41; when recall served as the covariate, the F for agreement remained significant at 11.24. When these analyses were repeated with agreement serving as the covariate, the marginally significant F of 2.36 for the repetition main effect on favorable thoughts was reduced slightly to 1.78; the marginally significant F of 2.61 on neutral/irrelevant thoughts became a significant F of 6.04; the nonsignificant F of 1.57 on counterarguments remained a nonsignificant F of .68; and the highly significant F of 11.49 on recall remained a significant F of 5.80. Interpretation of these analyses should be cautious, since the reliabilities of the measures are unknown. Nevertheless, the results are consistent with the notion that topic-relevant thinking mediates the effect of repetition on attitude, and they are inconsistent with the learning-leads-to-liking hypothesis.

Discussion

Although various research, including Experiment 2, has demonstrated that repetition influences learning and affect similarly, additional evidence provided in Experiments 1 and 2 suggests that this relationship in mere exposure research using persuasive communications is neither the result of learning leading to liking nor of liking leading to learning. Specifically, the within-cell correlations between learning and agreement in the experiments were near zero and nonsignificant. Additionally, the analyses of covariance (Footnote 2) suggested that although learning did not mediate agreement, the cognitive responses elicited by the communications did. Of course, it is possible that the learning-leads-to-liking hypothesis in mere exposure research applies only to stimuli that possess few cognitive associations (e.g., nonsense stimuli). The present experiments provide no evidence from which to evaluate this hypothesis.

The present experimental results indicated that regardless of the position advocated, message repetition led to (a) increasing, then decreasing agreement with the advocacy; (b) decreasing, then increasing counterargumentation; and (c) increasing topic-irrelevant thinking. Of importance is that the analyses of variance revealed that the sequence of the topic-relevant thoughts generated in response to the repeated messages paralleled the observed attitude change. It was hypothesized that the effects of message repetition on agreement were mediated by a two-stage attitude-modification process in which the repetition of the message arguments provided more opportunities to elaborate cognitively upon them and to realize their cogency and favorable implications. Hence, counterargumentation declined at the moderate exposure frequency. At high exposure levels, however, tedium and/or reactance may have motivated the individual to again attack the now offensive communication. Thus, counterargumentation was renewed, and agreement decreased at high exposure levels.

Validity of thought listings. It might be argued that the thought listings do not provide evidence about the nature of the subjects' thoughts, but instead reflect what subjects thought the experimenter wanted them to list (Orne, 1962). This interpretation is considered unlikely in the present case because (a) the experimenter clearly *requested* a list of everything about which subjects had been thinking during the message; (b) although a complex pattern of counterargumentation was produced (first decreasing, then increasing), reports of topic-irrelevant thinking increased as exposure frequency increased; and (c) a between-subjects design was employed, making unlikely the successful intuiting of the experimenter's hypotheses. Furthermore, the position of the advocacies did not alter the exposure effects on the listed thoughts and attitude; although this result replicates previous research on mere exposure (Zajonc et al., 1974), it is unlikely that such an effect is common knowledge to naive subjects.

It might also be argued that subjects are unable to provide veridical information about the nature of their thoughts. Cacioppo, Harkins, and Petty (in press) and Cacioppo and Petty (Note 2) have argued that subjects are able to report accurately their recent and current thoughts and ideas and that the fact that subjects may be unaware of the *reasons for*, or *consequences of*, their thoughts (cf. Nisbett & Bellows, 1977; Nisbett & Wilson, 1977) does not invalidate the use of their responses in a content analysis. A thought-listing analysis assumes only that the thoughts and ideas of subjects are as accessible as are attitudes, judgments, facts, and so on (cf. Wyer, 1974). The burden then lies with researchers to determine whether a subject's cognitive responses are predictive of, or influential in, affect and behavior. It should be noted that although the present results are consistent with the hypothesized attitude-modification process and inconsistent with the learning-affect model, no definitive evidence concerning the causal role of cognitive responses is provided. Future research might fruitfully employ distraction to attenuate the dominant type of thought generated and/or vary message-argument quality to differentiate the nature of the dominant thought elicited.

Cognitive dynamics of message repetition. We have proposed that message repetition guides a sequence of cognitive reactions to a persuasive communication. But are the repeated presentations of the stimulus per se responsible for this sequence? We think not. It should be apparent that in the present experiments, the number of repetitions of a message was confounded with the available time to think about the message arguments, generate new topic-relevant thoughts, and so forth. This (additional) time to think has been shown to be necessary for individuals to process more deeply and elaborate more fully the content of an impending communication (Petty & Cacioppo, 1977), the persuasive arguments for a group decision (Burnstein & Vinokur, 1975, 1977), and the qualities of a stimulus (Tesser & Leone, 1977). For example, Tesser and his colleagues (e.g., Sadler & Tesser, 1973; Tesser & Conlee, 1975) have demonstrated that if persons are asked to think about some object or issue, the amount of time spent thinking is related to the amount of attitude polarization that results. Thus, objects that are liked initially are evaluated more favorably with increased thought, and initially disliked objects are evaluated more negatively. If we had simply presented subjects with the topics of the pro-attitudinal (increasing expenditures by instituting a visitor's tax) and counterattitudinal (increasing expenditures by increasing tuition) advocacies, *without the accompanying persuasive arguments,* we expect that we would have replicated the attitude polarization effect: More time to think would have produced increased agreement with the pro-attitudinal and decreased agreement with the counterattitudinal advocacy. However, we found that moderate repetitions of (and time to think about) the persuasive message arguments led to increased agreement with both the pro- and counterattitudinal advocacies. This finding supports the notion that the first stage of the attitude-modification process of message repetition involves enhanced opportunities to process the *content* of the message (which was the same for both pro- and counterattitudinal groups). Moreover, we found that high exposure frequencies of (and lengthier time to think about) the message arguments resulted in decreased agreement with the advocacy. This result is consistent with the notion that the second stage of the attitude-modification process involves the development of tedium and/or reactance, which modifies the subsequent information-processing activity.

We assumed that individuals would be motivated to process personally involving stimuli (cf. Cialdini, Levy, Herman, Kozlowski, & Petty, 1976); thus, we employed communicative stimuli that were personally relevant to the individuals in the experiment. We also assumed that these individuals, if able, would be motivated to elaborate more fully the arguments contained in each advocacy. Thus, the provision of additional opportunities via message repetition to process the persuasive arguments contained in the advocacies would result in greater realizations of their validity, and not until the repetitions became excessive would this process yield to one motivated by reactance or tedium.

However, there are at least two other ways to account for the observed sequence of cognitive responses. First, since each of the arguments contained in the advocacies was in favor of an increase in spending, this general orientation may have served as a retrieval cue for initially accessing and generating topic-relevant cognitions that further supported an increase in spending. As the memorial pool of favorable thoughts was depleted, however, counterarguments would become increasingly accessible. Hence, according to this account, the *direction* of the message arguments is important rather than their cogency or persuasiveness.

A second plausible mechanism has been discussed by Burnstein and Vinokur (1975, 1977). The simple statement that others are advocating a position discrepant from one's own "may induce a person to reconstruct a line of reasoning which he thinks could have produced such (an advocacy)" (Burnstein, Vinokur, & Trope, 1973, p. 244). This model differs from Tesser's in that in this instance, the individual is motivated to consider, at least initially, all possible arguments that support the advocacy, since it has been made

salient to the individual that others endorse this view. Again, as the proargument pool in memory is exhausted, counterarguments become increasingly accessible. While this position is similar in many respects to our own, we place a greater emphasis on the cognitive responses elicited by the content and context of message repetition.

Cognitive dynamics and incidental recall. Finally, a word might be said about the incidental recall findings of Experiments 1 and 2. In each experiment, it was found that persons recalled more message arguments used to support a counterattitudinal than a proattitudinal position. Neither familiarity nor novelty can account for this effect, because the *same* message arguments were used to support each position. Nor can a consistency explanation account for this effect, because best remembered were the arguments used to support a position *inconsistent* with the beliefs of the subjects.

Perhaps a more cogent account of the incidental recall findings can be provided by an analysis of the extent of processing of the pro- and counterattitudinal advocacies. Accumulated research has supported Craik and Lockhart's (1972) proposal that the depth or extent of processing influences the recall of a stimulus (e.g., Craik & Tulving, 1975; Rogers et al., 1977; Cacioppo & Petty, Note 1). It is interesting to note in light of this evidence that as discovered in Experiment 2, the counterattitudinal advocacies elicited a greater number of *topic-relevant thoughts* than did the proattitudinal advocacies. This effect was obtained even though the message arguments comprising each advocacy were identical. If one assumes that the number of listed topic-relevant thoughts indicates the extent to which the individuals elaborated cognitively upon the message arguments, then superior recall of the message arguments supporting the counterattitudinal rather than the proattitudinal advocacy would be expected. The results are in accord with this analysis.

Reference Notes

1. Cacioppo, J. T., & Petty, R. E. Electrophysiologic concomitants of hot cognition. In F. J. McGuigan (Chair), *Overt measures of covert mental processes.* Symposium presented at the meeting of the American Psychological Association, Toronto, Canada, August 1978.

2. Cacioppo, J. T., & Petty, R. E. Inductive techniques for the assessment of cognitive processes. In M. Mahoney (Chair), *The assessment of cognitions.* Symposium presented at the National Conference on Cognitive Behavior Therapy Research, New York, October 1978.

References

Berlyne, D. E. Novelty, complexity, and hedonic value. *Perception & Psychophysics,* 1970, *8,* 279–286.

Berlyne, D. E. *Aesthetics and psychobiology.* New York: Appleton-Century-Crofts, 1971.

Blalock, H. M., Jr. *Causal inferences in nonexperimental research.* Chapel Hill: University of North Carolina Press, 1964.

Brehm, J. *A theory of psychological reactance.* New York: Academic Press, 1966.

Brickman, P., & D'Amato, B. Exposure effects in a free-choice situation. *Journal of Personality and Social Psychology,* 1975, *32,* 415–420.

Brickman, P., Redfield, J., Harrison, A. A., & Crandall, R. Drive and predisposition as factors in the attitudinal effects of mere exposure. *Journal of Experimental Social Psychology,* 1972, *8,* 31–44.

Brock, T. C. Communication discrepancy and intent to persuade as determinants of counterargument production. *Journal of Experimental Social Psychology,* 1967, *3,* 269–309.

Burnstein, E., & Vinokur, A. What a person thinks upon learning he has chosen differently from others: Nice evidence for the persuasive-arguments explanation of choice shifts. *Journal of Experimental Social Psychology,* 1975, *11,* 412–426.

Burnstein, E., & Vinokur, A. Persuasive argumentation and social comparison as determinants of attitude polarization. *Journal of Experimental Social Psychology,* 1977, *13,* 315–332.

Burnstein, E., Vinokur, A., & Trope, Y. Interpersonal comparison versus persuasive argumentation: A more direct test of alternative explanations for group-induced shifts in individual choice. *Journal of Experimental Social Psychology,* 1973, *9,* 236–245.

Cacioppo, J. T., Harkins, S. G., & Petty, R. E. The nature of cognitive responses to persuasive appeals. In R. E. Petty, T. M. Ostrom, & T. C. Brock (Eds.), *Cognitive responses in persuasion.* Hillsdale, N.J.: Erlbaum, in press.

Cialdini, R. B., Levy, A. H., Herman, C. P. Kozlowski, L. T., & Petty, R. E. Elastic shifts of opinion: Determinants of direction and durability. *Journal of Personality and Social Psychology,* 1976, *34,* 663–672.

Cochran, W. G., & Cox, G. M. *Experimental designs* (2nd ed.). New York: Wiley, 1957.

Craik, F. I. M., & Lockhart, R. S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior,* 1972, *11,* 671–684.

Craik, F. I. M., & Tulving, E. Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 1975, *104*, 268–294.

Crandall, R. Field extension of the frequency–affect findings. *Psychological Reports*, 1972, *31*, 371–374.

Gaito, J. Unequal intervals and unequal *n* in trend analysis. *Psychological Bulletin*, 1965, *63*, 125–127.

Greenwald, A. G. Cognitive learning, cognitive response to persuasion, and attitude change. In A. G. Greenwald, T. C. Brock, & T. M. Ostrom (Eds.), *Psychological foundations of attitudes*. New York: Academic Press, 1968.

Greenwald, A. G., & Sakumura, J. Attitude and selective learning: Where are the phenomena of yesteryear? *Journal of Personality and Social Psychology*, 1967, *7*, 387–397.

Grush, J. E. Attitude formation and mere exposure phenomena: A nonartifactual explanation of empirical findings. *Journal of Personality and Social Psychology*, 1976, *33*, 281–290.

Hamm, N., Baum, M. R., & Nikels, K. W. Effects of race and exposure on judgments of interpersonal favorability. *Journal of Experimental Social Psychology*, 1975, *11*, 14–24.

Harrison, A. A. Mere exposure. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10). New York: Academic Press, 1977.

Harrison, A. A., & Crandall, R. Heterogeneity–homogeneity of exposure sequence and the attitudinal effects of exposure. *Journal of Personality and Social Psychology*, 1972, *21*, 234–238.

Harrison, A. A., & Hines, P. Effects of frequency of exposure at three short times on affective ratings and exploratory behavior. *Proceedings of the 78th Annual Convention of the American Psychological Association*, 1970, *5*, 391–392. (Summary)

Heingartner, A., & Hall, J. Affective consequences in adults and children of repeated exposure to auditory stimuli. *Journal of Personality and Social Psychology*, 1974, *29*, 719–724.

Heise, D. R. Problems in path analysis and causal inference. In E. F. Borgatta (Ed.), *Sociological methodology*. San Francisco: Jossey-Bass, 1969.

Insko, C. A., Lind, E. A., & La Tour, S. Persuasion, recall and thoughts. *Representative Research in Social Psychology*, 1976, *7*, 66–78.

Insko, C. A., Turnbull, W., & Yandell, B. Facilitative and inhibiting effects of distraction on attitude change. *Sociometry*, 1974, *37*, 508–528.

Johnson, H. H., & Watkins, T. H. The effects of message repetition on immediate and delayed attitude change. *Psychonomic Science*, 1971, *22*, 101–103.

Matlin, M. W. Response competition as a mediating factor in the frequency–affect relationship. *Journal of Personality and Social Psychology*, 1970, *16*, 536–552.

McCullough, J. L., & Ostrom, T. M. Repetition of highly similar messages and attitude change. *Journal of Applied Psychology*, 1974, *59*, 395–397.

Miller, R. C. Mere exposure, psychological reactance, and attitude change. *Public Opinion Quarterly*, 1976, *40*, 229–233.

Mitchell, A., & Olson, J. Cognitive effects of advertising repetition. *Advances in Consumer Research*, 1977, *4*, 213–220.

Moreland, R. L., & Zajonc, R. B. A strong test of exposure effects. *Journal of Experimental Social Psychology*, 1976, *12*, 170–179.

Nisbett, R. E., & Bellows, N. Verbal reports about causal influences on social judgments: Private access versus public theories. *Journal of Personality and Social Psychology*, 1977, *35*, 613–624.

Nisbett, R. E., & Wilson, T. D. Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 1977, *84*, 231–259.

Orne, M. T. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 1962, *17*, 776–783.

Osterhouse, R. A., & Brock, T. C. Distraction increases yielding to propaganda by inhibiting counterarguing. *Journal of Personality and Social Psychology*, 1970, *15*, 344–358.

Pearlman, D., & Oskamp, S. The effects of picture content and exposure frequency on evaluations of negroes and whites. *Journal of Experimental Social Psychology*, 1971, *7*, 503–514.

Petty, R. E. *A cognitive response analysis of the temporal persistence of attitude change induced by persuasive communications*. Unpublished doctoral dissertation, Ohio State University, 1977.

Petty, R. E., & Brock, T. C. Effects of responding or not responding to hecklers on audience agreement with a speaker. *Journal of Applied Social Psychology*, 1976, *6*, 1–17.

Petty, R. E., & Cacioppo, J. T. Forewarning, cognitive responding, and resistance to persuasion. *Journal of Personality and Social Psychology*, 1977, *35*, 645–655.

Petty, R. E., Wells, G. L., & Brock, T. C. Distraction can enhance or reduce yielding to propaganda: Thought disruption versus effort justification. *Journal of Personality and Social Psychology*, 1976, *34*, 874–884.

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology*, 1977, *35*, 677–688.

Sadler, O., & Tesser, A. Some effects of salience and time upon interpersonal hostility and attraction during social isolation. *Sociometry*, 1973, *36*, 99–112.

Saegert, S. C., & Jellison, J. M. Effects of initial level of response competition and frequency of exposure on liking and exploratory behavior. *Journal of Personality and Social Psychology*, 1970, *16*, 553–558.

Sawyer, A. G. Repetition and cognitive response. In R. E. Petty, T. M. Ostrom, & T. C. Brock (Eds.), *Cognitive response in persuasion*. Hillsdale, N.J.: Erlbaum, in press.

Scheffé, S. *The analysis of variance*. New York: Wiley, 1959.

Smith, G. F., & Dorfman, D. D. The effect of stimulus uncertainty on the relationship between fre-

quency of exposure and liking. *Journal of Personality and Social Psychology*, 1975, *31*, 150–155.

Stang, D. J. Six theories of repeated exposure and affect. *JSAS Catalog of Selected Documents in Psychology*, 1973, *3*, 126. (Ms. No. 482)

Stang, D. J. Methodological factors in mere exposure research. *Psychological Bulletin*, 1974, *81*, 1014–1025.

Stang, D. J. The effects of mere exposure on learning and affect. *Journal of Personality and Social Psychology*, 1975, *31*, 7–13.

Stang, D. J., & O'Connell, E. J. The computer as experimenter in social psychological research. *Behavior Research Methods and Instrumentation*, 1974, *6*, 223–232.

Tesser, A., & Conlee, M. C. Some effects of time and thought on attitude polarization. *Journal of Personality and Social Psychology*, 1975, *31*, 262–270.

Tesser, A., & Leone, C. Cognitive schemas and thought as determinants of attitude change. *Journal of Experimental Social Psychology*, 1977, *13*, 340–356.

Weiss, R. F. Role playing and repetition effects on opinion strength. *Journal of Social Psychology*, 1971, *85*, 29–35.

Wilson, L. R., & Miller, H. Repetition, order of presentation, and timing of arguments and measures as determinants of opinion change. *Journal of Personality and Social Psychology*, 1968, *9*, 184–188.

Winer, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1971.

Wyer, R. S., Jr. *Cognitive organization and change: An information-processing approach*. Potomac, Md.: Erlbaum, 1974.

Zajonc, R. B. The attitudinal effects of mere exposure. *Journal of Personality and Social Psychology Monograph*, 1968, *9*(2, Pt. 2).

Zajonc, R. B., Markus, H., & Wilson, W. R. Exposure effects and associative learning. *Journal of Experimental Social Psychology*, 1974, *10*, 248–263.

Zajonc, R. B., & Rajecki, D. Exposure and affect: A field experiment. *Psychonomic Science*, 1969, *17*, 216–217.

Received January 30, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 1, 110-115

Effects of Controllable Versus Uncontrollable Factors on Responsibility Attributions: A Single-Subject Approach

Daniel Arkkelin, Thomas Oakley, and Clifford Mynatt
Bowling Green State University

A "lens-model" methodology was used to assess the influence of different informational "cues" on responsibility attributions for an automobile accident. Multiple regression analyses indicated that although there were large individual differences, in Experiments 1 and 2 the major determinants of responsibility judgments were car speed and the condition of the car's brakes. In Experiment 3, speeding and brakes information was deleted. The major determinant of attributions was the driver's past record, although the importance of severity of consequences was somewhat greater than in the first two experiments. In Experiment 4, driving record was deleted. The majority of subjects indicated that not enough information was available. Overall, severity was a relatively unimportant cue. Observers tended, instead, to rely on factors over which the driver had control. The lens-model approach seems quite suitable for studying attributional judgments, particularly since it provides a detailed description of individual judgment strategies.

Attribution research has typically utilized experimental designs in which only a limited number of attributional determinants are manipulated. Additionally, between-subjects statistical analyses are used in which individual differences are treated as "error variance." The present studies were conducted to explore an alternative approach that permits the assessment of the effects of several independent variables and the determination of the relative importance of each variable for individual subjects.

Fischhoff (1976) has pointed out that attribution theory is not the only area of psychological research concerned with inferential behavior. An attribution can be conceptualized as a special case of a general judgment process wherein individuals integrate information to arrive at judgments.

The authors would like to express their appreciation to Michael Doherty for his assistance throughout the project and for his comments on an early draft of the manuscript.

Requests for reprints should be sent to Daniel Arkkelin, who is now at the Department of Psychology, Wartburg College, Waverly, Iowa 50677.

Investigators of human judgment have used multiple regression analysis extensively to quantitatively describe the manner in which informational cues are weighted and combined by a judge to make a decision (Slovic & Lichenstein, 1971). Brunswik's (1952) lens model has served as the framework for many of these investigations (see Dudycha & Naylor, 1966; Hammond, Stewart, Brehmer, & Steinmann, 1975). Keeley and Doherty (1971) describe the typical paradigm in which a set of cues, X_i, are presented to a decision maker, whose response, Y_s, varies as a function of the cue values. The individual makes a large number of judgments in response to various combinations of the cue values. Multiple regression analysis is then applied to "capture the policy" of the judge, resulting in a regression equation that is a model of the judge's integrating and weighting strategy, or policy. Specifically, the regression weights are indexes of the extent to which the judge employs each cue.

Since the relative size of the regression weights is partially determined by their ordi-

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3701-0110$00.75

nal position in the regression equation when cues are intercorrelated, Darlington's (1968) "usefulness index" is commonly used to assess the relative importance of each cue. This index is the amount R^2 decreases if a given cue is deleted and the regression equation recalculated. The usefulness indexes can then be tested for statistical significance.

While this methodology can be applied to any area of attribution research, attribution of responsibility for an accident was chosen for the present investigations, since there are several unresolved questions concerning this topic. Several investigators have reported that observers tend to assign increasing responsibility to a target person as the severity of an accident increases (e.g., Chaiken & Darley, 1973; McKillip & Posavac, 1975; Shaver, 1970; Walster, 1966). Other studies have not replicated this effect (e.g., Shaver, 1970; Walster, 1967). Additionally, researchers have not reported the *magnitude* of the severity effect when it has occurred or its *relative* magnitude in comparison to other potential attributional determinants. The following experiments used the lens-model methodology to assess the relative influence of seven cues on individual observers' responsibility attributions.

Experiment 1

Method

Subjects. Ten students recruited from introductory psychology classes served as subjects.

Materials. Pilot subjects were asked to list what information they would need to determine a driver's responsibility for an automobile accident. The most frequently requested items were road conditions, traffic conditions, the driver's past record of accidents, car speed, and the condition of the brakes of the driver's car. In addition to these "cues," two others, severity of consequences to victims and severity of consequences to the driver, were included in the cue set.

Statements describing two levels of each cue were constructed (e.g., high or low severity of consequences, light or heavy traffic, etc.) resulting in 128 combinations of the cues. An example of one particular combination follows:

> While driving to work last week, Elaine W. was involved in an automobile accident. Elaine was unharmed, but the people in the other car were critically injured. The stretch of road on which the accident occurred was full of potholes, and there was very little traffic at the time of the accident. Elaine's record indicates that she has been in several accidents since she got her driver's license. It was determined that at the time of the accident she was not speeding and that the brakes on her car were in need of repair. On the basis of the above information, how responsible do you think Elaine should be held for the accident?

A 7-point scale was at the bottom of each profile, with the statement "not at all responsible" at the left extreme and "completely responsible" at the right extreme.

Since multiple regression analysis does not require that all possible cue combinations actually be presented to the subjects, 40 of the 128 profiles were randomly selected, and booklets containing the chosen profiles were constructed. The booklets contained the following frequencies of each level of each cue: severity-driver (high = 20, low = 20); severity-victim (high = 20, low = 20); road conditions (good = 19, poor = 21); traffic (light = 23, heavy = 17); driving record (good = 13, poor = 27); speed (excessive = 14, moderate = 26); brakes (good = 12, poor = 28).

Procedure. Each subject was given a booklet of profiles and a cover sheet indicating that each page contained a short story, with the pattern of information differing from page to page. Subjects were asked to make a judgment about the driver on the scale at the bottom of each page and instructed not to look back at their previous judgments.

Results

The numerical values (0 vs. 1) of the two levels of each cue were entered as predictor variables in a stepwise multiple regression analysis of each subject's 40 responsibility judgments. Table 1 presents the R^2 values ($M = .66$) and significant ($p < .05$) usefulness indexes (UI) for each subject.[1]

Inspection of this table reveals a marked reliance on the cues of speeding (9 of 10 subjects employed this cue to a significant extent, $M_{UI} = .31$) and brakes (9 subjects, $M_{UI} = .27$). Information regarding consequences to victims was relied upon to a significant extent by only one subject ($M_{UI} = .01$), and information about consequences to the driver was not employed by any subject.

Two conclusions are clearly warranted from Table 1: (a) Subjects' "policies" were

[1] Table 1 is included to provide an example of how data obtained with this technique are typically summarized. Tables for the remaining experiments are available from the first author upon request.

Table 1
Usefulness Indexes (UI) of Each Cue for Each Subject in Experiment 1

Subject no.	Cue							
	Severity (driver)	Severity (victim)	Road conditions	Traffic	Record	Speeding	Brakes	R^2
1				.09		.35	.44	.70
2							.23	.35
3			.06	.05		.21	.51	.66
4			.06	.08		.23	.30	.60
5						.32	.42	.75
6		.07				.46	.08	.65
7					.09	.31	.31	.84
8						.40	.09	.56
9					.08	.49		.73
10				.04		.34	.33	.75
Frequency	0	1	2	4	2	9	9	
M_{UI}	.00	.01	.01	.03	.02	.31	.27	

Note. No entry indicates that the usefulness index was not significant ($p < .05$).

similar, to the extent that 8 of 10 people utilized both speeding and brakes information to a significant degree, whereas severity of consequences to either the driver or victims was of minimal importance; and (b) there were large individual differences regarding both the magnitude of importance of particular cues and which cues were relied upon by any given subject.

Experiment 2

Method

Since speeding and brakes were last in order of appearance in each story and consequences to driver and victims occurred first and second, it seemed possible that the striking reliance on speeding and brakes may have been an artifact of serial position (i.e., a "recency effect"). Thus, Experiment 1 was replicated using the same profiles, except that consequences to the driver and victims appeared last and next to last in each story, while speeding and brakes were second and third in order of appearance. Ten introductory psychology students were presented the new booklets, following the same procedure as in Experiment 1.

Results

The R^2 values were similar to those obtained in Experiment 1 ($M = .72$). The overwhelming importance of speeding and brakes information observed in Experiment 1 was again apparent. All subjects utilized speeding information to a significant extent ($M_{UI} = .41$). Seven subjects employed information about brakes ($M_{UI} = .10$). Thus, the possibility of an artifact due to serial position does not seem likely. The frequency of usage of consequences was somewhat greater than in Experiment 1 (four subjects employed consequences to victims, $M_{UI} = .06$; two subjects employed consequences to the driver, $M_{UI} = .02$), but the importance of these cues was still much less than that of speeding and brakes.

Experiment 3

Method

Since the first two experiments indicated that speeding and brakes were the major determinants of responsibility attributions, and given the slight increase in the frequency of utilization of consequence information in Experiment 2, it was of interest to determine whether severity of consequences would become a major determinant of attributions if speeding and brakes information were not available. Hence, new profile booklets were constructed in which the speeding and brakes cues were deleted from the same 40 profiles employed in Experiment 2 (i.e., severity information still appeared last in the stories), and 10 additional subjects were asked to make responsibility judgments about these stories.

Results

The R^2 values were again similar to those of the first two experiments ($M = .60$). Examination of the significant usefulness indexes revealed that the most important de-

terminant of attributions was driving record (eight subjects employed this cue to a significant extent, $M_{UI} = .41$). Severity of consequences to the victims was the second most important cue (five subjects, $M_{UI} = .05$). Severity of consequences to the driver was used by only one subject.

Experiment 4

In order to determine how important severity of consequences would be in the absence of information regarding speeding, brakes, and driving record, 10 more subjects were presented the same 40 profiles, in which driving record was now deleted as well. Four of these subjects circled *1*s (i.e., "not at all responsible") on all profiles, and two subjects circled all *4*s (i.e., the midpoint on the scale). These subjects also indicated on an open-ended postexperimental questionnaire that there was not enough information available on which to base their judgments. Of the remaining four subjects, the R^2 for one was nonsignificant, and there was no identifiable pattern of cue utilization for the other three subjects.

General Discussion

Role of Severity in Responsibility Attribution

The results of Experiments 1 and 2 clearly indicate that the two most important determinants of attributions were the cues regarding whether the driver was speeding prior to the accident (19 of 20 subjects employed this cue to a significant degree) and whether the brakes of the car were in good or poor condition (16 of 20 subjects). The importance of these two cues was far greater than that of severity of consequences to the victims (5 subjects) or severity of consequences to the driver (2 subjects). This underscores the conclusion drawn by Shaver (Note 1) that "overall severity-dependent responsibility attribution (is) the exception rather than the rule" (p. 9).

Role of Controllable Versus Uncontrollable Factors

It is interesting that the two most frequently utilized cues (speeding and brakes) were factors directly under the driver's control. That is, an individual can be expected to take responsibility for driving at a safe speed and for maintaining the brakes of his or her car. On the other hand, factors such as road and traffic conditions or the severity of the consequences of an accident are not under a person's direct control. It is also interesting that few subjects employed more than two or three of the seven possible factors. It appears that most subjects simply chose to attend to that information which they deemed to be most relevant to the problem of assigning responsibility for the accident, while ignoring the other cues. When the two most relevant cues were eliminated, subjects relied almost unanimously on driving record (8 of 10 subjects). In accord with the hypothesis that observers rely on information regarding factors over which the driver has control, the overall importance of driving record in Experiment 3 is not surprising, since this was the only factor present over which the driver could be thought to have *any* control (in the sense of suggesting, for example, carelessness). The tendency for subects to rely on factors that are under the driver's control is consistent with Lerner's (1970) proposition that we have a need to believe that people deserve bad things that befall them, as well as with Wortman's (1976) argument that we make attributions in order to enhance our feelings of control over the environment.

Methodological Issues

In interpreting the results of the present experiments, several methodological caveats should be borne in mind. First, the relative "importance" of the various factors manipulated may be due not to the factors themselves but to the *range* of levels employed (i.e., the particular wording used to specify the different levels). Different wording might have produced a very different ordering of the variables in terms of importance (see Ebbesen & Konečni, 1975). Although the levels employed here seem representative of high and low values, it might be useful to consider several different levels in future research. Additionally, a different pattern of cue utilization may be obtained with other dependent

measures (e.g., sanction or guilt; cf. Fishbein & Ajzen, 1973; Vidmar & Crinklaw, 1974). It is also possible that judges may integrate cues in an interactive fashion. However, the R^2 values obtained here indicate that the majority of the variance in responsibility judgments was accounted for by a linear combination of the cues, and there is little evidence in the literature on human judgment and decision making of nonlinear cue utilization (Goldberg, 1968, 1970). Nevertheless, interaction terms can be included in the multiple regression equation (Kerlinger & Pedhazur, 1973).

Finally, there is some evidence in our data for a serial position effect on severity utilization. That is, consequences appeared to be somewhat more readily utilized when they came last in serial order than when they came early. Recent work dealing with children's moral judgments (Austin, Ruble, & Trabasso, 1977; Feldman, Klosson, Parsons, Rholes, & Ruble, 1976) has demonstrated a recency effect on consequence utilization due to presentation order. While this result suggests a possible resolution for the inconsistencies in past responsibility attribution research (i.e., perhaps consequences information was presented last in those experiments reporting the severity effect, whereas it may have been presented first in those studies failing to find the effect), a review of the procedures of past research indicates that this is not the case (cf. Chaiken & Darley, 1973; McKillip & Posavac, 1975; Medway & Lowe, 1975; Shaver, 1970; Stokols & Schopler, 1973; Walster, 1966). Also, no serial position effect was observed for speeding and brakes utilization, and that observed for severity was minimal. However, the small effect does suggest that perhaps future research should counterbalance the serial position of a particular cue.

Applicability of the Lens-Model Approach

More important than the specific implications of these findings for responsibility attribution are the general implications of the lens-model approach. The usefulness of this methodology seems apparent. It permits the parsimonious assessment of proportions of variance accounted for by a large number of attributional determinants for *individual subjects*, thus providing a way to systematically investigate individual differences. For example, individuals may differ in their tendency to rely on severity of consequences in assigning responsibility for an accident. In fact, our experiments suggest that this may indeed be the case. Even among subjects who relied on severity, some used this cue to a great extent, whereas others relied on it very little. Such individual differences (as well as differences with respect to how *other cues* are employed) could be subjected to experimental investigation, perhaps leading to the identification of attributional "styles" and the construction of theories regarding the characteristic manner in which people integrate information in making attributions.

Reference Note

1. Shaver, K. G. *Intentional ambiguity in the attribution of responsibility*. Unpublished manuscript, 1973. (Available from the Department of Psychology, College of William and Mary, Williamsburg, Virginia 23185.)

References

Austin, V. D., Ruble, D. N., & Trabasso, T. Recall and order effects in children's moral judgments. *Child Development*, 1977, *48*, 470–474.

Brunswik, E. *The conceptual framework of psychology*. Chicago: University of Chicago Press, 1952.

Chaiken, A. L., & Darley, J. Victim or perpetrator? Defensive attribution of responsibility and the need for order and justice. *Journal of Personality and Social Psychology*, 1973, *25*, 268–275.

Darlington, R. B. Multiple regression in psychological research and practice. *Psychological Bulletin*, 1968, *69*, 161–182.

Dudycha, L. W., & Naylor, J. C. Characteristics of the human inference process in complex choice behavior. *Organizational Behavior and Human Performance*, 1966, *1*, 110–128.

Ebbesen, E. B., & Konečni, V. J. Decision making and information integration in the courts: The setting of bail. *Journal of Personality and Social Psychology*, 1975, *32*, 805–821.

Feldman, N. S., Klosson, E. C., Parsons, J. E., Rholes, W. S., & Ruble, D. N. Order of information presentation and children's moral judgments. *Child Development*, 1976, *47*, 556–559.

Fishbein, M., & Ajzen, I. Attribution of responsibil-

ity: A theoretical note. *Journal of Experimental Social Psychology*, 1973, *9*, 148–153.

Fischhoff, B. Attribution theory and judgment under uncertainty. In J. H. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research*. Hillsdale, N.J.: Erlbaum, 1976.

Goldberg, L. R. Simple models or simple processes? Some research on clinical judgments. *American Psychologist*, 1968, *23*, 483–496.

Goldberg, L. R. Man versus model of man: A rationale, plus some evidence, for a method of improving on clinical inferences. *Psychological Bulletin*, 1970, *73*, 422–432.

Hammond, K. R., Stewart, T. R., Brehmer, B., & Steinmann, D. O. Social judgment theory. In M. F. Kaplan & S. Schwartz (Eds.), *Human judgment and decision processes*. New York: Academic Press, 1975.

Keeley, S. M., & Doherty, M. E. Probabilistic and multiple regression modeling of the biologist's decision processes and its implications for medical diagnosis. *Bulletin of Mathematical Biophysics*, 1971, *33*, 439–449.

Kerlinger, F. N., & Pedhazur, E. J. *Multiple regression in behavioral research*. New York: Holt, Rinehart & Winston, 1973.

Lerner, M. J. The desire for justice and reactions to victims. In J. Macauley & L. Berkowitz (Eds.), *Altruism and helping behavior*. New York: Academic Press, 1970.

McKillip, J., & Posavac, E. J. Judgments of responsibility for an accident. *Journal of Personality*, 1975, *43*, 248–265.

Medway, F. J., & Lowe, C. A. Effects of outcome valence and severity on responsibility attribution. *Psychological Reports*, 1975, *36*, 239–246.

Shaver, K. G. Defensive attribution: Effects of severity and relevance on the responsibility assigned for an accident. *Journal of Personality and Social Psychology*, 1970, *14*, 101–113.

Slovic, P., & Lichenstein, S. Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance*, 1971, *7*, 649–744.

Stokols, D., & Schopler, J. Reactions to victims under conditions of situational detachment: The effects of responsibility, severity, and expected future interaction. *Journal of Personality and Social Psychology*, 1973, *25*, 199–209.

Vidmar, N., & Crinklaw, L. D. Attributing responsibility for an accident: A methodological and conceptual critique. *Canadian Journal of Behavioral Sciences*, 1974, *6*, 113–129.

Walster, E. Assignment of responsibility for an accident. *Journal of Personality and Social Psychology*, 1966, *3*, 73–79.

Walster, E. "Second-guessing" important events. *Human Relations*, 1967, *20*, 239–250.

Wortman, C. B. Causal attributions and personal control. In J. H. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research*. Hillsdale, N.J.: Erlbaum, 1976.

Received February 7, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 1, 116-123

Marital Roles, Sex Differences, and Interpersonal Attraction

Joseph E. Grush
Northern Illinois University

Janet G. Yehl
University of South Carolina at Spartanburg

This investigation tested the hypothesis that role and dispositional factors can alter the usual link between similarity and attraction. In two replications, females and males with traditional or nontraditional attitudes toward sex roles and marriage rated similar or dissimilar opposite-sex strangers on three attraction measures. Results showed that sex and traditionality interacted with similarity in determining ratings of the strangers' general likability and personal role attraction (e.g., desirability as a dating partner), but not their functional role attraction (e.g., desirability as a debater on sex roles). The discussion suggests that previous attempts to find interactions have often failed because of a lack of correspondence between dispositional factors and similarity manipulations. The discussion also suggests that reinforcement and informational explanations could account for the overall findings, but that communication factors of extremity and discrepancy could not.

Investigators using Byrne's (1971) attraction paradigm have generally found a strong positive relationship between attitude similarity and interpersonal attraction. This relationship has been found for different modes of stimulus presentation (Byrne & Clore, 1966), different populations (Byrne, Griffitt, Hudgins, & Reeves, 1969), and different issues of varying levels of importance (Clore & Baldridge, 1968). The generalizability of the similarity–attraction relationship has been further demonstrated by investigations that have manipulated similarity on the basis of abilities (Zander & Havelin, 1960), economic status (Byrne, Clore, & Worchel, 1966), emotional states (Zimbardo & Formica, 1963), and personality traits (Byrne, Griffitt, & Stefaniak, 1967).

According to Clore and Byrne's (1974) reinforcement–affect model, attraction toward another person is determined by the proportion of reinforcements and punishments associated with that person. Although similarity between judge and target has frequently been found to be reinforcing, the focus of the model is on interpersonal reinforcement rather than similarity per se. Thus, a recent corollary of the model postulates that role and dispositional factors can alter the usual similarity–attraction relationship by altering the reinforcement contingencies associated with another person (Clore & Byrne, 1977).

In the corollary, roles refer to societal positions in which the occupant is expected to behave in certain prescribed ways. Thus, certain roles may attenuate the reinforcing properties of similarity because the expectations for those roles do not include similarity. Dispositions refer to consistent differences among individuals in their orientation toward responding to the same stimulus. Thus, individuals with different socialization (reinforcement) histories may differ in their attraction responses to stimulus persons who are similar or dissimilar to themselves.

The present investigation was designed to test whether role and dispositional factors would moderate the impact of similarity on attraction. Female and male students, with traditional or nontraditional attitudes toward marital roles, judged opposite-sex strangers who had similar or dissimilar attitudes. Tra-

The authors would like to thank David Sujak and Linda Fort for their help in completing the preliminary study and Kathleen Keeley for her timely suggestions and competent assistance in conducting the laboratory experiment.

Requests for reprints should be sent to Joseph E. Grush, Department of Psychology, Northern Illinois University, DeKalb, Illinois 60115.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3701-0116$00.75

ditionality was determined by scores on a marital role decisions questionnaire that is described later (see Preliminary Study). Traditional respondents, for example, believe that wives should be concerned with child rearing and domestic matters, whereas husbands should focus on career and financial matters. Nontraditional respondents believe that these roles should be jointly shared by both spouses. Judgments of the opposite-sex strangers were made on measures of *personal role attraction* (e.g., desirability as a marriage partner), *functional role attraction* (e.g., desirability as a debater on marital roles), and *general attraction* (i.e., likability).

The personal and functional role attraction measures were included to assess different types of (anticipated) interactions between judge and target. Personal role items assessed attraction in role relationships in which judge and target would directly interact at a personal or intimate level. Functional role items assessed attraction in role relationships in which the judge would primarily observe and evaluate the target person's performance of certain public roles. Finally, the general attraction item was included to determine if global attraction tends to mirror personal or functional role attraction.

The major hypothesis was that sex and traditionality of the judges would interact with similarity of the target persons in determining personal and general attraction but not functional attraction. This hypothesis was made up of three distinct predictions relating to the three attraction measures.

Prediction 1 was that nontraditional females and traditional males would display a distinct preference for similar partners on personal role attraction, whereas traditional females and nontraditional males would not. This prediction was based on the assumption that nontraditional women and traditional men have irreconcilable differences that would lead to constant conflicts and make personal relationships between the two quite punishing. Prediction 1 also assumed that traditional women and nontraditional men could find similar or dissimilar partners equally reinforcing in personal roles but for different reasons. With similar partners, reinforcement would come from the mutuality of shared beliefs. With dissimilar partners, traditional women would find it reinforcing to know, for example, that they could pursue careers and receive help at home if they so desired. Nontraditional men would find it reinforcing to know, for example, that they could fully pursue their own careers and avoid the drudgery of household chores if they chose to do so.

Prediction 2 was that general attraction would parallel personal role attraction. This prediction was based on research that shows that general likability tends to mirror role attraction when the latter involves direct interactions between judge and target (Grush, Clore, & Costin, 1975).

Prediction 3 was that similarity manipulations, either alone or in conjunction with characteristics of the judges, would not affect ratings of functional role attraction. This prediction was based on the assumption that debates and discussions of controversial issues are by nature interesting and informative (i.e., reinforcing) to the extent that both sides of the issues are presented. In other words, ratings of another person's suitability to perform these roles should be based more on whether they have a point of view to express than whether they agree with us.

Preliminary Study

Scale Construction

The preliminary study was conducted to develop the Marital Role Decisions Questionnaire (MRDQ) that would be used in the main experiment to select individuals who had traditional or nontraditional orientations toward sex roles and marriage. Possible items for this questionnaire were initially elicited from 20 married couples who responded to an open-ended questionnaire that asked them to describe (a) "important decisions that you and your spouse (or couples you know) had to make where the two of you had opposing viewpoints" and (b) "the choices that were available to you and your spouse (or couples you know)."

This elicitation procedure produced nearly 40 possible items. These 40 items were then reduced to 14 by (a) eliminating trivial or idiosyncratic decisions, (b) discarding de-

Table 1
Comparison of Mean Scores on the Marital Role Decisions Questionnaire for Five Known Groups

Females		Males	
Group	Score	Group	Score
Members of NOW[a]	21.08$_a$		
College students	40.75$_b$	College students	43.92$_b$
Wives of policemen	51.00$_c$	Policemen	57.33$_c$

Note. Scores on the Marital Role Decisions Questionnaire could range from 12 to 84, with higher scores indicating a more traditional orientation toward sex roles and marriage. Means with different subscripts differ significantly from each other beyond the .05 level (two-tailed).
[a] National Organization of Women.

cisions thought not to be amenable to traditional and nontraditional responses, and (c) constructing composite descriptions of decisions that were similar in content. The 14 resulting decisions covered the areas of *domestic* (whether household chores should be shared), *financial* (whether home should be purchased or rented), *social* (whether functions should be attended separately or together), *familial* (whether children should be placed in a nursery), and *career* (who works to put whom through professional school) considerations. Except for domestic considerations, each content area was represented by 3 items. These decisions were then rewritten as dilemmas, each with a 7-point response format: slight, moderate, or strong agreement with a traditional or a nontraditional alternative for resolving the dilemma and a "cannot decide" response.

Reliability and Validity

To test the reliability and validity of the MRDQ, the 14-item scale was administered to a small number of individuals from five "known groups." The number of individuals and the five groups from which they came were 13 female members of a local chapter of the National Organization of Women (NOW); 12 female college students; 12 spouses of policemen; 13 male college students; and 12 policemen. It is important to note that these samples were independent. That is, the female and male college students were from different schools, and the spouses of policemen and the policemen were not married to one another.

Coefficient alpha was computed to determine the reliability of the MRDQ. With each item treated as a parallel measure, reliability for the 14-item MRDQ was .78. However, inspection of the data revealed that 2 items (1 social and 1 financial dilemma) functioned independently of the other 12 items. The deletion of those 2 items produced a 12-item MRDQ scale with a reliability of .79.

The validity of the 12-item MRDQ scale was assessed by determining whether the responses of the known groups significantly differed in the expected direction. The results of these comparisons are presented in Table 1. As can be seen there, the mean MRDQ scores for the three groups of women significantly differed from each other: The NOW women were more nontraditional than college women, who in turn were more nontraditional than the spouses of policemen. As Table 1 also indicates, the two groups of male respondents significantly differed from each other: Male college students were more nontraditional than policemen. Comparisons between the sexes for comparable groups revealed no significant differences, although female college students and the spouses of policemen were slightly more nontraditional than their male counterparts. In summary, then, the preliminary study showed that the 12-item MRDQ had sufficiently good psychometric properties for use as a research instrument.[1]

[1] Copies of the MRDQ, as well as instructions and scoring procedures for its use, are available from the first author upon request.

Main Experiment

Method

Design

The experiment had a 2 (female or male) × 2 (traditional or nontraditional) × 2 (similar or dissimilar) between-subjects factorial design. The first two factors involved characteristics of the experimental participants, who served as judges in an impression-formation task. The factor of traditionality was determined by participants' responses on the MRDQ. Individuals scoring above the true midpoint of the scale were classified as "traditionals," whereas individuals scoring below the true midpoint were classified as "nontraditionals." The third factor was whether the target person's score on the MRDQ was similar or dissimilar to the judge's score.

Participants

One hundred twenty-eight college students (64 females and 64 males) served as experimental participants, with 16 individuals participating in each cell of the 2 × 2 × 2 design.[2] All students were enrolled in introductory psychology courses at Northern Illinois University and received extra course credit for their voluntary participation in the experiment. Students were recruited during the first week of classes, at which time volunteers completed the MRDQ. The experiment was conducted several weeks later, with students participating in groups of 5 to 10.

Procedure

Under the guise of an impression-formation task, each student examined the MRDQ responses of an opposite-sex stranger. Consistent with the cover story, no other information about the strangers was given except that they were also college students and of the same age group as the participants. The strangers' responses on the MRDQ were manipulated to be either similar or dissimilar to the students' response patterns. After examining the strangers' MRDQ responses, the students recorded their impressions of the strangers on items intended to measure personal role, functional role, and general attraction. Finally, all students were thanked for their participation and debriefed about the general purpose and methodology of the investigation.

Similarity Manipulations

The target person's response pattern on the MRDQ was manipulated to be either 10 (similarity) or 30 (dissimilarity) points discrepant and opposite in direction from the student's response pattern. The rules for constructing the target's MRDQ responses were (a) always change the target's responses in the direction that is opposite the student's overall score on the MRDQ; (b) make three sequential changes of 10 points each; (c) for each sequential change, randomly alter as many different items as possible; and (d) avoid using the "cannot decide" response. Once these rules were employed, each student was randomly assigned to the similar or dissimilar condition. Similarity and dissimilarity were based on 10 and 30 points, respectively, because the standard deviation of scores on the MRDQ was 10.32 for this population.

Dependent Measures

The dependent measures included items intended to measure personal role, functional role, and general attraction. There were three items intended to measure personal role attraction, or the extent to which the judge would enjoy interacting with the target person: How suitable is the target person as a work, dating, and marriage partner? There were three items intended to measure functional role attraction, or the extent to which the judge would enjoy observing the target person perform public roles: How suitable is the target person as a debater, panelist, and discussant on sex roles and marriage? A single item measured how much the judge would generally like the target person. Four filler items assessed the degree to which the target person was intelligent, moral, socially adjusted, and knowledgeable about current affairs. All 11 items were assessed on 7-point response scales, with anchors of "very much" and "very little."

Results

Reliability

Item analysis based upon pooled within-cell correlations was used to assess whether items designed to measure personal and functional role attraction could be legitimately combined. Ratings of the target persons' suitability as work, dating, and marriage partners were highly intercorrelated with each other (mean $r = .65$). Scores on these items were thus averaged to create a measure of *personal role attraction*. Ratings of the target persons' suitability as debaters, panelists, and

---

[2] In actuality, two replications of the experiment were conducted during two consecutive semesters. Since an analysis for a replicated experiment produced no reliable differences between the two semesters, the data for both semesters were combined to increase the power of the subsequent analyses.

Table 2
Mean Ratings of Personal Role, General, and Functional Role Attraction for the Eight Experimental Groups

Experimental group	Attraction measure				
	Personal role		General		Functional role rating
	Rating	p	Rating	p	
Female–nontraditional					
Similar	5.50	<.001	5.19	<.01	5.19
Dissimilar	2.87		4.13		5.23
Female–traditional					
Similar	5.25	*ns*	5.31	*ns*	5.51
Dissimilar	5.46		5.13		6.19
Male–nontraditional					
Similar	5.08	*ns*	5.13	*ns*	5.04
Dissimilar	4.50		4.81		4.71
Male–traditional					
Similar	4.45	<.02	5.06	<.01	4.90
Dissimilar	3.54		3.88		5.27

Note. Ratings on each attraction measure could range from 1 to 7, with higher ratings indicating greater attraction. Given that there were four directional predictions concerning the simple tests, a conservative alpha level would be .0125 rather than .05.

discussants of sex roles and marriage were also highly intercorrelated (mean $r = .60$). Scores on these items were thus averaged to create a measure of *functional role attraction.* Since personal role items were only minimally correlated with functional role items (mean $r = .17$), discriminant evidence was also sufficient to treat these two sets of items as separate measures of attraction. Finally, the likability item was treated as a single measure of *general attraction,* because its mean correlations with personal and functional role attraction were .53 and .28, respectively.

Principal Hypothesis

The three attraction measures were treated as separate dependent variables in a multivariate analysis of variance. The results of this analysis showed that the multivariate F for the Sex × Traditionality × Similarity interaction was significant, $F(3, 118) = 4.42$, $p < .01$. The corresponding univariate Fs for personal role and general attraction were also significant: $F(1, 120) = 12.34$, $p < .001$, and $F(1, 120) = 5.73$, $p < .05$, respectively. The univariate F for functional role attraction was not significant, $F(1, 120) = .01$, $p > .90$. These findings supported the principal hypothesis that sex and traditionality of the judges would interact with similarity of the targets in determining personal role and general attraction but not functional role attraction.[3]

Specific Predictions

Simple tests compared the personal role and general attraction ratings of similar and dissimilar target persons made by the various judges. The results of these comparisons are reported in Table 2. As can be seen there, nontraditional females displayed a distinct

[3] As an alternative test of the principal hypothesis, a profile multivariate analysis of variance was conducted. This analysis included two linear combinations of the attraction measures: (a) the average of personal and general attraction versus functional attraction and (b) personal versus general attraction. The results of this analysis were statistically significant, $F(2, 119) = 4.94$, $p < .01$. A step-down analysis indicated that the first contrast accounted for a significant amount of variance in the overall analysis, $F(1, 120) = 9.44$, $p < .005$, whereas the addition of the second contrast did not, $F(1, 120) = .49$. Thus, the results of this analysis also supported the principal hypothesis.

preference for similar as opposed to dissimilar target persons on personal role and general attraction ($p < .001$ and $p < .01$, respectively). Traditional men also preferred similar partners on personal role and general attraction ($p < .02$ and $p < .01$, respectively). By contrast, traditional women and nontraditional men displayed no such preference for similarity on either attraction measure ($ps > .15$). These findings supported Predictions 1 and 2 of the principal hypothesis.

Although Table 2 contains the means for functional role attraction, simple tests were not conducted because they were unwarranted. As already noted, for example, the triple interaction for functional role attraction was not significant. Moreover, the main effect for similarity was not significant, $F(1, 120) = .97$, $p > .30$. The Sex × Similarity and the Traditionality × Similarity interactions were also nonsignificant: $F(1, 120) = .77$, $p > .35$, and $F(1, 120) = 3.07$, $p > .08$, respectively. These findings supported Prediction 3 of the principal hypothesis.

Discussion

This study tested the hypothesis that role and dispositional factors can moderate the effects of similarity on attraction. Results showed that nontraditional women and traditional men displayed more liking (general attraction) and stronger preferences for similar versus dissimilar others as work, dating, and marriage partners (personal role attraction). By contrast, traditional women and nontraditional men were equally attracted to similar and dissimilar others on these same measures. Results also showed that all judges considered similar and dissimilar others to be equally suited to perform the roles of debaters, discussants, and panelists on topics of sex roles and marriage (functional role attraction). In summary, this study demonstrated that characteristics of judges and dimensions of judgment can affect interpersonal attraction as much as similarity of target persons.

The significance of the present study lies in the fact that it is apparently the first to produce clearly replicated results (see Footnote 2) of the moderating effects of dispositional factors on attraction responses.[4] Thus, the question naturally arises as to why the present study succeeded when past efforts have generally failed in this regard (see Byrne, 1971; Fishbein & Ajzen, 1972).

One possibility is that previous studies typically used Byrne's Interpersonal Judgment Scale (IJS), whereas the present study employed different attraction measures. Since the likability and work partner items of the present study are psychologically similar to the two items comprising the IJS, an analysis of these combined items was conducted to test this possibility. Not surprisingly, the results of this analysis were exactly parallel to those for personal role and general attraction: The Sex × Traditionality × Similarity interaction was significant, $F(1, 120) = 13.80$, $p < .001$; nontraditional women displayed a distinct preference for similarity, $F(1, 30) = 33.05$, $p < .001$, as did traditional men, $F(1, 30) = 7.19$, $p < .05$; and traditional women and nontraditional men displayed no such preference ($Fs < 1.50$). Thus, the success of the present study cannot be attributed to its use of a novel measure of attraction.

A second possibility is that previous studies failed to obtain moderating effects because the relationship between dispositional factors and similarity manipulations was more tenuous than in the present study. For example, in the studies of authoritarianism, participants' scores on the *F* scale and their responses on the attitude items were either unrelated (Byrne, 1965) or only minimally correlated (Sheffield & Byrne, 1967). Thus, it is not surprising that authoritarianism did not interact with the similarity of others' responses on issues that were generally unrelated to the authoritarian ideology. In other

---

[4] Although moderating effects on attraction have been obtained with need for affiliation, different investigations have produced different results (cf. Byrne, 1961, 1962). Somewhat similarly, the dispositional variable of social avoidance and distress has significantly interacted with similarity in affecting attraction in one study (Smith, 1972) but not another (Gouaux, Lamberth, & Friedrich, 1972).

studies, the tenuous relationship may have occurred in terms of unrealistic expectations about the personal benefits of attitude similarity. For example, it seems unreasonable to expect that individuals with high need for approval (Ettinger, Nowicki, & Nelson, 1970) can be so positively affected by attitude similarity that they will display more attraction toward agreeing strangers than that displayed by individuals with low need for approval.

By contrast, the present study ensured a close relationship between dispositional variables and similarity manipulations by using the same items to assess the dispositional variable and to manipulate similarity. Furthermore, information contained in these items was directly relevant to the dependent measures of attraction. For example, faced with the task of rating a target person's suitability as a potential marriage partner, participants could directly assess whether their own orientations toward marriage were compatible with the target person's orientation. Thus, the present study was probably successful in finding interactions because there was a closer correspondence between dispositional factors and similarity manipulations than that which existed in previous studies.

Alternative Interpretations

While the present findings were predicted from a reinforcement perspective, they are compatible with informational approaches to attraction. For example, Ajzen (1974) views attraction as being a summed function of the subjective probabilities that a target person has certain attributes and the affective values of those attributes. Thus, if nontraditional women and traditional men consider similar others to have more positive attributes for personal roles than dissimilar others, they would evaluate similar others more positively than dissimilar others. If these same judges also consider both similar and dissimilar others to have equally positive attributes for functional roles, they would evaluate both target persons equally well.

The present findings are also compatible with Kaplan and Anderson's (1973) information integration approach to attraction. According to this approach, variations in the judges' initial impressions or in the scale values and weights assigned to pieces of information about the target persons could account for the results obtained here. For example, the personal role findings could be due to nontraditional women and traditional men having more negative initial impressions of dissimilar others than traditional women and nontraditional men have of them. Additionally, differences between personal and functional role findings could be due to differences in the scale values or weights assigned to the same target when that person is judged on different dimensions.

A final interpretation warrants comment. The similarity manipulations of this study bear a strong resemblance to discrepancy manipulations in persuasion research. Thus, if nontraditional women and traditional men held extreme views on marriage, their dislike of dissimilar others would merely replicate the well-known derogation effect of discrepant communications and communicators (Hovland, Harvey, & Sherif, 1957; Johnson, 1966). To determine whether any group of judges held extreme views on marriage, mean MRDQ scores were computed for each group. The mean scores of both groups of nontraditional women were nearly two standard deviations away from the midpoint of the scale. However, the mean scores of all other groups, including the two groups of traditional men, were within a standard deviation of the midpoint.

Thus, extremity and discrepancy may account for the nontraditional women's dislike of dissimilar others on personal role and general attraction. However, these factors cannot explain the traditional men's derogation of dissimilar others on these same measures. Nor can they explain the nontraditional women's liking of dissimilar others on functional role attraction. This latter point is particularly important because functional attraction assessed the target persons' suitability to perform roles (e.g., debater) that are similar to the communicator roles for which derogation effects have been obtained. In summary, extremity and discrepancy may account for part of the present findings but not their overall pattern.

Conclusion

McGuire (1968) has argued that the principles governing human affairs are more likely to be contingent-interactive effects rather than condition-free main effects. In this regard, the present study represents a modest beginning in the area of interpersonal attraction. That is, the present study demonstrated that interpersonal attraction can be a multivariate function of the type of person making the judgment, the type of person being judged, and the type of dimension upon which the judgment is made. As investigators continue to pursue the interactive nature of interpersonal attraction more fully, it is hoped that the principles underlying evaluative judgments of others will begin to emerge more explicitly.

References

Ajzen, I. Effects of information on interpersonal attraction: Similarity versus affective value. *Journal of Personality and Social Psychology,* 1974, *29,* 374–380.

Byrne, D. Interpersonal attraction as a function of affiliation need and attitude similarity. *Human Relations,* 1961, *14,* 283–289.

Byrne, D. Responses to attitude similarity–dissimilarity as a function of affiliation need. *Journal of Personality,* 1962, *30,* 164–177.

Byrne, D. Authoritarianism and response to attitude similarity–dissimilarity. *Journal of Social Psychology,* 1965, *66,* 251–256.

Byrne, D. *The attraction paradigm.* New York: Academic Press, 1971.

Byrne, D., & Clore, G. L. Predicting interpersonal attraction toward strangers presented in three different stimulus modes. *Psychonomic Science,* 1966, *4,* 239–240.

Byrne, D., Clore, G. L., & Worchel, P. Effect of economic similarity–dissimilarity on interpersonal attraction. *Journal of Personality and Social Psychology,* 1966, *4,* 220–224.

Byrne, D., Griffitt, W., Hudgins, W., & Reeves, K. Attitude similarity–dissimilarity and attraction: Generality beyond the college sophomore. *Journal of Personality and Social Psychology,* 1969, *19,* 155–161.

Byrne, D., Griffitt, W., & Stefaniak, D. Attraction and similarity of personality characteristics. *Journal of Personality and Social Psychology,* 1967, *5,* 82–90.

Clore, G. L., & Baldridge, B. Interpersonal attraction: The role of agreement and topic interest. *Journal of Personality and Social Psychology,* 1968, *9,* 340–346.

Clore, G. L., & Byrne, D. A reinforcement–affect model of attraction. In T. L. Huston (Ed.), *Perspectives in interpersonal attraction.* New York: Academic Press, 1974.

Clore, G. L., & Byrne, D. The process of personality interaction. In R. B. Cattell & R. M. Dreger (Eds.), *Handbook of modern personality theory.* Washington, D.C.: Hemisphere, 1977.

Ettinger, R. F., Nowicki, S., & Nelson, D. A. Interpersonal attraction and the approval motive. *Journal of Experimental Research in Personality,* 1970, *4,* 95–99.

Fishbein, M., & Ajzen, I. Attitudes and opinions. *Annual Review of Psychology,* 1972, *23,* 487–544.

Gouaux, C., Lamberth, J., & Friedrich, G. Affect and interpersonal attraction: A comparison of trait and state measures. *Journal of Personality and Social Psychology,* 1972, *24,* 53–58.

Grush, J. E., Clore, G. L., & Costin, F. Dissimilarity and attraction: When difference makes a difference. *Journal of Personality and Social Psychology,* 1975, *32,* 783–789.

Hovland, C. I., Harvey, O. J., & Sherif, M. Assimilation and contrast effects in reactions to communication and attitude change. *Journal of Abnormal and Social Psychology,* 1957, *55,* 244–252.

Johnson, H. H. Some effects of discrepancy level on responses to negative information about one's self. *Sociometry,* 1966, *29,* 52–66.

Kaplan, M. F., & Anderson, N. H. Information integration theory and reinforcement theory as approaches to interpersonal attraction. *Journal of Personality and Social Psychology,* 1973, *28,* 301–312.

McGuire, W. J. Personality and susceptibility to social influence. In E. F. Borgatta & W. W. Lambert (Eds.), *Handbook of personality theory and research.* Chicago: Rand McNally, 1968.

Sheffield, J., & Byrne, D. Attitude similarity–dissimilarity, authoritarianism, and interpersonal attraction. *Journal of Social Psychology,* 1967, *71,* 117–123.

Smith, R. E. Social anxiety as a moderator variable in the attitude similarity–attraction relationship. *Journal of Experimental Research in Personality,* 1972, *6,* 22–28.

Zander, A., & Havelin, A. Social comparison and interpersonal attraction. *Human Relations,* 1960, *13,* 21–32.

Zimbardo, P., & Formica, R. Emotional comparison and self-esteem as determinants of affiliation. *Journal of Personality,* 1963, *31,* 141–162.

Received February 20, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 1, 124-130

Simple Minitheories of Love

P. M. Bentler and G. J. Huba
University of California, Los Angeles

Two causal models of love were developed as alternatives to the Tesser and Paulhus theory. These models were tested for adequacy of fit using maximum-likelihood methods. While the Tesser–Paulhus models can be rejected empirically, the alternate models provide acceptable statistical representations of four variables measuring love at two times separated by 2 weeks. One formulation is based on the idea of unidimensionality of interpersonal attraction. The second formulation represents a refinement of the Tesser–Paulhus view of love. The formulation based on the idea that interpersonal attraction is primarily a unidimensional construct provides the more parsimonious and interpretable theory.

Causal modeling methods have been used in this journal to evaluate a theory of love (Tesser & Paulhus, 1976). As a method of theory testing with correlational data, causal modeling represents a set of statistical techniques for investigating sample data in the light of hypothesized population models of the influence of certain variables on others. In general, a model represents a simultaneous series of statements about the regression of particular variables on various other ("causal," explanatory) variables. The most popular form of causal modeling, path analysis, is thus a sophisticated form of simultaneous multiple correlation/regression analysis, while recent developments in causal modeling embed path analysis into a factor-analytic type of framework. In such a framework, the regressions are at the level of the unmeasured, latent variables or factors, and the factors themselves are related by factor-analytic assumptions to the observed, manifest variables in the system. The goals of causal modeling include both the testing of a proposed model against data and the development of models that adequately account for data. The interrelations among causal modeling, data analysis, theory testing, and construct validation have been discussed by Bentler (1978).

Although causal models must be evaluated by a variety of criteria including meaningfulness, they can be tested statistically via the chi-square statistic: The hypothesis that the sample variances and covariances are drawn from a population having a hypothesized causal structure is tested against the alternative hypothesis that the variables are simply correlated. If the chi-square is large relative to degrees of freedom, the proposed causal structure must be rejected, because the observed sample data would be extremely unlikely to be obtained if the hypothesized model were true in the population. The chi-square statistic, in turn, is developed on the basis of various mathematical and statistical assumptions. A very general mathematical system of equations for causal models (Bentler, 1976) reduces to a set of equations described by Wiley (1973) and Jöreskog (1977), for which a computer program, LISREL III, exists, to obtain maximum-likelihood estimates of model parameters and chi-square statistics (Jöreskog & Sörbom, 1976). For a proof of these relations, see Bentler & Woodward (1978); for a more general model and computer program, see Weeks (1978). In this article, we use these recent approaches to reevaluate Tesser and Paulhus's theory and to evaluate two new minitheories of love.

Tesser and Paulhus (1976, p. 1099) ob-

The authors wish to thank Abraham Tesser for helpful comments on the manuscript.

Requests for reprints should be sent to P. M. Bentler, Department of Psychology, University of California, Los Angeles, California 90024.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3701-0124$00.75

tained the variances and covariances among four measures associated with the concept of love, assessed for 202 subjects on two occasions separated by 2 weeks. The first variable indicates how often the participant *thought* about a target love person during the previous 2 weeks (T_1 and T_2, where "1" and "2" indicate the measurement occasion). *Love* (L_1 and L_2) was assessed by Rubin's (1970) nine-item love scale. *Reality constraints* (RC_1 and RC_2) were assessed via a scale of new information obtained about the target person that confirmed (high RC) or contradicted (low RC) expectations. Finally, *dating* (D_1 and D_2) was indexed by the subject's report of number of dates with the target person during the previous 2 weeks. Any proposed miniature theory should account for the variances and covariances among variables on both occasions in a statistically acceptable manner; the means of the variables are irrelevant to the causal modeling formulation.

Tesser and Paulhus developed two causal models for their data. In these models, all variables on the first measurement occasion are allowed to correlate freely, and certain variables at the second occasion are functions of particular first-occasion variables, some simultaneous Occasion 2 influences, and residual or unknown effects. The details of the Tesser–Paulhus (1976) original model (p. 1098) and the final model (p. 1100) need not be described here, since, unfortunately, the Tesser–Paulhus models do not fully fit their data. Although Tesser and Paulhus (1976, 1978) apparently believe that a successful model has been developed to fully describe the measures and the 36 variances and covariances, the models only fit 28 of the dispersions, and arguments about the meaning of the models (Smith, 1978; Tesser & Paulhus, 1978) are limited to the subset of relationships. It can be shown that the original model, fit to all the data, yields as an index of fit a χ^2 value of 25.2, with 9 degrees of freedom, indicating that the sample covariance matrix is extremely unlikely ($p = .003$) to be drawn from a population having the proposed causal structure. Although Tesser and Paulhus's final model is reasonable for 28 of the 36 variances and covariances (i.e., by excluding reality constraints on the second testing), it is not capable of explaining all the data. When reality constraints on the second occasion are included, one obtains a χ^2 of 55.6, with 14 degrees of freedom ($p < .0001$), thus requiring similar rejection of the final model.[1]

Since the Tesser–Paulhus theory is not consistent with the data, one may inquire whether any other theory could be developed that would be successful in describing their data. Such a theory should properly be based on prior psychological theorizing. The purpose of this note is to explore two alternative miniature theories of love, based on research in interpersonal attraction, and to test these models in the Tesser–Paulhus data using maximum-likelihood estimation and chi-square goodness-of-fit tests. The first formulation is based on the well-known idea of unidimensionality of interpersonal attraction, whereas the second approach is a refinement of the Tesser–Paulhus view of love. Both formulations will be shown to be capable of explaining all the data involved.

A Unidimensional, Latent-Variable Approach to Love

Interpersonal attraction is often defined as "attitudinal positivity" (Huston & Levinger, 1978), which implies a general construct of favorability–unfavorability towards a partner. Consequently, it is not surprising that "most researchers have assumed that attraction is a unidimensional variable" (Berscheid & Walster, 1978, p. 3). As a consequence, one possible model for the Tesser–Paulhus love data would propose that the four variables being measured represent indicators of generalized love and attraction

[1] It should be noted that the general mathematical models and computer programs for their implementation were not widely distributed at the time Tesser and Paulhus did their research. Given the theoretical formulations and software available at the time, the Tesser–Paulhus solution of deleting one variable was quite appropriate, and their model fits the data within that context. Newer methodology, however, makes it possible to develop a more complete and robust model for these data.

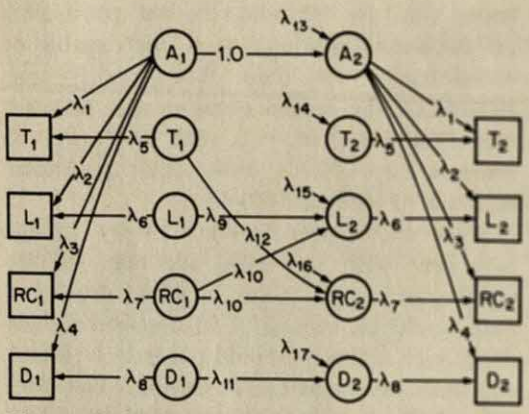


Figure 1. Schematic representation of a simple latent-variable (λ) model linking general attraction (A), love (L), thought (T), reality constraints (RC), and dating (D) at two different times (1 and 2).

toward a partner. Unfortunately, such a model was never evaluated for the Tesser–Paulhus data. In such a model, one would suppose that all of the observed variables are reflections of a single latent construct (a "common factor") of attraction (A) towards the target person. Additionally, it seems reasonable to posit that each of the variables also measures its own special quality or attribute (a "unique factor"), so that one can propose that the four overt variables on the first occasion represent the five latent variables of general attraction, A_1, and the specific attributes T_1, L_1, RC_1, and D_1. That is, each observed variable is composed of, or in part attributable to, general attraction and uncorrelated specific influences including measurement error. Since the times of measurement were separated by the short period of 2 weeks, the most simple unidimensional model of these love measures would additionally propose that the identical measurement structure exists on the second occasion (i.e., the loadings of the four variables on the common and unique factors are the same at both times). Furthermore, the most parsimonious way of interrelating the measures across time is to posit a simple reliability model, in which the second-occasion constructs are linear functions of only the associated first-occasion construct plus a residual. Thus, for example, $T_2 = wT_1 + e$, where w is a weight and e is a residual (i.e., the unique part of thinking about an individual at the second time is adequately predicted by regressing it only on the unique part of thinking about the individual at the earlier time). This particular latent-variable model does not adequately fit the data, $\chi^2(18) = 40.5$, $p < .002$, thus lending some credence to the Berscheid-Walster (1978) criticisms of simple unidimensionality. However, some slight modifications of this specialized model, constraining the parameters of measurement structure to be identical, does yield a formulation based on the unidimensionality of attraction that adequately fits all the observed variances and covariances.

Figure 1 presents the modified unidimensional reliability model. Since it yields a nonsignificant goodness-of-fit index, $\chi^2(19) = 24.2$, $p = .19$, one can conclude that the observed sample data are consistent with the population model shown. The estimates for the model parameters, their standard errors, and critical ratios (formed by dividing the parmeter estimates, minus the null value of zero, by the respective standard errors) are shown in Table 1.

Table 1
Latent-Variable Model Parameter Estimates, Standard Errors, and Critical Ratios (CR)

Parameter	Estimate	*SE*	CR
	Within-occasion parameters		
λ_1	3.18	.18	17.29
λ_2	16.16	1.10	14.68
λ_3	.29	.10	3.06
λ_4	1.52	.18	8.37
λ_5	1.83	.14	13.35
λ_6	10.31	.71	14.53
λ_7	1.77	.09	20.09
λ_8	2.48	.13	19.21
	Cross-time parameters		
λ_9	.70	.08	8.21
λ_{10}	.20	.05	4.08
λ_{11}	.80	.05	15.33
λ_{12}	−.28	.08	−3.49
	Residual parameters		
λ_{13}	.17	.04	4.07
λ_{14}	.55	.16	3.35
λ_{15}	.79	.14	5.45
λ_{16}	.80	.12	6.87
λ_{17}	.48	.07	6.74

The figure uses a convention, now quite standard in the methodological literature, of representing observed variables by squares. The latent constructs, which are unmeasured, are represented in circles. The entire minitheory proposes directional causal influences, with the causal flow originating in the latent variables and yielding the overt variables (the measurement model) as well as cross-time effects on the latent variables (the structural model). The coefficients λ_i represent the parameters of the model to be estimated (the estimates are given in Table 1). It can be seen that at both occasions, there are five latent constructs, the specific constructs associated with each variable (e.g., T_1) and the general attraction construct, A_1. The parameters of the measurement structure, λ_1 to λ_8, are constrained to be identical on both occasions. The cross-time causal effects are given by parameters λ_9 to λ_{12} (with the two λ_{10} parameters set equal), while λ_{13} to λ_{17} represent regression residuals (i.e., the unpredictable parts of the Time 2 latent constructs). The specific estimated values for the parameters, as reported in Table 1, can be used to reproduce the data to within statistical accuracy. However, the important feature of Table 1 lies in the critical ratios (CR), which are like familiar z ratios, each representing a ratio of the parameter estimate minus the null value of zero to its standard error. Using most familiar cutoff values for this ratio shows that all parameters are important to the model.

The substantive meaning of the simple model in Figure 1 is the following: The general attraction construct on the second occasion, A_2, is caused by a simple weighted sum of general attraction (A_1) and a residual (λ_{13}). That is, general attraction is stable over a 2-week period and is not altered by specific aspects of thought, love, reality constraints, or dating. Each specific construct on the second occasion is given by a weighted sum of the identical construct on the previous occasion plus a residual (e.g., $L_2 = \lambda_9 L_1 + \lambda_{15}$). Thus, each of the specific parts of the measured variables is stable across 2 weeks. The single exception is given by thought on the second occasion (T_2), which is independent of a causal influence of thought at the initial time (T_1); apparently, the specific aspect of thoughts (T_2) cannot be predicted from any variable or construct at initial testing. Finally, there are two cross-variable causal influences across time, with T_1 affecting RC_2 and RC_1 influencing L_2. The meaning of these influences appears to be that the more subjects think about their dates at Time 1, the *more* they subsequently believe that recent, new information about their friend contradicts their expectations; and the more they initially believe that information about their friend confirms expectations, the more they tend to love their friend later.

It is possible that the first cross-variable effect reflects an unrealistic quality associated with excessive initial brooding, which becomes contradicted with real, intervening experience, and that the second cross-variable effect reflects the loving consequences of early cognitive security. While the last two links were necessitated by the data and not postulated in the simple reliability model, the affective consequences of having expectations validated seem consistent with the laboratory studies cited by Clore and Byrne (1977) and are well justified by Tesser and Paulhus (1976). Furthermore, the link between large amounts of initial thought about the partner and believing that new information contradicts expectations reflects a point that is made by Tesser and Paulhus in their original article (1976, p. 1095) in citing Kerckhoff and Davis (1962), who note that a "point so often stressed in the literature is that couples go through a period of idealization and perception distortion which may lead to disillusionment (or 'reality shock') at a later date" (p. 302). It must be remembered, furthermore, that the cross-time links are between those parts of the variables *not* attributable to general attraction or evaluation; within this context, both links seem quite interpretable. Of course, these explanations associated with the model must await testing with further data in the context of a larger causal model including measures of the hypothesized influences. Although our statistical methods are reassuring regarding the internal validity of the model, external validity or generalizability to

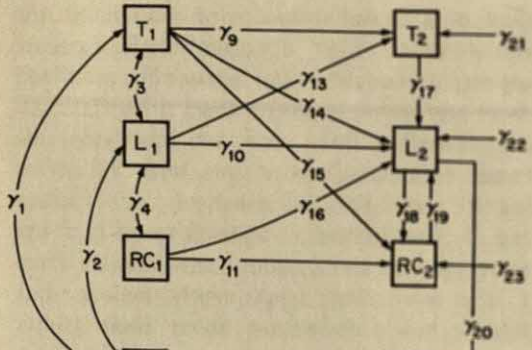


Figure 2. Schematic representation of a simple manifest-variable (γ) model linking love (L), thought (T), reality constraints (RC), and dating (D) at two different times (1 and 2).

other situations must await further research. Smith (1978) has made a similar point.

A Manifest-Variable Approach to Love

Although the above model based on general attraction is a viable one, it is also possible that a model without such a construct may fit the data. Among such models, the simplest one is surely a reliability model for the variables across time. This model proposes that the four variables at Time 1 could be freely intercorrelated but that all causation across time consists of simple links such that a given variable depends causally only on its predecessor (e.g., $T_2 = wT_1 + e$). As might be expected from the previous result, this simultaneous regression model for the four observed, manifest variables measured on two occasions also could not account for all features of the data, $\chi^2(18) = 139.7$, $p < .0001$. Cross-variable causation is obviously essential to enable the model to fit the data, and thus a 24-parameter model was developed that would contain the features of the simple model, while incorporating some of the most necessary causal influences proposed by Tesser and Paulhus. This model is presented in Figure 2, and the parameter estimates, standard errors, and critical ratios are given in Table 2. This model, when estimated by maximum likelihood, has an acceptable index of fit, $\chi^2(12) = 15.9$, $p = .20$. Thus, although the manifest-variable model has 24 parameters as compared to the previous latent-variable model with 17 parameters, the manifest-variable model also cannot be *statistically* rejected for these data.

The manifest-variable model can be considered to be in the spirit of the Tesser–Paulhus approach. It considers the variables at Time 1 to be exogenous (noncausal) variables, with parameters γ_1 to γ_4 associated with covariances among the variables and γ_5 to γ_8 with their variances (not shown in the figure, but given in the table). Unlike the Tesser–Paulhus model, however, the current model found that the estimated covariances of T_1 and RC_1 and of RC_1 and D_1 could be

Table 2
Manifest-Variable Model Parameter Estimates, Standard Errors, and Critical Ratios (CR)

Parameter	Estimate	*SE*	CR
Exogenous parameters			
γ_1	4.43	.79	5.60
γ_2	24.62	4.24	5.81
γ_3	49.79	5.98	8.33
γ_4	4.46	1.68	2.65
$\gamma_5 = \sigma^2(T_1)$	12.89	1.29	10.02
$\gamma_6 = \sigma^2(L_1)$	370.50	36.62	10.12
$\gamma_7 = \sigma^2(RC_1)$	3.24	.32	10.03
$\gamma_8 = \sigma^2(D_1)$	8.24	.82	10.02
Cross-time parameters			
γ_9	.44	.06	6.83
γ_{10}	.66	.06	10.44
γ_{11}	.17	.07	2.43
γ_{12}	.79	.05	16.85
γ_{13}	.08	.01	7.08
γ_{14}	−1.09	.34	−3.19
γ_{15}	−.13	.05	−2.70
γ_{16}	1.52	.43	3.51
Endogenous parameters			
γ_{17}	2.45	.33	7.44
γ_{18}	.03	.01	2.90
γ_{19}	−1.78	.59	−3.01
γ_{20}	.03	.01	4.71
Residual parameters			
γ_{21}	5.04	.50	10.02
γ_{22}	108.04	11.35	9.52
γ_{23}	2.84	.30	9.51
γ_{24}	3.12	.31	10.02

set to zero, so that no parameters are needed for these covariances. (Stated differently, the observed correlations of .129 and .086 could not be considered to be significantly different from 0.) The causal connections across time again required consideration of effects of specific variables on themselves (γ_9 to γ_{12}) as well as some cross-variable effects (γ_{13} to γ_{16}). Two such effects, associated with γ_{15} and γ_{16}, were described in the context of the previous model and would seem to have a similar interpretation in the current model. The remaining effects seem to indicate that loving leads to thinking about one's loved one (γ_{13}) and that extensive thinking at Time 1 leads to *reduced* loving at Time 2. It is as if thinking does not have a positive effect on loving at a later time; but this hypothesis is contraindicated by the endogenous, within-Time-2 effect represented by γ_{17}, which is positive. The remaining endogenous effects, γ_{18} to γ_{20}, are not consistent with the Tesser and Paulhus concept of retrospective assessment on T, RC, and D, which they suggested would cause love, a contemporaneously measured variable. Rather, we find L_2 having an effect on RC_2 (γ_{18}) and on D_2 (γ_{20}), in addition to the effect of RC_2 on L_2 (γ_{19}). These causal paths, though statistically necessary to the model, do not appear to us to be particularly interpretable. Finally, the parameters γ_{21} to γ_{24} reflect the residuals of the causal model.

Discussion

Although we have developed a path analysis model at the level of the manifest variables that would indeed account for the observed data reported by Tesser and Paulhus, we agree with Smith (1978) that causal models of love should not be formulated at the level of overt variables. Not only is the model of Figure 2 likely to be modified solely as a consequence of arbitrary changes in the reliability of the variables involved, but considering all variables at Time 1 as exogenous and outside the explanatory system seems to be shortsighted. The unidimensional latent-variable model not only accounts for the overt variables at Time 1, it also provides for a very simple representation of effects across time and does not require complicated reciprocal causation among variables at Time 2. Since the latent-variable model provides an accurate description of the data with seven fewer parameters, parsimony also favors the latent-variable model. Furthermore, this latent-variable model is consistent with many research approaches to interpersonal attraction as a unidimensional construct (cf. Huston, 1974). Finally, although the manifest-variable model is *statistically* acceptable for the data, there are clear theoretical ambiguities that make this formulation of dubious validity. For example, the opposing signs associated with the effects of thought on love (γ_{14} vs. γ_{17}) are not consistent with the observed positive intercorrelations among all four thought and love variables. While the correlation between thought at Time 1 and love at Time 2 is .612, the associated regression weight is -1.09, suggesting an uninterpretable suppressor effect.

A glance at Figures 1 and 2 will provide rapid visual evidence for the dissimilarities among the models, but certain similarities should also be observed. In particular, both models propose an effect for a given variable on itself across time. In the manifest-variable model, all these effects are important, while in the latent-variable model all but the reliability of the *specific* parts of thought (T_1 and T_2) are necessary. The two cross-variable paths (λ_{12} vs. γ_{15} and λ_{10} vs. γ_{16}) have also been noted and discussed previously. Since these effects appear to be similar across two completely different formulations of the phenomenon, it is likely that they represent fundamental features of the experience of love.

In contrast to the original models of Tesser and Paulhus (1976), which were debated by Smith (1978) and Tesser and Paulhus (1978), the current formulations are consistent with all the data collected. While we do share the concerns of Smith (1978) that causal modeling should be theory based and a priori, the unidimensional model we have presented is consistent with current psychological formulations of general interpersonal attraction as operationalized in existing laboratories around the country, as well as being a slight modifi-

cation of the most mathematically simple model for the data. Of course, it must be recognized that in moderately sized samples such as the present one, minor modifications can be made in the present models that would yield indexes of fit approximately comparable to those we have obtained; such modifications should, however, incorporate all the measurement and structural influences found to be necessary here.

Since causal models of nonexperimental data can be modified substantially when other crucial causal influences are simultaneously considered, the current models must not be considered as final theoretical formulations of love, but rather as potentially interesting representations of the Tesser–Paulhus data.[2] As additional longitudinal data are collected about love, it should prove useful to include the present models as a subset of some larger, more comprehensive formulation in order to account for changes over time in attraction and emotional responses to significant others, and to attempt to reject the proposed simple model in explaining comprehensively the phenomena of interpersonal attraction and love.

[2] Unfortunately, the sample of Tesser and Paulhus is too small to allow a cross-validation study of the models.

References

Bentler, P. M. Multistructure statistical model applied to factor analysis. *Multivariate Behavioral Research*, 1976, *11*, 3–25.

Bentler, P. M. The interdependence of theory, methodology, and empirical data: Causal modeling as an approach to construct validation. In D. B. Kandel (Ed.), *Longitudinal research on drug use*. New York: Wiley, 1978.

Bentler, P. M., & Woodward, J. A. A Head Start re-evaluation: Positive effects are not yet demonstrable. *Evaluation Quarterly*, 1978, *2*, 493–510.

Berscheid, E., & Walster, E. H. *Interpersonal attraction*. Reading, Mass.: Addison-Wesley, 1978.

Clore, G. L., & Byrne, D. The process of personality interaction. In R. B. Cattell & R. M. Dreger (Eds.), *Handbook of modern personality theory*. Washington, D.C.: Hemisphere, 1977.

Huston, T. L. (Ed.). *Foundations of interpersonal attraction*. New York: Academic Press, 1974.

Huston, T. L., & Levinger, G. Interpersonal attraction and relationships. *Annual Review of Psychology*, 1978, *29*, 115–156.

Jöreskog, K. G. Structural equation models in the social sciences: Specification, estimation, and testing. In P. R. Krishnaiah (Ed.), *Proceedings of the symposium on applications of statistics*. Amsterdam: North-Holland, 1977.

Jöreskog, K. G., & Sörbom, D. *LISREL III: Estimation of linear structural equation systems by maximum-likelihood methods*. Chicago: National Educational Resources, 1976.

Kerckhoff, A. C., & Davis, K. E. Value consensus and need complementarity in mate selection. *American Sociological Review*, 1962, *27*, 295–303.

Rubin, Z. Measurement of romantic love. *Journal of Personality and Social Psychology*, 1970, *16*, 265–273.

Smith, E. R. Specification and estimation of causal models in social psychology: Comment on Tesser and Paulhus. *Journal of Personality and Social Psychology*, 1978, *36*, 34–38.

Tesser, A., & Paulhus, D. L. Toward a causal model of love. *Journal of Personality and Social Psychology*, 1976, *34*, 1095–1105.

Tesser, A., & Paulhus, D. On models and assumptions: A reply to Smith. *Journal of Personality and Social Psychology*, 1978, *36*, 40–42.

Weeks, D. G. *Structural equation systems on latent variables within a second-order measurement model*. Unpublished doctoral dissertation, University of California, Los Angeles, 1978.

Wiley, D. The identification problem for structural equation models with unmeasured variables. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences*. New York: Seminar, 1973.

Received March 2, 1978

Journal of Personality and Social Psychology
1979, Vol. 37, No. 1, 131-146

Statistically Combining Independent Studies: A Meta-Analysis of Sex Differences in Conformity Research

Harris M. Cooper
University of Missouri—Columbia

Traditional (literary) reviews of research in social psychology are compared with a statistical approach. It is concluded on both abstract and practical grounds that adoption of the statistical approach would lead to theoretical progress for the research area covered. A meta-analysis "package" is described and then applied to the question of whether there are sex differences in degree of conformity. The meta-analysis is yoked to a literary analysis, and conclusions of differing direction and confidence appear. Problems in application are encountered, and appropriate courses of action are discussed. Finally, limitations on the power of the procedure are outlined.

The traditional way to review research in social psychology has been to take a literary approach. That is, summary statements about research areas are usually based on impressions gleaned by the reviewer from a reading of related studies. This article takes issue with the efficiency of such an approach. An example of a *statistical* technique for combining the results of independent experiments is provided, and this approach is contrasted with a literary summary. The quantitative procedures describe numerically the characteristics of a body of evidence and give a probability level related to the observed pattern of results. Although the actual numerical manipulations require little statistical sophistication, the theoretical progress possible through their employment may be great (progress is here defined as more precise and confident statements about segments of our world). The reasons for believing that progress will follow the adoption of quantitative procedures are both abstract and practical. The abstract reasons are presented first, and the practical reasons are given after the illustration is in hand.

Theoretical progress within a scientific discipline is intimately tied to two other kinds of advancement, one relating to methodology and the other relating to the volume of research being produced. Methodologically, advancement can occur in either research design or analysis, but both advancements typically involve the development of increasingly precise measurement instruments. In the case of design refinements, precision of measurement permits the observation of events that were inaccessible before. An example of this kind of advancement would be, say, the introduction of videotape to the study of nonverbal behavior. Analysis refinements, on the other hand, allow for the more exact description of observed phenomena. Such an advance would be the development of factorial analysis of variance. Factorial designs allow researchers to reduce error (in comparison to one-way designs) and to study interaction. An increased ability to define a concept operationally, to measure it soundly and powerfully, and to study it in relation to other concepts will clearly influence related theoretical developments.

The second kind of advance related to theoretical progress is the accumulation of

Thanks are extended to Robert Arkin, Bruce Biddle, Anthony Greenwald, Robert Rosenthal, and four reviewers for comments on this article.

Requests for reprints should be sent to Harris M. Cooper, Center for Research in Social Behavior, University of Missouri, 111 East Stewart Road, Columbia, Missouri 65201.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3701-0131$00.75

cation of the most mathematically simple model for the data. Of course, it must be recognized that in moderately sized samples such as the present one, minor modifications can be made in the present models that would yield indexes of fit approximately comparable to those we have obtained; such modifications should, however, incorporate all the measurement and structural influences found to be necessary here.

Since causal models of nonexperimental data can be modified substantially when other crucial causal influences are simultaneously considered, the current models must not be considered as final theoretical formulations of love, but rather as potentially interesting representations of the Tesser–Paulhus data.[2] As additional longitudinal data are collected about love, it should prove useful to include the present models as a subset of some larger, more comprehensive formulation in order to account for changes over time in attraction and emotional responses to significant others, and to attempt to reject the proposed simple model in explaining comprehensively the phenomena of interpersonal attraction and love.

[2] Unfortunately, the sample of Tesser and Paulhus is too small to allow a cross-validation study of the models.

References

Bentler, P. M. Multistructure statistical model applied to factor analysis. *Multivariate Behavioral Research*, 1976, *11*, 3–25.

Bentler, P. M. The interdependence of theory, methodology, and empirical data: Causal modeling as an approach to construct validation. In D. B. Kandel (Ed.), *Longitudinal research on drug use*. New York: Wiley, 1978.

Bentler, P. M., & Woodward, J. A. A Head Start re-evaluation: Positive effects are not yet demonstrable. *Evaluation Quarterly*, 1978, *2*, 493–510.

Berscheid, E., & Walster, E. H. *Interpersonal attraction*. Reading, Mass.: Addison-Wesley, 1978.

Clore, G. L., & Byrne, D. The process of personality interaction. In R. B. Cattell & R. M. Dreger (Eds.), *Handbook of modern personality theory*. Washington, D.C.: Hemisphere, 1977.

Huston, T. L. (Ed.). *Foundations of interpersonal attraction*. New York: Academic Press, 1974.

Huston, T. L., & Levinger, G. Interpersonal attraction and relationships. *Annual Review of Psychology*, 1978, *29*, 115–156.

Jöreskog, K. G. Structural equation models in the social sciences: Specification, estimation, and testing. In P. R. Krishnaiah (Ed.), *Proceedings of the symposium on applications of statistics*. Amsterdam: North-Holland, 1977.

Jöreskog, K. G., & Sörbom, D. *LISREL III: Estimation of linear structural equation systems by maximum-likelihood methods*. Chicago: National Educational Resources, 1976.

Kerckhoff, A. C., & Davis, K. E. Value consensus and need complementarity in mate selection. *American Sociological Review*, 1962, *27*, 295–303.

Rubin, Z. Measurement of romantic love. *Journal of Personality and Social Psychology*, 1970, *16*, 265–273.

Smith, E. R. Specification and estimation of causal models in social psychology: Comment on Tesser and Paulhus. *Journal of Personality and Social Psychology*, 1978, *36*, 34–38.

Tesser, A., & Paulhus, D. L. Toward a causal model of love. *Journal of Personality and Social Psychology*, 1976, *34*, 1095–1105.

Tesser, A., & Paulhus, D. On models and assumptions: A reply to Smith. *Journal of Personality and Social Psychology*, 1978, *36*, 40–42.

Weeks, D. G. *Structural equation systems on latent variables within a second-order measurement model*. Unpublished doctoral dissertation, University of California, Los Angeles, 1978.

Wiley, D. The identification problem for structural equation models with unmeasured variables. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences*. New York: Seminar, 1973.

Received March 2, 1978

Journal of Personality and Social Psychology
1979, Vol. 37, No. 1, 131-146

Statistically Combining Independent Studies: A Meta-Analysis of Sex Differences in Conformity Research

Harris M. Cooper
University of Missouri—Columbia

Traditional (literary) reviews of research in social psychology are compared with a statistical approach. It is concluded on both abstract and practical grounds that adoption of the statistical approach would lead to theoretical progress for the research area covered. A meta-analysis "package" is described and then applied to the question of whether there are sex differences in degree of conformity. The meta-analysis is yoked to a literary analysis, and conclusions of differing direction and confidence appear. Problems in application are encountered, and appropriate courses of action are discussed. Finally, limitations on the power of the procedure are outlined.

The traditional way to review research in social psychology has been to take a literary approach. That is, summary statements about research areas are usually based on impressions gleaned by the reviewer from a reading of related studies. This article takes issue with the efficiency of such an approach. An example of a *statistical* technique for combining the results of independent experiments is provided, and this approach is contrasted with a literary summary. The quantitative procedures describe numerically the characteristics of a body of evidence and give a probability level related to the observed pattern of results. Although the actual numerical manipulations require little statistical sophistication, the theoretical progress possible through their employment may be great (progress is here defined as more precise and confident statements about segments of our world). The reasons for believing that progress will follow the adoption of quantitative procedures are both abstract and practical. The abstract reasons are presented first, and the practical reasons are given after the illustration is in hand.

Theoretical progress within a scientific discipline is intimately tied to two other kinds of advancement, one relating to methodology and the other relating to the volume of research being produced. Methodologically, advancement can occur in either research design or analysis, but both advancements typically involve the development of increasingly precise measurement instruments. In the case of design refinements, precision of measurement permits the observation of events that were inaccessible before. An example of this kind of advancement would be, say, the introduction of videotape to the study of nonverbal behavior. Analysis refinements, on the other hand, allow for the more exact description of observed phenomena. Such an advance would be the development of factorial analysis of variance. Factorial designs allow researchers to reduce error (in comparison to one-way designs) and to study interaction. An increased ability to define a concept operationally, to measure it soundly and powerfully, and to study it in relation to other concepts will clearly influence related theoretical developments.

The second kind of advance related to theoretical progress is the accumulation of

Thanks are extended to Robert Arkin, Bruce Biddle, Anthony Greenwald, Robert Rosenthal, and four reviewers for comments on this article.

Requests for reprints should be sent to Harris M. Cooper, Center for Research in Social Behavior, University of Missouri, 111 East Stewart Road, Columbia, Missouri 65201.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3701-0131$00.75

evidence concerning a specific topic. Repeated observation allows for greater confidence in uncovered relations, when observations show consistent or similar results. Repeated results also increase confidence in the generalizability of observed phenomena, with intended and unintended differences in replications testifying to the robustness of relations. Statements that can be made with great confidence undoubtedly motivate future research to take a more precise theoretical tack. Methodological and accumulation advances are connected, in that increased precision of measurement reduces the number of observations necessary for relations to be accepted at a given level of confidence. In sum, then, it can be argued that refined measurements and the accumulation of aligned studies are two major spurs to theoretical progress.

With these relations in mind, one can argue that the traditional literature review in social psychology lacks analytic precision in at least three ways. First, literary reviews may be more susceptible to the idiosyncracies of a particular reviewer's perspective than are reviews using consensually validated statistical procedures. As Gene Glass (1976) states:

> A common method for integrating several studies with inconsistent findings is to carp on the design or analysis deficiencies of all but a few studies—those remaining frequently being one's own work or that of one's students or friends—and then advance the one or two "acceptable" studies as the truth of the matter. (p. 4)

Second, the literary review usually ignores the issue of relationship strength by rarely assessing the size of the effect under study. Lastly, the typical review imprecisely weights conclusions with respect to the volume of available evidence (Light & Smith, 1971).

It seems, then, that replacing the present literary model of research view with a statistical model would almost necessarily increase measurement precision. In addition, this particular replacement relates directly to the issue of accumulating evidence. Areas having statistical combinations, therefore, should benefit theoretically from the more exact description, sounder inferences, and greater confidence in observed phenomena that such a technique allows. Again to quote Glass (1976), statistical analysis "connotes a rigorous alternative to the casual, narrative discussions of research studies which typify our attempts to make sense of the rapidly expanding research literature" (p. 3).

A methodological advance cannot occur unless the development of a field is compatible with it. Many areas of social psychology are, I feel, compatible with the advancement proposed here. This judgment is based on two observations. First, certain areas have generated enough research on identical or conceptually similar topics that statistical procedures can be reliably applied to them. Many areas in social psychology can now cite 20, 40, or more similar hypothesis tests, as the example to be presented here illustrates. These areas seem "ready" for more precision. A second reason for the proposed compatibility between field and method is an increasing awareness concerning the importance of legitimate data analysis. Partially in response to McGuire's (1973) call for redirection, social psychology training programs tend now to require advanced course work in statistical procedures (I am myself a product of this "trend"). In addition, the major journals in the field increasingly present research analyzed with complex statistical techniques. While the procedures presented here require only a knowledge of the fundamental probability axioms behind nearly all published analyses, this increased exposure to statistics may remove some of the "inertia" that typically surrounds a newly introduced analytic procedure.

The term *newly introduced* was used above because the methods presented do not make their debut in this article. Before turning to the procedures themselves, however, let me summarize the arguments for their adoption. First, it has been noted that measurement advances and aligned research efforts can be expected to lead to theoretical progress. Second, within social psychology, more precise methodologies are available than the method presently used for drawing inferences from multiple tests of related hypotheses. Social theory, therefore, may benefit if these more precise analytic techniques are employed more frequently. Certain areas may be

"ready" for such precision because the volume of research available permits statistical testing and because researchers may now feel quite positive toward rigorous analytic procedures.

Meta-Analytics

Glass (1976) has suggested that data analysis can be categorized into three distinct activities: primary, secondary, and meta-analysis. *Primary* analysis occurs when the data of a given study are analyzed for the first time. *Secondary* analysis is the reanalysis of data. This may occur so that more appropriate statistical procedures can be applied or so that new questions can be answered with an old data set. Finally, *meta-analysis* occurs when the results of independent experiments are combined "for the purpose of integrating the findings" (p. 3).

Methods for meta-analysis have a history as long as most commonly used statistical procedures. In 1938, Fisher and Pearson both presented methods for obtaining an overall probability for a series of studies (Fisher, 1938, pp. 104–106; Pearson, 1938). In social psychology, Mosteller and Bush (1954) presented a meta-analysis procedure in the *Handbook of Social Psychology*. While early instances of meta-analysis applications are rare, they now appear with some frequency. Rosenthal (1976, p. 441) presents a meta-analysis testing the influence of interpersonal expectations on behavior. Over 300 studies are included. He reports that nearly 50,000 failures to find an influence would be needed to reverse, at the .05 level of significance, the conclusion that expectations affect behavior. Glass (1976) cites meta-analyses relating to teaching methods, television instruction, and socioeconomic status relations to IQ, to name a few. Smith and Glass (1977) present a meta-analysis integrating the literature on the outcomes of psychotherapy. Their conclusion, based on 375 studies, is that "therapy of any type under . . . average conditions can be expected to move the typical client from the 50th to the 75th percentile of the untreated population" (Glass, 1976, p. 7). Glass also concludes that these studies exhibit only a "trivial" advantage for behavioral over nonbehavioral therapies.

As the above examples illustrate, a meta-analysis is conducted on a group of studies related either because (a) they share a common conceptual hypothesis or (b) they share operations for the realization of the independent or dependent variables, regardless of conceptual focus. Meta-analytics result in a single set of numbers describing the body of literature involved. Typically, a significance level is generated, which addresses the question "What is the probability that a set of studies exhibiting these results could have been generated if no actual relation existed?" Other statistics are usually presented also, such as Rosenthal's assessment of the number of studies needed to reverse a conclusion and Glass's description of the percentile relation between experimental groups.

Meta-analytic procedures have been developed that range in sensitivity. The least sensitive test simply "counts" the number of probability levels below .50 (chance for the predicted direction) and asks whether this number is too large. The most sensitive procedures require that the raw data from included experiments be available (Light & Smith, 1971). Rosenthal (1978) has presented a detailed description of available meta-analytic methods, including their relative advantages and limitations. This effort will not be repeated here. The method to be illustrated presently asks only that probability levels from included studies be available. These levels are first transformed to Z scores (standard normal deviates) and then related to the number of studies involved. This procedure is presented because it is more precise than the counting method and more accessible than methods using raw data. In addition, the illustrated method requires no assumptions concerning data beyond those typically made by the authors of the original studies (i.e., linearity, normality, homogeneity of variance, independence of events) and requires that each study exhibit unit variance. The method was introduced by Stouffer in 1949 and then modified by Mosteller and Bush in 1954.

The weighted and unweighted Stouffer

method. The Stouffer (1949, p. 45) method for combining studies uses the following formula:

$$Z_{ma} = \frac{Z_{s1} + Z_{s2} + \cdots Z_{sn}}{\sqrt{N_s}}, \qquad (1)$$

where Z_{ma} = the standard normal deviate for the meta-analysis; $Z_{s1} \cdots Z_{sn}$ = the standard normal deviate for each included study; and N_s = the total number of studies included. The method asks that the reviewer take the following steps: (a) Record the probability level reported in each study associated with the relevant hypothesis; (b) turn to the Z score (standard normal deviate) table appearing in any elementary statistics text and find the Z score associated with each probability level; (c) sum these Z scores and divide this sum by the square root of the number of studies involved; and (d) refer this Z score back to the table and record the appropriate probability level. This probability describes the likelihood that the included studies' results were generated by chance. The user should be alert to the fact that a direction for the hypothesis should be chosen before the procedure is begun (Mosteller & Bush, 1954, p. 330). This means that a reported two-tailed p value supporting the hypothesis under consideration should be halved before the associated Z score is retrieved. A disconfirming p value should be halved and its associated Z score *subtracted* from the sum of the supporting studies.

The Stouffer method has many advantages, not the least of which is ease of employment. As Mosteller and Bush (1954) point out, however, even more precision can be built into this procedure if the reviewer weights (or differentially assigns importance to) studies within the body of literature. The Stouffer method allows the reviewer to weight each standard normal deviate by the size of the sample on which it is based or by other desirable characteristics such as internal or external validity. Thus, a reviewer can make adjustments based on methodological rigor before the statistic is generated. Studies that are viewed as "sound" can be weighted more heavily than "unsound" studies, if the reviewer so desires.

The formula for the weighted Stouffer method, when sample size is the weighting factor, is:

$$Z_{ma} = \frac{N_{s1}Z_{s1} + N_{s2}Z_{s2} + \cdots N_{sn}Z_{sn}}{\sqrt{N_{s1}^2 + N_{s2}^2 + \cdots N_{sn}^2}}, \qquad (2)$$

where $N_{s1} \cdots N_{sn}$ = the number of subjects in each included study, and all other quantities are defined as before. In this case, the modification requires that each study's Z score be multiplied by the number of subjects in the study before the scores are summed. This numerator is then divided by the square root of the squared and summed individual study sizes. The formula given here is weighted by sample sizes rather than another criterion because sample size is the most objective weighting criterion. Other less objective criteria have been convincingly employed, however (e.g., Stickell, 1963). The sample-size weighting means that larger studies are given greater importance. In comparison to the unweighted method, the weighted method will lead to a lower probability level when larger studies produce larger Z scores and a higher probability when smaller studies produce larger Z scores.

Studies needed to reverse a conclusion. We can state with some confidence that no literature review will uncover every relevant hypothesis test ever conducted (Rosenthal, in press, has referred to this as the "file drawer problem"). The Stouffer method has the advantage of allowing us to easily compute what could be called a fail-safe number. With the unweighted procedure we can ask the question "How many studies totaling a null hypothesis confirmation would be needed to reverse the conclusion that a relationship exists?" To state this numerically, we ask, "How many studies showing a *summed* Z score total of zero would be necessary to raise the overall probability level to about .05, or perhaps .10?" For the $p < .05$ level, the fail-safe formula would be:

$$N_{fs.05} = \left(\frac{Z_{s1} + Z_{s2} + \cdots Z_{sn}}{1.645}\right)^2 - (N_s), \qquad (3)$$

where $N_{fs.05}$ = the number of additional studies needed to increase the meta-analysis probability to above .05, and all other quan-

tities are defined as before. This formula is a direct algebraic manipulation of the unweighted formula. It requires that (a) the sum of the Z scores for known studies be divided by 1.645 (the Z score associated with $p < .05$, two-tailed); (b) this quantity be squared; and (c) the number of known studies be subtracted from it. The resulting quantity is the number of *additional* studies showing a summed null relationship needed to increase the meta-analysis probability above the .05 level of significance. Other significance levels can be used by replacing the denominator on the right with the critical Z value. Obviously, the fail-safe N should be presented only when a significant overall probability level can be reported.

The fail-safe N is an important descriptive statistic in that it allows a reader to easily evaluate the "strength" exhibited in a review against the felt completeness of the reviewer's sampling procedure. However, a limitation of the fail-safe N should also be pointed out. It is an appropriate guide for the reader *only* if the assumption of a summed null relation in undiscovered studies is acceptable. It is always possible that a smaller number of studies exist that have a summed Z score of equal but *negative* value to the sum of those reviewed. The plausibility of this alternative also should be considered by a reader. The illustration below includes both a weighted and unweighted Stouffer procedure and a fail-safe N. All three procedures are recommended for future reviews.

Describing relationship strength. Observers of psychological research have long bemoaned the infrequent reporting of measures of relationship strength, or effect sizes, in the literature (e.g. Bakan, 1966; Levy, 1967). As O'Brien and Shapiro (1968) point out, the usual criteria for gauging the importance of a study, its level of significance, are numbers bound tightly to the size of the sample involved. Significance testing, then, which compares an observed relation to the chance of no relation, becomes less informative as evidence supporting a phenomenon accumulates. The question turns from "whether" to "how much." Effect-size measures that are free of sample-size influence, therefore, ought to be an integral part of all literature reviews, especially when the reviewer concludes that, in fact, a relation exists.

There is an effect-size measure appropriate for every kind of research design or analysis. Jacob Cohen (1977) presents a definitive catalogue of these measures. It is the present author's hope that Cohen's work will continuously grow in importance. Here Cohen's d index is illustrated:

$$d = \frac{M_a - M_b}{SD}, \tag{4}$$

where M_a and M_b = the means of the two groups compared, and SD = the pooled estimate of the population standard deviations. The d index is chosen because it is simple, it is "scale free" (i.e., the standard deviation adjustment indicates that studies using different measurement scales can be compared), and it is applicable to a plurality of the studies in social psychology. It requires that only two means be compared, but even in most factorial designs we find that each independent variable is represented by two conditions. If more complex hypotheses are under consideration, alternate effect-size indexes are readily available in Cohen's book.

The d index is a number that tells how far apart the means of our two groups are in terms of their common standard deviations. For instance, if $d = .2$, it means that two tenths of a standard deviation separate the two sample means. In Cohen's terminology, a d index equaling .2 corresponds to a "small" effect size, with .5 being "moderate," and .8 being "large." This classification may leave something to be desired in terms of intuitive appeal. For this reason, Cohen (1977) also presents "percentage of distribution overlap" measures like the one in use by Smith and Glass (1977), cited earlier. The overlap measure illustrated here, called U_3, tells what percentage of the population with the smaller mean is exceeded by 50% of the population with the larger mean. Put more conversationally, "how much of the smaller-meaned group is exceeded by the average person in the larger-meaned group?" A table for converting the d index to U_3 is presented by Cohen (1977, p. 22). As an example, assume Groups

A and B have been compared and a d index of .2 is found, with Group A exhibiting the larger mean. The U_3 measure tells us that the average person in Group A has a score greater than 58% of the people in Group B.

The reader may wonder how this effect-size analysis can be carried out when means and/or standard deviations are not reported in an experimental write-up. Friedman (1968) has shown that when these quantities are not reported, the d index can be estimated from the formula:

$$d = \frac{2t}{\sqrt{df}}, \qquad (5)$$

where t = the reported t value (of $\sqrt{F}$ if analysis of variance is used), and df = the associated degrees of freedom for error. Instances will exist where neither means, standard deviations, nor t values will be reported. This problem is addressed in conjunction with the illustration presented below.

To summarize, a meta-analysis "package" has been suggested that includes the following statistics: (a) a Stouffer Z score; (b) a weighted Stouffer Z score; (c) a fail-safe N; (d) a d index of relationship strength; and (e) a percentage of overlap of distributions measure, U_3. The remaining tasks, then, are to illustrate the application of these procedures and to address the problems one encounters in doing so.

An Illustration: Do Females Conform More Than Males?

An area in social psychology where enough data have been accumulated to warrant meta-analysis is that of sex differences in conformity. Maccoby and Jacklin (1974) uncovered 47 independent tests of this relationship. Sex and conformity is chosen here because (a) the independent variable realization need not be explained; (b) the research designs and dependent measures are familiar to most social psychologists, so limited space is needed for their description; and most importantly, (c) Maccoby and Jacklin's literary combination of studies provides an excellent counterpart to the statistical combination. By choosing an area in which a traditional review already exists and by employing only the earlier cited studies in the meta-analysis, it is possible to compare the conclusions and precision of the two techniques. Essentially, Maccoby and Jacklin's conclusions are supported, though differences in some inferences are also apparent.

As hypotheses for the present study, then, Maccoby and Jacklin's (1974) *interpretation* of the sex differences in conformity literature will be used. The following passages, I believe, best summarize their inferences from the body of evidence:

> 1. All that can be said at this point is that the results are inconsistent, and that when Asch experiments are considered in conjunction with other conformity studies, neither sex shows an overall tendency to be more susceptible to social influence from peers. (p. 271)
>
> 2. In face-to-face encounters, when an individual must openly disagree with the opinion of others, as is the case in the Asch situation, women somewhat more often conform to others' judgments, but inconsistency of the findings and the frequency of sex similarity are striking. (p. 268)
>
> 3. Studies on the effects of persuasive communications have a more consistent outcome . . . On these grounds no overall difference in susceptibility to social influence would be expected. In any case . . . none is found. (p. 268)

Hypothesis confirmation in this meta-analysis means not only that the predicted relations are supported but also that Maccoby and Jacklin's earlier inferences have been upheld.

The studies sampled. The literature serving as data was Maccoby and Jacklin's 47 cited studies. No attempt was made to supplement or update the earlier review. This was done so that the statistical combination was yoked to the literary combination according to evidential base.

The data included 35 studies appearing in the *Journal of Personality and Social Psychology*, 5 in *Developmental Psychology*, 4 in *Child Development*, 2 in the *Journal of Social Psychology*, and one in the *Journal of Experimental Child Psychology*. The mean year of publication is 1969, with the earliest year of appearance 1958, and the latest, 1974. Twenty-one of the studies were published between 1968 and 1970. Twenty-four studies

used college-age samples, with the rest choosing younger subjects.

The most striking characteristic of the sample is that only published studies are included. Maccoby and Jacklin (1974) note that "some information is lost owing to the selection that occurs in the publication process itself" (p. 3). Most distressing is that, as Maccoby and Jacklin point out, "negative findings probably constitute the most frequent omissions from the literature" (p. 5). Essentially, two kinds of bias are operating here. First, we are less likely to see null results in published form. Second, we are less likely to see results that contradict previously published findings. This selection bias operates both on the decision of researchers to submit reports and on journal reviewers. The bias in the sampling process highlights the importance of calculating the fail-safe *N* before making inferences. Readers of a review should ask themselves, "How many studies are missing, given the sampling procedure employed?" A fail-safe *N* surpassing this estimate should increase the readers' confidence in the conclusions.

Sampling procedures more exhaustive than Maccoby and Jacklin's choice are available. Convention programs, computer-assisted searches, and *Dissertation Abstracts International* are all alternate or supplemental sources to professional journals. Rosenthal (1976) applies the interesting technique of separately analyzing dissertation research, under the assumption that this sample contains the least "confirmatory" bias (mitigating the "confirmatory" bias in published papers, by the way, is the argument that published research is conducted by the more capable researchers).

The analysis design. To test the first hypothesis (that no overall difference in conformity associated with gender appears in the literature), the studies were analyzed as a single group. To test the second and third hypotheses (relating to situation distinctions), the studies were divided into three categories. The first category contained Asch-type (Asch, 1956) experiments, in which subjects had face-to-face interaction with either other subjects or an experimental confederate. The second category included studies where conformity to a fictitious group norm or to a simulated other was measured. The final category was composed of persuasive communication or social reinforcement (Insko, 1965; Insko & Cialdini, 1969) experiments. The Asch-type studies used primarily agreement with a group of peers' erroneous perceptual judgment as the dependent variable. Fictitious group norm studies used primarily opinion change, and persuasive communication studies used mainly attitude change to measure conformity. It is not clear whether Maccoby and Jacklin's (1974) distinction concerning face-to-face interaction intended the categorization employed here. Their remark may refer to both in-person and simulated-other experiments. In either case, the results support the face-to-face distinction.

This three-way classification led to the exclusion of three studies because they were correlational (Baumrind & Black, 1967; Samorjzyk, 1969; Whiting, 1963) and two studies because they related to behavior in experimental situations (Cook et al., 1970; Dillehay & Jernigan, 1970). In addition, three studies were excluded because they were conducted on non-American subject populations (Beloff, 1958; Bronfenbrenner, 1970; Frager, 1970). Of the excluded studies, Bronfenbrenner reported that Russian males were more conforming than Russian females, and Cook et al. reported that females were more compliant to experimental deception. No other excluded study reported significant sex differences. Finally, the study by Sistrunk and McDavid (1971) was not included because of its special place in the literature, to be noted later. The exclusions encompassed 19% of the sample but permitted the following generalizations to be made about the studies remaining in the analysis: (a) They all involved peer influence; (b) they all were performed using white Americans (or results reported here are relevant to this population); (c) they all included an experimenter-controlled introduction of the conformity-producing stimulus.

After the 9 studies not fitting into a category were dropped, 38 studies remained. Sixteen studies fell into the Asch-type category, 8 into the group norm category, and 14

Table 1
Asch[a]-Type Conformity Experiments

Author	Year	*N*	Maccoby & Jacklin's (1974) *p*	Retrieved *Z* score	Retrieved *d* index	Retrieved U_3 (%)
Allen & Newston	1972	366	.50	−.84	−.08	46
Bishop & Beckman	1971	144	.50	−1.64	−.28	39
Costanzo & Shaw	1966	96	.50	1.55	.35	63
Hamm & Hoving	1969	42	.0025	3.29	1.06	85
Mock & Toddenham	1971	280	.05	3.29	.47	67
Carrigan & Julian	1966	96	.005	3.29	.96	83
Dodge & Muench	1969	122	.50	0	.00	50
Schneider	1970	96	.50	.67	.16	56
Landsbaum & Willis	1971	64	.50	.01	.00	50
Gerard, Wilhelmy, & Conolley	1968	154	.025	1.96	.33	63
Stricker, Messick, & Jackson	1967	190	.50	0	.00	50
Endler	1966	120	.025	2.05	.41	66
Glinski, Glinski, & Slatin	1970	56	.50	0	.00	50
Hollander, Julian, & Haaland[b]	1965	112	.025	2.58	.49	69
Julian, Regula, & Hollander	1968	240	.025	3.29	.56	70
Willis & Willis	1970	96	.50	0	0	50
M	1968.9	142.1	.32	1.22	.28	60

Note. Positive *Z* values denote more female conformity.
[a] Asch (1956).
[b] Hollander et al. used a chi-square statistic; *Z*, *d*, and U_3 are estimated from their probability level.

into the persuasive communication category. Tables 1 through 3 present this categorization. The mean year of journal appearance for each grouping is nearly identical. This indicates that there was no research paradigm shift in sex and conformity studies over the covered period. Also, the tables show that the average Asch-type experiment was conducted using about 17 more subjects than the average group norm experiment (Asch-type = 141.1; group norm = 124; persuasive communication = 132.7). All three experiment-size means, however, are well within a standard deviation of each other (Asch-type SD = 87.6; group norm SD = 72.8, persuasive communication SD = 88.5). This analysis of category differences helps rule out alternative hypotheses, should different results be found for different categories.

Obtaining data about experiments. A number of problems arise when the reviewer attempts to extract from each experimental report the statistics needed for combination (here, *t* values and *p* levels). The most glaring problem is that many studies finding no differences do not report the statistics upon which this inference is based. In the present analysis, 12 of the 38 categorized studies (32%) reported no statistics. Creating even more frustration was the finding that 4 studies (11%) reported *t* (or *F*) values without mentioning the direction of the effect! In each case, a simple finding of "no difference" is reported. Although the premium on space in professional journals is appreciated, much of this "planned obsolescence" that researchers build into their reports can be avoided. Instances were found in which analysis of variance summary tables were presented with nonsignificant *F* values replaced by lines. This does not save space. The statement that "no differences were found between Groups A and B" can be replaced with the statement that "a nonsignificant difference was found, with Group A higher than Group B." At the least, directional statements along with their statistical basis ought to appear for all main effects tested. The increase in length that this may entail will be amply repaid through the increased amount of time the report will contribute to the area of interest. It should also be noted that the present

Table 2
Experiments With Fictitious Group Norms

Author	Year	*N*	Maccoby & Jacklin's (1974) *p*	Retrieved *Z* score	Retrieved *d* index	Retrieved U_3 (%)
Hamm	1970	216	.50	0	.0	50
Sistrunk & McDavid	1971	80	.50	0	.0	50
Lefurgy & Woloshin	1969	53	.50	0	.0	50
Wyer	1966	80	.50	1.64	.41	65
Sampson & Hancock	1967	251	−.025	−2.05	−.30	−62
Sistrunk[a]	1971	64	.50	0	.0	50
Endler & Hoy	1967	120	.50	.52	.0	50
Wyer	1968	128	.50	−.33	−.04	50
M	1968	124	.44	−0.03	.01	51

Note. Negative *Z* values denote more male conformity.
[a] Sistrunk reports more conformity among black females than males but no differences among whites.

42%-incomplete reporting rate is probably higher than the rate that will be found in other areas. This is because the experiments included here typically studied sex differences only secondarily. Most researchers will take more care in reporting primary findings.

The best strategy for dealing with incomplete data reports is to assume an exact finding of no difference. Thus, such studies are treated below as having uncovered a *t* value of 0 with a probability of .50. As meta-analysis combined probabilities become lower, we can assume that this procedure increases "conservative" bias in our estimations and inferences.

A second difficulty encountered when statistics are being collected is that some studies employ multiple dependent measures. In such an event, alternate courses of action are open to the reviewer. First, all dependent variables can be included. This procedure means that some experiments will contribute to the overall probability more than others. If the reviewer does not wish this to be so, each dependent variable within a study can be weighted so that the sum of each study's measures is equal. A third strategy is to choose the major dependent variable (i.e., the one closest to the sample description) and exclude all others. This is the strategy used here.

Other problems related to multiple measures arise when (a) separate papers report different dependent measures collected on the same subjects or when (b) the same researcher (laboratory) reports several studies. These problems (as well as the sampling bias referred to above) diminish the *independence* of data involved in meta-analysis. At present, the more elegant statistical adjustments for nonindependence of observations (most notably, adjustments in degrees of freedom) can only be subjectively approximated in meta-analytics. This can be accomplished through the reviewers' weighting decisions. Thus, if Researcher X has produced four similar studies in a 2-year period, a cautious meta-analyst might individually weight these studies less heavily than a similar single study produced from a different laboratory.

Another problem in data collection arises when both parametric and nonparametric statistics appear in the literature. Only one instance of nonparametric use was encountered in this analysis (Hollander, Julian, & Haaland, 1965). The procedure used was to extrapolate all important values from the reported probability level. A more equal mix of statistics based on differing assumptions might lead to the use of other procedures (i.e., combining raw data from expected-frequency-type studies and treating these separately).

A final problem in data retrieval relates to levels of significance. Care should be taken to recompute the relevant significance levels presented in the individual reports. Probabilities falling between .05 and .01 are often

reported as $p < .05$. For testing the null hypothesis, this procedure is legitimate, but more accurate Z scores will be retrieved if more accurate probabilities are computed. With the introduction of high-speed computers, the rounding of p levels is practiced less often. However, the reviewer cannot easily discern whether this has been done.

The reader may have noted that the procedures outlined above (gathering t values, examining means, recomputing probabilities) represent a significant departure from traditional literary reviews regardless of whether studies are statistically combined. It would be interesting, then, to gauge separately the increased precision created by means of data retrieval and by the combining procedures. For this reason, the Stouffer methods have been performed on both the p levels gathered by the procedures outlined above and the p levels presented in Maccoby and Jacklin's (1974) annotated bibliography. These latter probabilities take one of four values: .50, .05, .01, and .001.

Results. In brief, the hypotheses can be stated as follows:

1. When the body of evidence is examined as a whole, neither sex will show an overall tendency to be more susceptible to social influence from peers.

2. In face-to-face encounters, women more often than men will conform to others' judgments. However, (a) there will be inconsistency of the findings, and (b) the frequency of sex similarity will be striking.

3. Studies on the influence of persuasive communications will show no overall difference in susceptibility to social influence.

Testing Hypothesis 1 first entailed combining Z scores over 46 studies (with Sistrunk & McDavid, 1971, excluded). Table 4 presents most resulting statistics (all p levels are reported as one-tailed). When the probabilities reported in Maccoby and Jacklin (1974) served as data, the unweighted Stouffer method led to a Z_{ma} of 2.13 with a one-tailed probability of less than .015. The weighted method produced a Z_{ma} of 1.18 with $p < .12$.

The Stouffer procedures were then repeated with only the Z scores of the 38 categorized studies included. Using the Maccoby and Jacklin probabilities led to an unweighted Z_{ma} of 2.35. Such a score would occur less than 1% of the time given that females actually do not conform more than males. When the Z scores retrieved from the primary

Table 3
Persuasive Communication (Attitude Change) Experiments

Author	Year	N	Maccoby & Jacklin's (1974) p	Retrieved Z score	Retrieved d index	Retrieved U_3 (%)
Dean, Austin, & Watts	1971	161	.50	.52	.12	54
Eagly & Telaak	1972	118	.50	0	0	50
Greenbaum	1966	100	.50	0	0	50
Insko	1965	70	.50	0	0	50
Insko & Cialdini	1969	152	.50	0	0	50
Linder, Cooper, & Jones	1967	53	.50	0	0	50
Marquis	1973	52	.50	0	0	50
Nisbett & Gordon	1967	152	.50	0	0	50
Osterhouse & Brock	1970	160	.50	0	0	50
Rosenkrantz & Crockett	1965	176	.50	0	0	50
Rule & Rehill	1970	90	.50	0	0	50
Silverman	1968	403	.50	1.41	.16	56
Silverman, Shulman, & Wiesenthal	1970	98	.50	0	0	50
Worchel & Brehm	1970	73	.50	0	0	50
M	1968.8	132.7	.50	.14	.02	51

Note. Positive Z values denote more female conformity.

Table 4
Summary of Combined Z Score Results of Three Types of Experiments

	Overall		Asch[a]-type		Fictitious group norm		Persuasive communication	
Procedure	Maccoby & Jacklin (1974)	Retrieved	Maccoby & Jacklin (1974)	Retrieved	Maccoby & Jacklin (1974)	Retrieved	Maccoby & Jacklin (1974)	Retrieved
Unweighted	2.35	3.44	4.11	4.87	.69	−.08	.00	.51
p	.01	.0005	.00005	.000005	.75	.53	.50	.31
Weighted	1.92	3.04	3.56	4.03	−1.23	−.91	.00	1.10
p	.03	.002	.0005	.00005	.89	.82	.50	.14
Fail-safe *N*	40	129	84	125	—	—	—	—

Note. All *Z* scores are based on 38 studies; *p*s are one-tailed.
[a] Asch (1956).

sources were used, the unweighted Z_{ma} rose to 3.44, an event that chance produces less than 5 in 10,000 times. If Maccoby and Jacklin's data are used, the fail-safe *N*, the number of null-totaling studies needed to raise the observed probability to above 5 chances in 100, is 40. Based on the retrieved data, a fail-safe *N* of 129 is found.

It seems, then, that taken as a whole, the body of evidence supports the conclusion that females conform more than males. If we restrict our inference to American subjects in experimenter-controlled, peer-influence situations, we can make this statement more confidently. This conclusion is at odds with Maccoby and Jacklin's, which saw "no overall tendency" for females to be more conforming. In this instance, the literary and statistical combinations led to differing interpretations of the data. In addition, the retrieving of data from primary sources led to a more confident conclusion of sex differences than did the annotated bibliography. The *p* levels produced by the two gathering methods differ by 89 uncovered studies needed to reverse the conclusion.

Tests of Hypotheses 2 and 3 show less dramatic differences between the approaches. Hypothesis 2 entailed analyzing separately the Asch-type and group norm experiments. The unweighted Z_{ma} for only Asch-type experiments was 4.11 with Maccoby and Jacklin's *p* levels and 4.87 with retrieved *p* levels. Both occur by chance less than 1 in 10,000 times. The weighted Z_{ma} scores are again lower, being 3.56 and 4.03, respectively, but are still highly significant ($p < .0005$ and $p < .00005$, respectively). The fail-safe *N* at $p = .05$ for Maccoby and Jacklin's data was 84, whereas the fail-safe *N* for retrieved data exceeded this by 41.

Analysis of group norm experiments produced the first set of nonsignificant results. As Tables 2 and 4 show, there was a slight tendency for group norm studies to report more conforming in males than females (as the negative Z_{ma} score indicates). This tendency was far from significant, however.

Hypothesis 2, then, was clearly supported. As Maccoby and Jacklin (1974) conclude, females do seem to conform more than males in face-to-face situations. The difference in conclusion strength introduced by reporting precision also seems negligible. What does go unsupported is Maccoby and Jacklin's conclusion that the results show inconsistency. In face-to-face encounters, only 2 of 16 studies report "small" effects contradicting the overall conclusion. The appearance of these anomalies could have been generated by undiscovered interactions or by chance. To label the overall results as "inconsistent" because of them is to attach undue importance to their probabilistic nature.

Also unsupported is Maccoby and Jacklin's conclusion that "the frequency of sex similarity" is striking. A more legitimate statement would be that the size of the sex difference may be smaller than the reviewers anticipated or the cultural stereotype predicted.

The effect-size analysis allows these impressions to be quantified. As Table 1 indicates, the average d index for Asch-type experiments was .28. This shows that on the average, the male and female distribution midpoints were .28 standard deviations apart (assuming that normality holds). In terms of overlap (U_3), it is found that the average female conformed more than 60% of the male population (of course, the sample and setting restriction also restrict the generalizability of this statement). Further, since the d index exhibits a standard deviation of .37, we can expect that future experiments will exhibit d indexes between .08 and .48 about 95% of the time. Alternatively, Asch-type studies should show the average female conforming more than between 54% and 69% of the male sample, 95% of the time.

Hypothesis 3 is supported by the data presented in Tables 3 and 4. The Maccoby and Jacklin (1974) data produced a Z_{ma} of 0, whereas the weighted procedure with retrieved data produced the largest, though still nonsignificant, Z_{ma} of 1.10 ($p < .14$). Neither reporting precision nor method seems to influence this conclusion.

To summarize, then, the statistical combination of studies led to a conclusion different from the literary one when it showed an overall tendency for females to conform more than males. Increased precision in data gathering also weighed against the null hypothesis. When the studies were categorized, Maccoby and Jacklin's conclusion that sex differences would arise only in face-to-face interactions was upheld. However, the interpretation of this effect as inconsistent and exhibiting striking similarity between the sexes was questioned. Finally, it should be noted that, as one would expect, the two methods differ most in the "least clear" instance. Although in this instance the ability to categorize studies meant that the differing overall conclusions were substantively inconsequential, this may not always be the case.

The Eagly (1978) Update

Since the completion of the above meta-analysis, another review of sex differences in conformity research has appeared (Eagly, 1978). This literary review, based on considerably more studies than Maccoby and Jacklin's, also proves instructive when compared to meta-analytics.

Eagly provides a three-way classification of conformity studies nearly identical to the one employed above. With regard to Asch-type (group pressure) studies, the available evidence is summarized as follows: Of the 61 studies, 38 (62%) reported no difference, 21 (34%) found females to be significantly more conforming, and 2 (3%) found males significantly more conforming (Eagly, 1978, p. 92).

Eagly's criterion for significance is a main effect reaching the .05 chance level for null acceptance. A meta-analytic chi-square produced by Eagly's data has a value of 55.94.[1] Such a statistic is almost impossible to produce by chance (a chi-square value of 10.83 appears by chance once in 1,000 observations). To reduce the observed frequency of sex differences to the exact chance expectation, 361 studies *each* showing a null relation would have to exist but be undiscovered. About 221 undiscovered null studies would be needed to raise the expected number of confirmations close enough to the presently observed number so that chance could not be ruled out at the .05 level.

For fictitious group norm studies (not involving group pressure) Eagly reports 22 studies, with 2 reporting more conformity in women and 1 reporting more conformity in men. Eagly's conclusion that no difference is found here is justified through meta-analysis. Only 19 undiscovered null studies would reduce the present observations to an exact chance rate. For attitude change (persuasion) studies, 51 studies showing no difference, 10 showing greater female persuasibility, and 1

[1] In generating this statistic, I have used the expected values of 5 for significant differences (Rosenthal, 1978, suggests that no expectation less than 5 be used) and 54 for null differences. These numbers correspond to a .08 level of significance, implying that the reported chi-square value is conservative. I ignored the below-expectation (3%) occurrence of greater male conformity. A similar procedure is used throughout this discussion.

showing greater male persuasibility are reported. Eagly (1978) interprets these studies as indicating a very slight balance favoring female influenceability (p. 95). The associated meta-analytic chi-square statistic is 5.45, which is significant between the .02 and .01 levels. Undiscovered null studies numbering 139 would reduce the observed frequencies to a chance rate. About 59 null studies would reduce the observed patterns' relation to chance to above the .05 level.

In sum, Eagly (1978) states that "the modal finding in the conformity literature is no difference between the sexes" (p. 93) and that "in group pressure settings, a substantial minority of studies has reported differences in the female direction" (p. 93). These statements are correct but may be misinterpreted. They ought to be accompanied by a statement emphasizing that except in the case of extremely strong effects and/or large-sized experiments, we would *expect* the modal result to be a null confirmation. Further, it is evident that in the case of sex differences in group pressure conformity, the available "substantial minority" of null rejections is so substantial as to be almost impossible to create by chance.

Discussion

With the illustration in hand, it seems appropriate to conclude by addressing some of the more practical advantages of statistical combinations, in contrast to the abstract arguments presented earlier. Some limitations should also be pointed out.

First, the issue of inconsistency in experimental results deserves fuller examination. Other authors make statements whose eloquence cannot be improved upon here. As Saveggia (1974) offers:

> methodological principle overlooked by writers of research . . . reviews is that research results are *probabilistic*. What this principle suggests is that, in and of themselves, the findings of any single research are meaningless—they may have occurred simply by chance. It also follows that, if a large enough number of researches has been done on a particular topic, chance alone dictates that studies will exist that report inconsistent and contradictory findings! Thus, what appears to be contradictory may simply be the positive and negative details of a distribution of findings. (p. 397)

The disadvantages of maintaining our present use of the word *inconsistent* are listed by Light and Smith (1971):

> First, a great deal of information, much of which might be potentially valuable, would be thrown away. Second, a decision would be postponed for at least the length of time required by the new research. Third, from the point of view of the next reviewer of the literature, this new research would simply be the fourteenth in a set of studies. No matter what the result, for the next reviewer the contradictions remain. Thus, for any research, it is worth making an attempt to find a way to combine and reconcile conflicting studies. (p. 430)

The adoption of statistical procedures, then, will lead to less frequent conclusions of inconsistency. The use of the word *inconsistent* should occur only when experiments reveal strong relations in *opposite* directions.

When this stricter definition of inconsistency *is* applicable to a research area, a second advantage of meta-analysis will become evident. Often, inconsistency is an indication that undiscovered interaction between variables exists. Statistical combinations allow for the testing of certain kinds of interaction without running additional subjects. Experimental differences in time, location, operational definitions, age of subject, and so on can be used to cluster research (Feldman, 1971). Statistical procedures can then test the importance of cluster distinctions.

Clustering experiments will also lead to fuller awareness of sample restrictions relevant to the results in hand. In the present example, categorizing research led to uncovering what is a fairly typical generalization restriction in social psychological research (i.e., the results apply to white Americans, 21 years old or younger, in peer influence situations). These restrictions become possible mediators for future research to investigate. They are, however, a function of the available evidence and not of the combining method.

It should be emphasized that the ability to test certain interactions with meta-analysis does not mean that all problems of conceptualization and methodological artifact can be resolved in this manner. Whether con-

formities in Asch-type and in persuasive communication experiments are truly phenomena of a similar nature is not addressed by meta-analysis. Further, as in evaluating a single study, alternative conceptualization of included independent variables may rival the one offered. Thus, Festinger and Carlsmith's (1959) F statistic cannot tell us whether dissonance or self-perception is behind the "more attitude change for less reward" finding. The Z_{ma} statistic in a present day meta-analysis on replications of this study would not address this issue either. Finally, studies sharing a certain methodological technique become open to rival hypotheses (we hope, of course, that undiscovered method weaknesses are distributed throughout the research process and are equally likely to weigh against as for a given hypothesis). The study by Sistrunk and McDavid (1971) provides an example of restricting an inference by varying a method. They noted that one characteristic of Asch-type studies was a consistent use of perceptual dependent variables. Their results showed that face-to-face conformity differences may be restricted to this certain type of judgment only.

A final advantage of meta-analysis deals with the power of research to uncover relationships. As earlier stated, the probability of a study's finding significance is tied to the number of subjects involved. An example of this is present in Table 1. The experiments by Costanzo and Shaw (1966) and by Gerard, Wilhelmy, and Conolley (1968) uncovered effects of almost equal magnitude. Yet, because Gerard et al. ran 58 more subjects, their effect size proved significant. Meta-analyses should, as a matter of course, contain effect-size estimates, and these should be used to guide future research.

To conclude, this article has attempted to introduce and provide argument for the adoption of more rigorous procedures for drawing inferences from bodies of evidence in social psychology. An example was presented employing as hypotheses a literary interpretation of an area of research and using the exact same studies as data. This procedure made it possible to compare the two methods with conceptual issues controlled. Some statistical conclusions differed in direction and con dence from the literary ones.

Pleas for the increased use of archival da are often heard. Add one more voice, stati that the best archival data we have are o *own* records.

References

Allen, V. L., & Newston, D. Development of co formity and independence. *Journal of Personali and Social Psychology*, 1972, *22*, 18–30.

Asch, S. E. Studies of independence and conformity A minority of one against a unanimous majorit *Psychological Monographs*, 1956, *70*(9, Whole N 416).

Bakan, D. The test of significance in psychologic research. *Psychological Bulletin*, 1966, *66*, 423–43

Baumrind, D., & Black, A. E. Socialization practic associated with dimensions of competence in pr school boys and girls. *Child Development*, 1967, *3* 291–327.

Beloff, H. Two forms of social conformity: Acquies cence and conventionality. *Journal of Abnorma and Social Psychology*, 1958, *56*, 99–104.

Bishop, B. R., & Beckman, L. Developmental con formity. *Developmental Psychology*, 1971, *5*, 536

Bronfenbrenner, U. Reaction to social pressure from adults versus peers among Soviet day school and boarding school pupils in the perspective of an American sample. *Journal of Personality and Social Psychology*, 1970, *15*, 179–189.

Carrigan, W. C., & Julian, J. W. Sex and birth-order differences in conformity as a function of need affiliation arousal. *Journal of Personality and Social Psychology*, 1966, *3*, 479–483.

Cohen, J. *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Academic Press, 1977.

Cook, T. D., et al. Demand characteristics and three conceptions of the frequently deceived subject. *Journal of Personality and Social Psychology*, 1970, *14*, 185–194.

Costanzo, P. R., & Shaw, M. E. Conformity as a function of age level. *Child Development*, 1966, *37*, 967–975.

Dean, R. B., Austin, J. A., & Watts, W. A. Forewarning effects in persuasion. *Journal of Personality and Social Psychology*, 1971, *18*, 210–221.

Dillehay, R. C., & Jernigan, L. R. The biased questionnaire as an instrument of opinion change. *Journal of Personality and Social Psychology*, 1970, *15*, 144–150.

Dodge, N., & Muench, G. A. Relation of conformity and the need for approval in children. *Developmental Psychology*, 1969, *1*, 67–68.

Eagly, A. H. Sex differences in influenceability. *Psychological Bulletin*, 1978, *85*, 86–116.

Eagly, A. H., & Telaak, K. Width of the latitude of acceptance as a determinant of attitude change.

Journal of Personality and Social Psychology, 1972, *23,* 383–397.

Endler, N. S. Conformity as a function of different reinforcement schedules. *Journal of Personality and Social Psychology,* 1966, *4,* 175–180.

Endler, N. S., & Hoy, E. Conformity as related to reinforcement and social pressure. *Journal of Personality and Social Psychology,* 1967, *7,* 197–202.

Feldman, K. A. Using the work of others: Some observations on reviewing and integrating. *Sociology of Education,* 1971, *44,* 86–102.

Festinger, L., & Carlsmith, B. Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology,* 1959, *58,* 203–210.

Fisher, R. A. *Statistical methods for research workers* (7th ed.). London: Oliver & Boyd, 1938.

Frager, R. Conformity and anticonformity in Japan. *Journal of Personality and Social Psychology,* 1970, *15,* 203–210.

Friedman, H. Magnitude of experimental effect and a table for its rapid estimation. *Psychological Bulletin,* 1968, *70,* 245–251.

Gerard, H. B., Wilhelmy, R. A., & Conolley, E. S. Conformity and group size. *Journal of Personality and Social Psychology,* 1968, *8,* 79–82.

Glass, G. V. Primary, secondary and meta-analysis research. *Educational Researcher,* 1976, *5*(10), 3–8.

Gliniski, R. J., Gliniski, B. C., & Slatin, G. T. Nonnaivety contamination in conformity experiments: Sources, effects, and implications for control. *Journal of Personality and Social Psychology,* 1970, *16,* 478–485.

Greenbaum, C. W. Effect of situation and personality variables on improvisation and attitude change. *Journal of Personality and Social Psychology,* 1966, *4,* 260–269.

Hamm, N. H. A partial test of a social learning theory of children's conformity. *Journal of Experimental Child Psychology,* 1970, *9,* 29–42.

Hamm, N., & Hoving, K. L. Conformity of children in an ambiguous perceptual situation. *Child Development,* 1969, *40,* 773–784.

Hollander, E. P., Julian, J. W., & Haaland, G. A. Conformity process and prior group support. *Journal of Personality and Social Psychology,* 1965, *2,* 852–858.

Insko, C. A. Verbal reinforcement of attitude. *Journal of Personality and Social Psychology,* 1965, *2,* 621–623.

Insko, C. A., & Cialdini, R. B. A test of three interpretations of attitudinal verbal reinforcement. *Journal of Personality and Social Psychology,* 1969, *12,* 333–341.

Julian, J. W., Regula, C. R., & Hollander, E. P. Effects of prior agreement by others on task confidence and conformity. *Journal of Personality and Social Psychology,* 1968, *9,* 171–178.

Landsbaum, J. B., & Willis, R. H. Conformity in early and late adolescence. *Developmental Psychology,* 1971, *4,* 334–337.

LeFurgy, W. G., & Woloshin, G. W. Immediate and long-term effects of experimentally induced social influence in the modification of adolescents' moral judgments. *Journal of Personality and Social Psychology,* 1969, *12,* 104–110.

Levy, P. Substantive significance of significant differences between two groups. *Psychological Bulletin,* 1967, *67,* 37–40.

Light, R. J., & Smith, P. V. Accumulating evidence: Procedures for resolving contradictions among different studies. *Harvard Educational Review,* 1971, *41,* 429–471.

Linder, D. E., Cooper, J., & Jones, E. E. Decision freedom as a determinant of the role of incentive magnitude in attitude change. *Journal of Personality and Social Psychology,* 1967, *6,* 245–254.

Maccoby, E. E., & Jacklin, C. N. *The psychology of sex differences.* Stanford, Calif.: Stanford University Press, 1974.

Marquis, P. C. Experimenter–subject interaction as a function of authoritarianism and response set. *Journal of Personality and Social Psychology,* 1973, *25,* 289–296.

McGuire, W. J. The yin and yang of progress in social psychology: Seven Koan. *Journal of Personality and Social Psychology,* 1973, *26,* 446–456.

Mock, R. L., & Toddenham, R. D. Race and conformity among children. *Developmental Psychology,* 1971, *4,* 349–365.

Mosteller, F. M., & Bush, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology: Vol. 1. Theory and method.* Cambridge, Mass.: Addison-Wesley, 1954.

Nisbett, R. E., & Gordon, A. Self-esteem and susceptibility to social influence. *Journal of Personality and Social Psychology,* 1967, *5,* 268–276.

O'Brien, T. C., & Shapiro, B. J. Statistical significance—What? *Mathematics Teacher,* 1968, *61,* 673–676.

Osterhouse, R. A., & Brock, T. C. Distraction increases yielding to propaganda by inhibiting counterarguing. *Journal of Personality and Social Psychology,* 1970, *15,* 344–358.

Pearson, E. S. The probability integral transformation for testing goodness of fit and combining tests of significance. *Biometrika,* 1938, *30,* 134–148.

Rosenkrantz, P. A., & Crockett, W. H. Some factors influencing the assimilation of disparate information in impression formation. *Journal of Personality and Social Psychology,* 1965, *2,* 397–402.

Rosenthal, R. *Experimenter effects in behavioral research* (2nd ed.). New York: Irvington, 1976.

Rosenthal, R. Combining results of independent studies. *Psychological Bulletin,* 1978, *85,* 185–193.

Rosenthal, R. The "file drawer problem" and tolerance for null results. *Psychological Bulletin,* in press.

Rule, B. G., & Rehill, C. Distraction and self-esteem effects on attitude change. *Journal of Personality and Social Psychology,* 1970, *15,* 359–365.

Samorajczyk, J. Children's responsiveness to imagination suggestions during school entry. *Developmental Psychology,* 1969, *1,* 211–215.

Sampson, E. E., & Hancock, F. T. An examination of

the relationship between ordinal position, personality, and conformity. *Journal of Personality and Social Psychology,* 1967, *5,* 398–407.

Schneider, F. W. Conforming behavior of black and white children. *Journal of Personality and Social Psychology,* 1970, *16,* 466–471.

Silverman, I. Role-related behavior of subjects in laboratory studies of attitude change. *Journal of Personality and Social Psychology,* 1968, *8,* 343–348.

Silverman, I., Shulman, A. D., & Wiesenthal, D. L. Effects of deceiving and debriefing psychological subjects on performance in later experiments. *Journal of Personality and Social Psychology,* 1970, *14,* 203–212.

Sistrunk, F. Negro–white comparisons in social conformity. *Journal of Social Psychology,* 1971, *85,* 77–85.

Sistrunk, F., & McDavid, J. W. Sex variable in conforming behavior. *Journal of Personality and Social Psychology,* 1971, *17,* 200–207.

Smith, M. L., & Glass, G. V. Meta-analysis of psychotherapy outcome studies. *American Psychologist,* 1977, *32,* 752–760.

Stickell, D. W. *A critical review of the methodology and results of research comparing televised and face-to-face instruction.* Unpublished doctoral dissertation, Pennsylvania State University, 1963.

Stouffer, S. A. *The American soldier: Vol. 1. Adjustment during army life.* Princeton, N.J.: Princeton University Press, 1949.

Stricker, L. J., Messick, S., & Jackson, D. N. Suspicion of deception: Implications for conformity research. *Journal of Personality and Social Psychology,* 1967, *5,* 379–389.

Taveggia, T. C. Resolving research controversy through empirical cumulation. *Sociological Methods and Research,* 1974, *2,* 395–407.

Whiting, B. B. (Ed.). *Six cultures: Studies in child rearing.* New York: Wiley, 1963.

Willis, R. H., & Willis, Y. A. Role playing versus deception: An experimental comparison. *Journal of Personality and Social Psychology,* 1970, *16,* 472–477.

Worschel, S., & Brehm, J. W. Effect of threat to attitudinal freedom as a function of agreement with the communicator. *Journal of Personality and Social Psychology,* 1970, *14,* 18–22.

Wyer, R. S., Jr. Effects of incentive to perform well, group attraction, and group acceptance on conformity in a judgmental task. *Journal of Personality and Social Psychology,* 1966, *4,* 21–26.

Wyer, R. S., Jr. Effects of task reinforcement, social reinforcement, and task difficulty on perseverance in achievement-related activity. *Journal of Personality and Social Psychology,* 1968, *8,* 269–276.

Received March 6, 1978 ■

Ad Hoc Reviewers

The Editor acknowledges with thanks the advice and consultation provided by the following persons, who served recently as ad hoc reviewers of manuscripts. (The number of manuscripts reviewed is in parentheses, if more than one.)

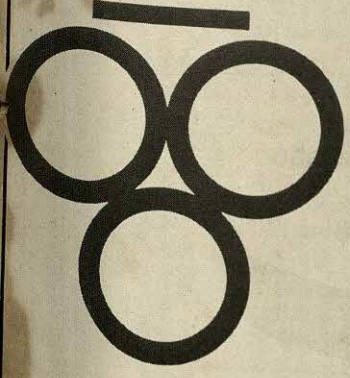
Vernon L. Allen
Robert Arkin
Paul Bakan
Albert Bandura
Andrew Baum
Michael Berbaum
Leonard Bickman
Ross W. Buck
E. Jocelyn Calvert-Boyanowsky
Jack Citrin
Simon Fraser
Daniel Geller
Kenneth Gergen
Gerald Greenberg
Ranald D. Hansen
Richard Harris (2)
Richard Izzett
Stuart A. Karabenick
Norbert Kerr
Henry Law
Melvin Lerner
Margaret Matlin
Edwin Megargee
Thomas Murray
Richard Nisbett
Milton Rokeach
Ralph Rosnow
Michael Saks
Edward E. Sampson
Paul Slovic
David Shaffer
Abraham Tesser
Joseph Veroff
Amiram Vinokur
Gary L. Wells
Charles E. Werts
David Wiley
Kip Williams

APA EMPLOYMENT BULLETIN

Possibly your key to greater career opportunities.

Each month 125 to 150 position openings are listed by geographical area and employment field for psychologists with a minimum of a Master's degree. Members may also publish availability notices. Coded identification numbers assure confidentiality. Ideal for employers, students, and Master's and Ph.D graduates.

A one year subscription is as little as $9.00 for members, $16.00 for non-members. If you're in a hurry, First Class delivery is available for an additional $4.00.



For more information, contact:
The Editor, Employment Bulletin,
or send your order to:
American Psychological Association
Subscription Dept.-Employment Bulletin
1200 Seventeenth St., N.W.
Washington, D.C. 20036

Methodology in Clinical Research

Special issue of the **Journal of Consulting and Clinical Psychology** *August 1978*

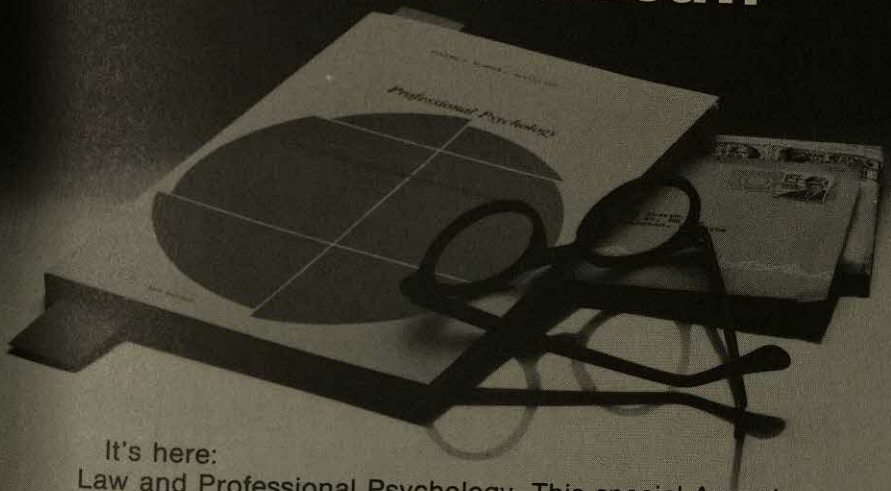
Major methodological aspects of clinical topic areas are described and discussed. Topic areas include smoking and addiction, marital and child treatment, sex roles, obesity, and others.

Contributors include, the editor, Brendan A. Maher, Alan E. Kazdin, Richard M. McFall, Peter E. Nathan, David Lansky, and Judith Worell to name a few.

Copies of the special issue are available at $6 each, prepaid. Discounts on bulk orders are also available. To order single copies or for more information on bulk orders write:

Subscription Department
American Psychological Association
1200 17th Street, N.W.
Washington, D.C. 20036

NOW: INFORMATION ABOUT THE LAW AND PSYCHOLOGY.



It's here:
Law and Professional Psychology. This special August 1978 issue of **Professional Psychology** brings you thirteen original articles full of fresh thought and constructive recommendations on the issues that vitally affect your day-to-day practice. You'll find precautions you may take to protect yourself as a professional and a plan for improving investigations of malpractice claims. Expert witness testimony, jury selection, civil commitment, confidentiality and privilege, minor's consent to treatment, and the use of psychological devices are critically examined.

Send your order today. And soon you'll receive this special issue.

Single copies of the issue are available for $5 each. All orders $25 or less must be prepaid. Make and send checks payable to:

American Psychological Association
Order Department
1200 17th Street, N.W.
Washington, D.C. 20036

Enclosed is $_____ for ___ copies of **Professional Psychology's** special issue—Law and Professional Psychology—at $5.00 each.

NAME _____

ADDRESS _____

CITY _____

STATE _____ ZIP CODE _____

PP-8

Student-involving introductory texts

NEW

EXPERIENCING SOCIAL PSYCHOLOGY: Readings and Projects

Ayala Pines and **Christina Maslach,** both of the University of California, Berkeley

This integrated reader/project manual features research articles and popular readings on current topics, classical problems, and new content areas such as social pathology, ethical dilemmas of researchers, and the problems of "burn-out" among helping professionals. The introduction shows students how to read such material critically, while thoroughly class-tested projects make this material concrete by involving students in a variety of independent activities, including data collection, theoretical critiques, and group exercises.

Random House/1979/320 pages paperbound/Order Codes: 31281, 32168 (I.M.)

PERSONALITY: An Introduction

John Lamberth, Herbert Rappaport, and **Margaret Rappaport,** all of Temple University

Each chapter of this comprehensive text was written by the author with the greatest expertise and interest in the area, making the whole text completely authoritative. Every major aspect of personality study — assessment techniques, development, research, and applications — is fully covered and viewed in the light of all theoretical approaches, giving you the most comprehensive, balanced, up-to-date presentation possible.

Alfred A. Knopf/1978/544 pages hardbound/Order Codes: 31190, 32167 (I.M.)

For more information, write to:

Random House/Alfred A. Knopf

College Review Desk J-59, 400 Hahn Road, Westminster, Md. 21157

JOURNAL OF

Personality AND Social Psychology

VOLUME 37 NUMBER 2 · FEBRUARY 1979

Self-Deception: A Concept in Search of a Phenomenon 147
Ruben C. Gur and Harold A. Sackeim

The concept of self-deception is operationalized, and the existence of self-deception is experimentally demonstrated.

Television Viewing and Fear of Victimization: Is the Relationship Causal? 170
Anthony N. Doob and Glenn E. Macdonald

Using regression and canonical correlation analyses, it was shown that television does not cause fear of criminal victimization.

Retrieval Selectivity in Memory-Based Impression Judgments 180
John H. Lingle and Thomas M. Ostrom

Personality Correlates of Susceptibility to Biasing Information 195
Elisha Y. Babad

Prediction and Perception of Social Motives 203
Judith E. Maki, Warren B. Thorngate, and Charles G. McClintock

Effects of External Evaluation on Artistic Creativity 221
Teresa M. Amabile

Menstrual Cycle Affects Kinesthetic Aftereffect, An Index of Personality and Perceptual Style 234
A. Harvey Baker, Brian L. Mishara, Irene W. Kostin, and Laurence Parker

Evaluative Judgments of Aspects of Life as a Function of Vicarious Exposure to Hedonic Extremes 247
Marshall Dermer, Sidney J. Cohen, Elaine Jacobsen, and Erling A. Anderson

Actor–Observer Differences in Information Inference and Causal Attribution 261
Susan V. Elsen

Sex Differences in Eavesdropping on Nonverbal Cues 273
Robert Rosenthal and Bella M. DePaulo

Young Children's Preferred Attentional Strategies for Delaying Gratification 286
Brian T. Yates and Walter Mischel

Strategy Signals in Face-to-Face Interaction 301
Starkey Duncan, Jr., Lawrence J. Brunner, and Donald W. Fiske

Ad Hoc Reviewers 300

Corrections to Lynch and Cohen 233

Editors Named for New *Journal of Personality and Social Psychology* 314

Notice on Author Alterations 194

Published monthly by the

AMERICAN PSYCHOLOGICAL ASSOCIATION

ISSN 0022-3514

Editor

Anthony G. Greenwald, *Ohio State University*

Topic Editors

Irwin Altman, *University of Utah*
Robert A. Baron, *Purdue University*
Daryl Bem, *Stanford University*
Eugene Burnstein, *University of Michigan*
Robert B. Cialdini, *Arizona State University*
Reid Hastie, *Harvard University*
Douglas N. Jackson, *University of Western Ontario*
Leslie Z. McArthur, *Brandeis University*

Consulting Editors

Icek Ajzen, *University of Massachusetts—Amherst*
Robert S. Baron, *University of Iowa*
Gordon Bear, *Ramapo College of New Jersey*
Ellen Berscheid, *University of Minnesota*
Michael Birnbaum, *University of Illinois at Urbana-Champaign*
Philip Brickman, *Institute for Social Research*
J. Merrill Carlsmith, *Stanford University*
John S. Carroll, *Loyola University of Chicago*
Jerry L. Cohen, *University of Illinois at Urbana-Champaign*
Sheldon Cohen, *University of Oregon*
Rick Crandall, *Texas Christian University*
William D. Crano, *Michigan State University*
Robyn M. Dawes, *University of Oregon*
Kay K. Deaux, *Purdue University*
Marshall Dermer, *University of Wisconsin—Milwaukee*
Edward Diener, *University of Illinois at Urbana-Champaign*
Anthony Doob, *University of Toronto, Toronto, Canada*
Susan V. Eisen, *McLean Hospital, Belmont, Massachusetts*
Baruch Fischhoff, *Decision Research, Eugene, Oregon*
Irene H. Frieze, *University of Pittsburgh*
David Funder, *Stanford University*
David L. Hamilton, *University of California, Santa Barbara*
Ranald D. Hansen, *Oakland University*
Richard J. Harris, *University of New Mexico*
Edward Helmes, *University of Western Ontario, London, Canada*
Donald S. Hiroto, *VA Hospital, Brentwood, Los Angeles, CA*
Edward E. Jones, *Princeton University*
Martin F. Kaplan, *Northern Illinois University*
Douglas T. Kenrick, *Montana State University*
Norbert L. Kerr, *University of California, San Diego*
Samuel Komorita, *University of Illinois at Urbana-Champaign*
Vladimir J. Konečni, *University of California, San Diego*
Robert M. Krauss, *Columbia University*
Ellen J. Langer, *Harvard University*
Mark R. Lepper, *Stanford University*
Melvin Manis, *University of Michigan (Ann Arbor)*
Hazel Markus, *University of Michigan (Ann Arbor)*
Margaret Matlin, *SUNY College at Geneseo*
Dale T. Miller, *University of British Columbia, Vancouver*
Thomas M. Ostrom, *Ohio State University*
Paul B. Paulus, *University of Texas at Arlington*
Richard Petty, *University of Missouri—Columbia*
Harry T. Reis, *University of Rochester*
Judith Rodin, *Yale University*
Seymour Rosenberg, *Livingston College, Rutgers*
Ralph L. Rosnow, *Temple University*
Lee Ross, *Stanford University*
Michael Ross, *University of Waterloo, Waterloo, Canada*
Diane Ruble, *Princeton University*
Michael J. Saks, *Nat'l Center for State Courts, Williamsburg, VA*
Michael F. Scheier, *Carnegie-Mellon University*
Barry R. Schlenker, *University of Florida*
Philip Shaver, *New York University*
Steven J. Sherman, *Indiana University*
Jerome E. Singer, *University of the Health Services, Bethesda, Maryland*
Eliot R. Smith, *University of California, Riverside*
Mark Snyder, *University of Minnesota*
Melvin L. Snyder, *University of Texas at Austin*
Lawrence Stricker, *Educational Testing Service*
Shelley E. Taylor, *Harvard University*
Abraham Tesser, *University of Georgia*
Bernard Weiner, *University of California, Los Angeles*
Gary L. Wells, *University of Alberta, Edmonton, Canada*
Stephen Worchel, *University of Virginia*
Robert S. Wyer, *University of Illinois at Urbana-Champaign*
Mark Zanna, *University of Waterloo, Waterloo, Canada*
Miron Zuckerman, *University of Rochester*

Assistant to the Editor
Holly Wong

APA Journal Staff

Anita DeVivo, *Executive Editor*
Ann I. Mahoney, *Manager, Journal Production*
Barbara R. Richman, *Production Supervisor*
Marjorie M. Henry, *Production Editor*
Robert J. Hayward, *Advertising Representative*
Juanita Brodie, *Subscription Manager*

The Journal of Personality and Social Psychology publishes original reports of research, methodology, theory, criticism, and literature review in the fields of personality and social psychology of normal processes and populations. Submitted manuscripts should delineate an important question and provide a clear answer to it. In the case of research reports, this will often require demonstration that the reported results are generalizable beyond a single research setting. Consult the editorial in the January 1976 issue (Vol. 33, pp. 1–7) for a detailed statement of policies and instructions to authors.

Manuscripts: Submit manuscripts in quadruplicate, prepared according to the *Publication Manual of the American Psychological Association* (2nd ed.). Copies should be highly legible and will not be returned to authors unless they are original typewritten copies. Mail manuscripts to the Acting Editor, Clyde Hendrick, Department of Psychology, Box 8185, University of Miami, Coral Gables, FL 33124 until March 31, 1979. After April 1 mail manuscripts according to instructions on page 314.

Required supplementary information: Manuscripts cannot be reviewed for publication unless the author provides supplementary information in the following categories: (a) a cover letter that indicates the relation (or the lack of relation) of the submitted manuscript to other papers by the same author(s) that are currently submitted, in press, or recently published; (b) one copy of each such related paper; (c) one copy of a statement of the procedures of the research reported in the submitted manuscript in sufficient detail to serve as a guide for conduct of an exact replication; and (d) one copy of the summary tables (e.g., correlation tables, analysis of variance summaries) for the *major* analyses reported. If it appears that the supplementary materials in categories (c) or (d) might be bulky, the author should exercise judgment and restraint in selecting what to send. It is not necessary to send supplementary material that repeats verbatim manuscript content.

After a manuscript is accepted, authors will be expected to provide assurance that their data are available in unanalyzed form for reanalyses by other investigators and that their research procedures were consistent with the American Psychological Association's published principles of research ethics, except insofar as may be detailed otherwise in the manuscript.

Copyright and Permission: All rights reserved. Written permission must be obtained from the American Psychological Association for copying or reprinting text of more than 500 words, tables, or figures. Permission is normally granted contingent upon like permission of the author, inclusion of the APA copyright notice on the first page of reproduced material, and payment of a fee of $10 per page, table, or figure. Abstracting is permitted with credit to the source. Permissions and fees are waived for authors who wish to reproduce their own material for personal use; fees only are waived for authors who wish to use their own material commercially. Permission and fees are waived for the photocopying of isolated articles for nonprofit classroom or library reserve use by instructors and educational institutions. Libraries are permitted to photocopy beyond the limits of U.S. copyright law: (1) those post-1977 articles with a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through the Copyright Clearance Center, P.O. Box 765, Schenectady, NY 12301; (2) pre-1978 articles, provided that the per-copy fee stated in the Publishers' Fee List is paid through the Copyright Clearance Center, P.O. Box 765, Schenectady, NY 12301. Address requests for reprint permission to the Permissions Office, APA, 1200 Seventeenth Street, N.W., Washington, D.C. 20036.

Subscriptions: Subscriptions are available on a calendar year basis only (January through December). Nonmember rates for 1979: $60 domestic, $64 foreign, $6 single issue. APA member rate: $20. Write to Subscription Section, APA.

Change of Address: Send change of address notice and a recent mailing label to the attention of the Subscription Section, APA, 30 days prior to the actual change of address. APA will not replace undelivered copies resulting from address changes; journals will be forwarded only if subscribers notify the local post office in writing that they will guarantee second-class forwarding postage.

Reprints: Authors may order reprints of their articles from the printer when they receive proofs.

Back Issues and Back Volumes: For information regarding back issues or back volumes write to Order Department, APA.

Microform Editions: For information regarding microform editions write to any of the following: Johnson Associates, Inc., P.O. Box 1017, Greenwich, Connecticut 06830; University Microfilms, Ann Arbor, Michigan 48106; or Princeton Microfilms, Princeton, New Jersey 08540.

Published monthly in one volume per year by the American Psychological Association, Inc., 1200 Seventeenth Street, N.W., Washington, D.C. 20036. Printed in the U.S.A. Second-class postage paid at Washington, D.C., and at additional mailing offices.

JOURNAL OF
Personality
AND
Social Psychology

Volume 37 February 1979 Number 2

Self-Deception: A Concept in Search of a Phenomenon

Ruben C. Gur
University of Pennsylvania

Harold A. Sackeim
Columbia University

The concept of self-deception has received little empirical investigation despite its broad implications for theories of personality and consciousness. Four criteria, based on a logico-linguistic analysis, are presented as necessary and sufficient for ascribing self-deception: To be self-deceived an individual must hold two contradictory beliefs; these beliefs are held simultaneously; one belief is not subject to awareness; and the nonawareness of this belief is motivated. Two experiments are described that examine whether misidentifications of the voices of oneself and others fulfill these criteria and are, therefore, instances of self-deception. In Experiment 1, it was found that when subjects were incorrect in their self-report identifications of voices, at some level of processing correct identifications had been made, and these subjects held contradictory beliefs. Obtrusive and unobtrusive measures indicated that for the most part, subjects were not aware of committing errors. Correlational data supported the contention that the errors were motivated. In Experiment 2, the hypothesized motivational contexts for the errors were manipulated by giving subjects pretreatments either of success or of failure. As predicted, the failure group committed more misidentifications of self, whereas the success group committed more misidentifications of others. Overall, the findings confirmed that some misidentifications of voices of self and others are instances of self-deception.

The concept of self-deception has received much attention in philosophical and literary writings (e.g., Camus, 1956; Fingarette, 1969; Gide, 1955; Sartre, 1958). Numerous psychologists have also discussed self-deception in accounting for a wide range of behavior. Meehl and Hathaway (1946) and Anastasi (1961) have argued that self-deception on the part of respondents contributes more to the lack of validity of self-report personality inventories than does other-deception, or conscious lying. In the area of motivation and personality, Hilgard (1949) claimed that self-deception constitutes a defining characteristic of all defense mechanisms. In the area of problem-solving behavior, Wason and Johnson-Laird (1972) in-

This research was supported by National Science Foundation Grant BNS-75-23061 to Jerre Levy, Ruben C. Gur, and Raquel E. Gur and by Alcohol, Drug Abuse, and Mental Health Administration National Research Service Award 1F31 MH 05779-01 from the National Institute of Mental Health to Harold A. Sackeim.

We thank Solomon Asch, Jonathan Baron, Donald D. Dorfman, Jonquil M. Drinkwater, Matthew H. Erdelyi, Henry Gleitman, Raquel E. Gur, Lauren J. Harris, Ernest R. Hilgard, Leonard Horowitz,

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3702-0147$00.75

Diary No. 66
Date 2/11/1979

voked the concept of self-deception as a possible explanation of the persistence of subjects in maintaining hypotheses in the face of disconfirmation. Murphy (1970, 1975) related the concept of self-deception to possible interpretations of the experimental findings on perceptual defense. In fact, Mischel (1974) has viewed all neurotic behavior as instances of self-deceptive acts.

Despite the considerable role assigned to the concept of self-deception in several areas of study, there have been no attempts to demonstrate that any given set of behavior conforms to what is meant by self-deception. Philosophers have frequently remarked that determining what is meant by self-deception and demonstrating that people do deceive themselves have major implications for views concerning the structure of consciousness (cf. Fingarette, 1969; Sartre, 1958). For instance, findings that indicate that people do lie to themselves may also indicate that people can hold beliefs of which they are not aware and that the selective awareness of beliefs can be motivated. Within psychology it has been a long-standing tradition to assume that people are necessarily aware of their cognitions (Wundt, 1912/1973). Recently, however, studies have appeared that shed strong doubt on the validity of this view (Dixon, 1971; Erdelyi, 1974; Nisbett & Bellows, 1977; Nisbett & Wilson, 1977; Sackeim, Packer, & Gur, 1977). The claim that selective nonawareness of cognition can be motivated—the essence of what is entailed by the concept of self-deception—is particularly controversial.

Elsewhere, we have presented a logico-linguistic analysis of what is meant by the concept of self-deception (Sackeim & Gur, 1978). On the basis of this analysis, four criteria have been offered as necessary and sufficient for the ascription of self-deception. Here, we will present two experiments designed to test whether a phenomenon, misidentification of voices of self or others, fits these criteria. Such findings may be viewed as an experimental demonstration of the existence of self-deception.

Criteria Necessary and Sufficient for Ascribing Self-Deception

It has been noted that when it is assumed that people are necessarily aware of their cognition, the concept of self-deception is paradoxical (Canfield & Gustafson, 1962; Demos, 1960; Fingarette, 1969; Gardiner, 1970; Penelhum, 1966; Sartre, 1958; Siegler, 1962). This paradox has been formulated by Sartre (1958) as follows:

> The one to whom the lie is told and the one who lies are one and the same person, which means that I must know in my capacity as deceiver the truth which is hidden from me in my capacity as the one deceived. Better yet I must know the truth very exactly in order to conceal it more carefully—and this not at two different moments, which at a pinch would allow us to reestablish a semblance of duality—but in the unitary structure of a single project. How then can the lie subsist if the duality which conditions it is suppressed? (p. 49)

In the psychological literature, a similar paradox was raised by critics of research on subliminal perception and perceptual defense (Bruner & Postman, 1949; Eriksen & Browne, 1956; Howie, 1952). The point was made that in order for a perceiver to avoid perceiving a stimulus, the stimulus must first be perceived. When it is assumed that perception implies awareness of the percept, notions like subliminal perception and perceptual defense are paradoxical. Proponents of the existence of these phenomena (Dixon, 1971; Erdelyi, 1974) have argued that it is erroneous to assume that cognition must be subject to awareness. In support of this posi-

Jerre Levy, Jacob Nachmias, M. Frank Norman, Daniel N. Osherson, Michael Parrish, Richard Passingham, Paul Rozin, Martin E. P. Seligman, Richard L. Solomon, David Williams, Julius Wishner, and Donna M. Zucchi for their comments on an earlier version of the manuscript and Anthony G. Greenwald and several anonymous reviewers for their comments on the present version. We also thank Joseph Liran, who aided in setting up the instrumentation, and Jo Ann Chalal, Harvey Doppelt, Ron Kamnik, Lisa Litvin, Alan Luxenberg, Joseph Piperato, Shelley Steinberg, and Marjory Warner, who served as experimenters.

The order of authors is random.

Requests for reprints should be sent to Ruben C. Gur, Department of Psychology, University of Pennsylvania, 3815 Walnut Street, Philadelphia, Pennsylvania 19104.

tion, Nisbett and Wilson (1977) have summarized a sizable body of literature that indicates that people may lack awareness for both the contents and processes involved in cognition.

Rejection of the assumption that cognition is necessarily subject to awareness is implicit in the common use of the term *self-deception.* Often when people describe an individual as self-deceived, they state that "all along" or "deep inside" the individual holds a belief that contradicts an avowed belief. It is further implied in the common use of the term that the self-deceived individual, who holds such contradictory beliefs, does not do so gratuitously or out of ignorance; rather, one belief is not subject to awareness, in order to provide some psychological gain. Just as people attribute motivational determinants in cases of conscious lying, they view instances of self-deceit as implicating motivation. Accordingly, we have offered the following criteria as necessary and sufficient for ascribing self-deception to any given phenomenon:

1. The individual holds two contradictory beliefs (that p and not that p).
2. These two contradictory beliefs are held simultaneously.
3. The individual is not aware of holding one of the beliefs.
4. The act that determines which belief is and which belief is not subject to awareness is a motivated act.

Self-Deception: Reactions to Self-Confrontation

In the search for a phenomenon that fits the concept of self-deception, the self-confrontation experience provides a likely candidate. It has been shown that when subjects are confronted with audio- or videotape feedback of the self and others, feedback of the self is associated with considerable psychophysiological reactivity (Holzman, Rousey, & Snyder, 1966; Olivos, 1967; Sackeim & Gur, 1978), changes in affect and constriction in ideation (Duval & Wicklund, 1972; Holzman & Rousey, 1966; Holzman et al., 1966; Rousey & Holzman, 1967), and changes in self-concept (Alkire & Brunse, 1974; Boyd & Sisney, 1967; Duval & Wicklund, 1972; Gur & Sackeim, in press; Storms, 1973). Self-confrontation manipulations have also been shown to influence dream content (Castaldo & Holzman, 1967, 1969) and other behaviors that do not appear to be mediated by conscious awareness (Huntley, 1940; Wolff, 1943).

Differences in the direction and intensity of reactions to self-confrontation have been associated with experimental manipulations of self-esteem and with individual differences in personality (Davis & Brock, 1975; Duval & Wicklund, 1972; Gibbons & Wicklund, 1976; see Sackeim & Gur, 1978, for a review). These findings have led to the conclusion that individuals who hold negative attitudes about the self and score high on measures of cognitive discrepancy[1] find self-confrontation to be aversive and will tend to avoid it. On the other hand, people low in cognitive discrepancy do not find self-confrontation aversive and in fact will seek it out. Several studies have shown that people often fail to recognize their own voice, and sometimes people identify others as the self (Holzman et al., 1966; Huntley, 1940; Olivos, 1967; Wolff, 1943). The fact that there are individual differences in reactions to self-confrontation, and that some people find self-confrontation aversive while others seek it out, may indicate a motivational basis for the errors in identification.

The question we posed for investigation is whether individuals are engaging in self-deception when they avoid self-confrontation by misidentifying the self as others and when they artificially seek out self-confrontation

[1] Cognitive discrepancy is defined here as the extent to which individuals hold discrepant attitudes and beliefs about themselves. These discrepant cognitions may involve conflicts between what individuals believe themselves to be and what they believe they should be or wish to be. On paper-and-pencil personality inventories, items that a priori would tap this dimension are questions concerning dissatisfaction with the self. For example, on the Neuroticism scale of the Eysenck Personality Inventory (Eysenck & Eysenck, 1963), such an item would be "Do you often worry about things you should not have done or said?"

by misidentifying others as the self. Both of these errors involve distortions of reality. Mistaking self for others reflects an avoidance or a denial of an aversive event. Individuals who misidentify others as self commit a "narcissistic," self-projecting response. Our hypothesis, then, is that at least some misidentifications of self or others are instances of self-deception.

Experiment 1

Operationalizing the Criteria for Self-Deception

In order to demonstrate that misidentifications of self and others are instances of self-deception, it must be shown that when people commit such errors they fulfill the four criteria for ascribing self-deception. The operationalization of the criteria is presently outlined in specific relation to the procedure used in Experiment 1. Subjects were administered a simple identification task in which they were asked to indicate whether audio stimuli were tapes of their own or other people's voices. Galvanic skin responses (GSR) to each voice and reaction time of identifications were recorded. Subjects had completed a battery of personality inventories before the experimental session.

The First Criterion: Holding Two Contradictory Beliefs

The first criterion for ascribing self-deception requires evidence that when subjects misidentify a voice they hold contradictory beliefs. Subjects' self-report identifications of voices provide one index of beliefs. In affirming that a given voice is, for instance, that of another, a statement has been made concerning a particular belief. Levels of GSR reactivity to voices may provide another index of beliefs. In every relevant study that examined psychophysiological reactivity, it has been found that levels of arousal are substantially higher following self-confrontation as compared to confrontation with others (Holzman et al., 1966; Murray, 1963; Olivos, 1967; Sackeim & Gur, 1978; Verwoerdt, Nowlin, & Agnello, 1965; Dickinson & Ray, Note 1; Sackeim, Note 2). If subjects hold contradictory beliefs when they incorrectly identify voices, then their levels of GSR reactivity should be high when they misidentify voices of self as others (false negative responses) and low when they misidentify voices of others as self (false positive responses). This would indicate that when subjects committed such errors, at some level of processing correct identifications had also been made. This use of psychophysiological responding to determine whether subjects hold contradictory beliefs is similar in concept to the use of psychophysiological measurement in the detection of conscious lying (Lykken, 1959, 1974; Orne, Thackray, & Paskewitz, 1972) and to the employment of such measures in subception experiments (Dixon, 1971; Eriksen, 1956, 1958; Lazarus & McCleary, 1951).

The Second Criterion: Simultaneity of Beliefs

The second criterion for the ascription of self-deception requires that the two contradictory beliefs be held simultaneously. In the present experimental context this criterion is fulfilled by examining the psychophysiological index of beliefs at the same time that self-report identifications are delivered.

The Third Criterion: Nonawareness of a Belief

Fulfillment of the third criterion requires evidence that one of the two contradictory beliefs is not subject to awareness and in the present investigation requires demonstration that subjects are not aware of misidentifying voices. Showing that subjects are unaware of stimulus conditions, cognitions, and/or behavior has been particularly difficult in several areas of psychology except through the use of highly obtrusive measures (Martin, Hawryluk, & Guse, 1974; Nisbett & Wilson, 1977). Martin et al. (1974) wrote, "The most pressing methodological problem with the present studies is the one that plagues all studies of learning without awareness. There seems to be no entirely defensible method of assessing subject awareness after the experiment" (p. 604). To overcome this problem in the present study, two measures

of awareness of errors were taken. The first measure relies on postexperimental self-reports of whether subjects are aware of having committed errors. This measure is subject to the criticisms mentioned by Martin et al. (1974). The second measure, however, provides for this context a possible solution to the methodological problem, an unobtrusive index of awareness.

The unobtrusive index of awareness rests on the claim that consciously holding a belief that a voice is that of the self leads to differential consequences in subsequent responding than not being aware of such a belief. When subjects either correctly or incorrectly identify a voice as the self, their subsequent identification of the voice of another will be faster than their responses on other trials of voices of others. In the present study, trials of voices are arranged so that every presentation of the voice of self is followed by presentation of the voice of a stranger. If individuals correctly identify the self, awareness of this fact should make the next identification of the voice of another faster than responses to voices of another not preceded by identifications of self. This is the case because any difference in the physical properties of the voice consciously believed to be the self and those of the subsequent voice will aid in identifying the latter voice as that of another. If subjects are unaware that a given voice is the self, that is, if they commit a false negative error, they should not be faster in identifying the following voice than they are on remaining trials of voices of others. By the same argument, if subjects are not aware of committing false positive errors, that is of misidentifying others as self, they should be faster in identifying voices that immediately follow the false positive errors than in identifying voices that are not preceded by such errors. Comparisons of reaction times on trials following particular response types can thus provide an unobtrusive index of awareness.

The Fourth Criterion: Motivational Determinants

To fulfill the fourth criterion, it must be shown that misidentifications of self and others are motivated. This criterion is the most difficult to establish, and Experiment 1 is viewed as an initial step in this direction.

Previous studies have found that manipulations of self-esteem influence the aversiveness of self-confrontation (Davis & Brock, 1975; Duval & Wicklund, 1972; Gibbons & Wicklund, 1976). Likewise, it has been shown that measures of personality and psychopathology, particularly assessments of discrepant cognitions about the self, predict levels of psychophysiological reactivity during self-confrontation and changes in affect following the confrontation (Sackeim, Note 2). Given the hypothesis that subjects commit false negative errors in order to avoid aversive consequences of self-confrontation, it might be expected that subjects who commit such errors would score higher on measures of cognitive discrepancy than subjects who do not make these errors. However, this prediction can not be made because of the very fact that it is hypothesized that these subjects are engaging in self-deception. In line with Meehl and Hathaway (1946), we have argued that self-deception be viewed as a generalized response set that depresses measures of psychopathology (Sackeim & Gur, 1978, 1979). On the other hand, individuals who commit false positive responses (i.e., misidentifications of other as self), do so because they find self-confrontation pleasurable. The motivational determinants of these errors are opposite in direction to those of the false negative responses. It is expected that subjects who commit such errors will be low on measures of cognitive discrepancy that predict the aversiveness of self-confrontation.

To examine further the motivational contexts of misidentifications of self and others and to test the face validity of the claim that both these errors may be instances of self-deception, subjects were administered a questionnaire designed to assess individual differences in tendencies to engage in self-deception. It was predicted that subjects who committed either type of error would score higher on this measure than subjects who did not misidentify voices of self or others.

In summary, the following set of conditions should be fulfilled in order to demonstrate that misidentifications of voices of self and others are instances of self-deception: Psychophysiological reactivity to the voice of self is greater than reactivity to the voices of others, *regardless* of the correctness of identifications of voices. Subjects demonstrate that they are unaware of committing errors. Subjects who commit false negative responses or false positive responses, by and large, report that they are not aware of having made such errors. On trials following both true positive and false positive responses, subjects are faster in identifying voices, whereas no savings occur following false negative responses. Subjects who commit false positive responses are lower on measures of cognitive discrepancy than subjects who make no such responses. Subjects who commit either false positive and/or false negative responses are higher on a measure of individual differences in tendencies to engage in self-deception than subjects who commit no errors.

Method

Subjects

The subjects were 60 University of Pennsylvania undergraduate volunteers (30 male, 30 female). Only those students who had always lived in the northeastern United States were accepted as subjects.

Procedure

Subjects were given a set of personality and psychopathology inventories to fill out at home, with assurance that all results would be kept confidential. Appointments were made for returning the forms and participating in the experimental part of the study. Before arriving in the laboratory, subjects were informed only that they would be participating in a study on "voice discrimination and personality."

Upon arrival, subjects were taken to an experimental room, where samples of their voices were recorded. Subjects were seated in a chair so that their mouths were no more than one foot (30.5 cm) from a microphone. They were given a paragraph to read in a "normal speaking voice." Each subject read the same paragraph, taken from Kuhn (1962, pp. 124–125). This paragraph was chosen because pilot studies had indicated that subjects needed to pay careful attention to the text to read the paragraph correctly. This insured that few subjects would be paying close attention to particular intonations or phrasings.

After the voice sample was recorded, subjects were escorted to another room for electrode placement. GSR electrodes were placed on the volar surface of the middle phalanges of the second and third fingers of the nondominant hand. Subjects were then escorted to the experimental room.

Upon entering the experimental room, subjects were seated in a chair positioned in front of a small table. GSR electrodes were then connected to leads from a polygraph, which was in the room adjacent to the experimental room. When this was accomplished, the following instructions were read to subjects:

> We would like you now to participate in a rather straightforward task. I will soon place earphones on your head and we will play a tape for you. On the tape, there are 30 voices. There is a period of silence separating each voice. We want you to listen to every voice carefully. As soon as a voice starts, you should decide whether you think the voice is your own or that of a stranger. As soon as you have made a decision, you should press one of the six buttons, which tells us whether you thought the voice was your own or that of a stranger and how certain you are of this decision. The three buttons on the left (right) correspond to the choice of a stranger. The three buttons on the right (left) correspond to the self.[2] If you press Button 1 on either side, it means you are not very certain of your decision. If you press Button 2 on either side, it means you are somewhat certain of your decision, whether it be self or other. If you press Button 3 on either side, it means that you are very certain of your decision. Remember, press the button as quickly as you can, indicating your decision as to whether the voice was your own or not and how certain you are.
>
> Many of the voices you may hear will be on for very short periods. Do not be surprised. Your own voice may appear many times, a few times or even not at all. Press only one button per voice.
>
> In summary, all the task requires is that you listen to each of the 30 voices, make a decision as to whether each voice is that of a stranger or your own, and press one of six buttons telling us what your decision was, and how certain you were. You should do this as quickly as possible. Do you have any questions?

---

[2] The positions of "self" and "other" buttons on the left and right sides of the panel were randomized. Both sets of buttons were numbered from left to right.

Remember, please keep as still as possible and, in particular, do not move your left/right arm. There will now be a resting period of a few minutes before the task begins. I will put the earphones on you so that you can get used to them and, after a short while, come back and tell you when we are ready to begin.

After a period of 5 to 7 minutes, the experimenter returned to inform the subject that the tape was about to be run. During the experiment, subjects were monitored through a one-way mirror.

Each tape contained five groups of six voices, differing in temporal duration (2-, 4-, 6-, 12- and 24-sec periods). Ordering of the groups of voices was sequential, starting with the 2-sec voices. A period of silence, ranging from 20 to 40 sec (M = 30.00 sec) occurred before and after every voice. The position of voices on the tapes was fixed, with the self appearing once in each group (Voices 4, 10, 15, 24, and 28). Two voices of others were also repeated, each appearing once in each group. Male subjects were played tapes that contained male voices, and female subjects heard only voices of women. The tapes for the two sexes were identical in format. Each subject heard the same other voices, and all stimuli began with the third sentence in the recorded paragraph. The voices of others were provided by 17 male and 17 female undergraduates at the University of Pennsylvania. All had been born and lived continuously in the northeastern United States. They did not differ in age from the subjects in this study.

After subjects responded to the 30 voices, the electrodes were removed, and the subjects were taken to another room. They were then given a postexperimental questionnaire and the Self-Deception Questionnaire (SDQ; Sackeim & Gur, 1979). After these tests were completed, subjects were debriefed.

Apparatus

Master tapes were constructed so that silence periods, before and after the self, consisted of unrecordable leader tape. In this way, it was assured that voices would be present on the tapes for the exact temporal duration intended. No master tape was played on more than 11 occasions. The audiotape that was used in this study was a 1.5-mil professional mastering tape.

Silver/silver chloride electrodes, 11 mm in diameter, were used, with K-Y jelly serving as the electrolyte. Electrodes were plated prior to the beginning of the experiment. No pair of electrodes with a resistance greater than 250 Ω or with bias potentials greater than .5 mV was used (Edelberg, 1972; Venables & Martin, 1967). GSR was recorded on a Grass Model 7P polygraph, using a low-level direct-current preamplifier. For measurement of GSR, a constant current of 50 μA was passed.

All voices on the master tapes were recorded on the right channel. Simultaneous with the onset of the audio stimuli on the right channel was the onset of a 1000-cycle/sec sine wave signal on the left channel. These signals, of 150 msec duration, were recorded from a low-frequency function generator. Tapes were played in stereo, with subjects receiving in both ears only the right channel, through modification of a pair of headphones.

The onset of the sine wave signal triggered a voice-operated relay that simultaneously started a reaction time clock and registered on one channel of the polygraph. When subjects pressed one of the six buttons on the panel before them, the reaction time clock was stopped, and a light was lit in the control room indicating which button had been pressed.

Materials

Subjects were administered the Eysenck Personality Inventory (EPI; Eysenck & Eysenck, 1963) and a revised version of the Preexamination Questionnaire (PEQ; Liebert & Morris, 1967; Morris & Liebert, 1969, 1970).[3] The 10 items from the Neuroticism scale of the EPI used by Sackeim (Note 2) to measure cognitive discrepancy were again used in this fashion, along with the Worry scale of the PEQ. These inventories were administered before the experimental session.

The postexperimental questionnaire was concerned with whether subjects had hearing difficulties, how many times they thought they heard tape recordings of their own voices, whether they thought they committed either type of error, how certain they were when they first heard their own voices, what their reaction was to the voice, and how frequently in the past they had been exposed to recorded versions of their voices.

In addition, subjects filled out a questionnaire specifically constructed to measure self-deception, the SDQ. The questionnaire consists of 20 questions, all of which are meant to be psychologically threatening (e.g., Have you ever enjoyed your bowel movements? Have you ever doubted your sexual adequacy?). The questionnaire is answered on a 1 (not at all) to 7 (very much so) Likert-type scale, with scores of 1 or 2 on individual questions keyed as self-deceptions. Total SDQ scores (range = 0–20) are used in the analysis of results.

Results

Effects of Confrontation With Self and Others

In line with previous studies using self-confrontation procedures (e.g., Holzman et

[3] The PEQ, which is usually administered prior to examinations, was revised so that subjects indicated how worried and anxious they were about filling out the set of personality and psychopathology inventories.

al., 1966; Olivos, 1967; Sackeim, Note 2), subjects showed greater psychophysiological reactivity to the voice of self than of the voices of others. A three-factor (Self versus Other Voice Conditions × Sex × Trials) repeated measures analysis of variance on GSR change-in-conductance scores [4] revealed a significant effect of self versus other conditions, $F(1, 58) = 42.43$, $p < .001$. Sex was also associated with GSR reactivity, with females more reactive than males, $F(1, 58) = 4.41$, $p < .05$. There was a main effect of order of trials, $F(4, 232) = 4.47$, $p < .005$, as well as an interaction between the factor of self versus other voice conditions and that of order of trials, $F(4, 231) = 11.82$, $p < .001$. This interaction is illustrated in Figure 1.

As seen in Figure 1, the significant interaction indicates that psychophysiological reactivity in the other voice condition habituated with increasing number of trials, whereas reactivity to self did not habituate.

A similar analysis of variance with reaction time as the dependent variable revealed that reaction time was longer on trials of self compared to trials in which the voices of others were presented, $F(1, 58) = 18.84$, $p < .001$. There was no significant effect for sex, $F(1, 58) < 1$. The main effect of order of trials was significant, $F(4, 232) = 20.10$, $p < .001$, as was the interaction between this factor and that of self versus other voice conditions, $F(4, 231) = 14.68$, $p < .001$. This interaction is illustrated in Figure 2.

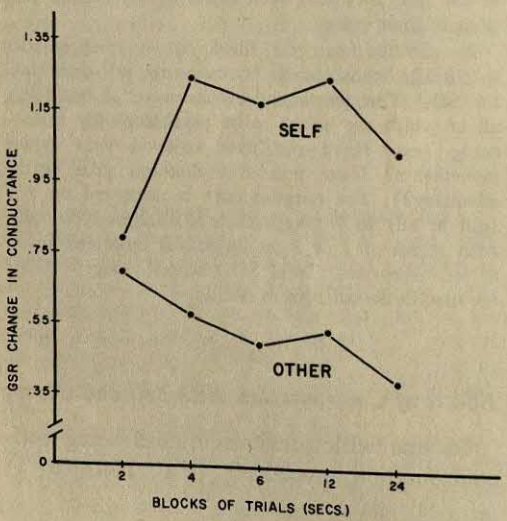


Figure 1. Galvanic skin response (GSR) reactivity (in micromhos) to voices of self and others as a function of block of trials. (Each block of trials increasingly differed in voice duration, as indicated on the abscissa.)

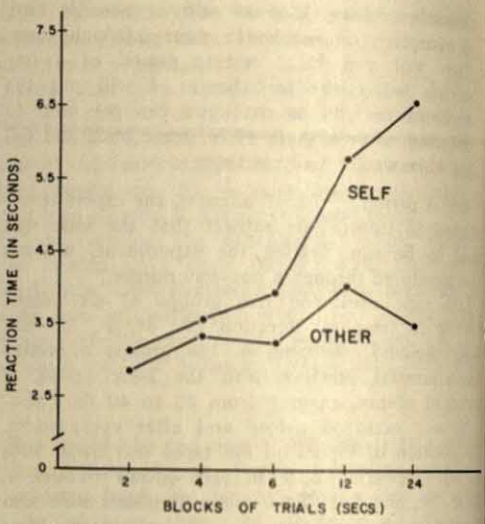


Figure 2. Reaction time to voices of self and others as a function of block of trials. (Each block of trials increasingly differed in voice duration, as indicated on the abscissa.)

Figure 2 suggests that the interaction occurred because the longer reaction time in later periods of the experiment was evidenced in identifications of self, but not in identifications of others.

Criteria for Ascribing Self-Deception

The first criterion: Holding two contradictory beliefs. To satisfy the first criterion for ascribing self-deception, it must be shown that GSR reactivity to the voice of self is as high when a subject denies that the voice is the self as when the subject correctly identifies the voice as the self. Likewise, it must be shown that GSR reactivity to the voices

[4] GSR reactivity scores are measured in units of microhmos (Edelberg, 1972). These scores are computed by converting resistance measures to change-in-conductance scores (i.e., $1000/R_2 - 1000/R_1$). R_2, in k-ohms, is the lowest level of resistance reached by subjects within 5 sec of voice onset. R_1, in k-ohms, is the level of resistance at voice onset.

of others is as low when a subject misidentifies others as self as when the subject correctly identifies others. As stated above, the motivational factors offered to account for false negative and false positive errors are opposite in direction. Subjects were therefore divided into four groups: subjects who made no errors ($n = 15$), subjects who made false negative errors only ($n = 14$),[5] subjects who made false positive errors only ($n = 18$), and subjects who made both false negative and false positive errors ($n = 13$). If subjects held contradictory beliefs when they misidentified voices, we would expect that in the group that committed false negative errors only, levels of GSR for true positive and false negative responses would not differ and would be higher than those for true negative responses. In the group that committed false positive errors only, levels of GSR for true negative and false positive responses should not differ and should be lower than levels of GSR for true positive responses. For the group of subjects who made no errors, levels of GSR should be higher on true positive as compared to true negative responses. Finally, the commission of both types of errors by the same subjects cannot be viewed as bona fide instances of self-deception, given the differing motivational factors that presumably produce each type of error. Accordingly, no expectations could be offered for this group.

Because of the significant interaction between self and other voice conditions and order of trials on GSR, subjects' change-in-conductance scores for all identifications had to be adjusted for the effect of this interaction.[6] To do that, for each subject, trial position was regressed on change-in-conductance scores separately for each voice condition. Adjusted scores, corrected for trial position, were obtained. The mean change-in-conductance scores, adjusted for trials, were computed within subjects for each response category. The results for the four groups of subjects are presented in Figure 3.

As expected, in the group that made no errors, GSR reactivity was greater for the voice of the self (true positive responses) as compared to the voices of others (true negative responses), $t(14) = 4.63$, $p < .001$, one-tailed. The crucial test of the self-deception hypothesis pertains to the behavior of subjects in the second and third groups. Consistent with this hypothesis, in the group whose only errors were false negative responses, levels of GSR were higher for both true positive and false negative responses compared to true negative responses, $t(12) = 2.91$, $p < .01$, one-tailed, and $t(12) = 2.14$, $p < .05$, one-tailed, respectively. Levels of GSR for true positive and false negative responses did not differ, $t(12) < 1$. The results for the group whose only errors were false positive responses were also consistent with the self-deception hypothesis. Levels of GSR were higher for true positive responses in comparison to both true negative responses, $t(17) = 4.05$, $p < .001$, one-tailed, and false positive responses, $t(17) = 2.20$, $p < .025$, one-tailed. False positive and true negative responses did not differ in levels of GSR, $t(17) < 1$. The results for the group that committed both types of error differ in pattern from the results of the two groups that committed one type of error. Therefore, it can not be claimed that subjects who committed both types of error held contradictory beliefs when the errors were being made. For this last group of subjects, none of the differences among the four response types reached statistical significance by two-tailed tests. The results for this group indicated that they were relatively insensitive in their self–other identifications on both verbal report and psychophysiological response measures.

In summary, those subjects who com-

[5] One subject, whose only errors were false negative responses, committed five such responses and did not make any true positive responses. Since the analyses of results bearing on the first criterion are within-subjects comparisons of mean change-in-conductance scores for the various response categories, the results for this subject were not included in these analyses.

[6] To determine whether GSR scores were also associated with the number of times subjects correctly identified voices as self, correlations were computed between number of identifications of self and mean GSR scores for all voices and separately for mean GSR scores for each response category. No associations were found.

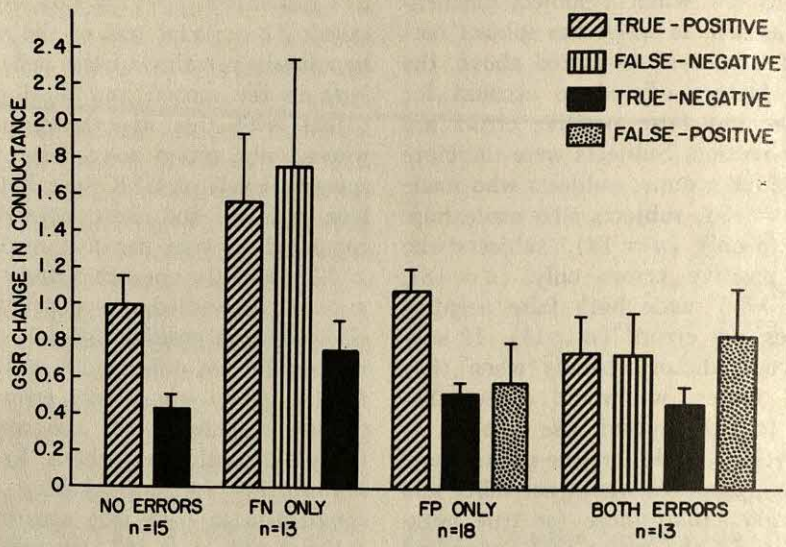


Figure 3. Mean galvanic skin response (GSR) reactivity (in micromhos), adjusted for order of trials, as a function of types of self-report identification for groups that committed no errors (NO ERRORS), false negative errors (FN ONLY), false positive errors (FP ONLY), and false negative and false positive errors (BOTH ERRORS).

mitted one type of error indicated by their GSR reactivity that when they misidentified voices, at some level of processing correct identifications had been made; these subjects, therefore, held contradictory beliefs.

The second criterion: Simultaneity of beliefs. Since the psychophysiological measurements were taken concurrent with subjects' identification of voices, the second criterion was satisfied by the very nature of the experimental procedure.

The third criterion: Nonawareness of misidentifications. Subjects' postexperimental reports as to whether they committed either type of error and their reports as to how many times they believed that the voice of self appeared on the tapes were in high agreement. Only 1 subject out of 27 who committed false negative responses reported having done so. On the other hand, of the 31 subjects who committed false positive responses, 16 claimed to have made such errors. Thus, on the basis of this obtrusive measure, virtually all subjects were not aware of having made false negative errors, whereas about half of the subjects who committed false positive errors lacked awareness of having made these responses.

The results for the unobtrusive measure of awareness were congruent with the findings from subjects' postexperimental reports. Within blocks of trials, reaction time on trials of other voices immediately following correct identifications of self (true positive responses) was first compared to reaction time for the remaining voices of others. This was done in order to establish that conscious belief that a voice was the self resulted in savings in reaction time on the subsequent trial. This was the case both for subjects who committed no false negative responses, $t(32) = 2.68$, $p < .01$, one-tailed, and for subjects who committed false negative errors, $t(25) = 2.12$, $p < .025$, one-tailed. Following false negative responses, however, reaction time was not faster on the subsequent trial of other, $t(25) = -1.91$, *ns*, indicating that savings occurred only after subjects expressed a conscious belief that a voice was the self.

A similar comparison was made between reaction time on trials that immediately followed misidentifications of voices of others as self (false positive responses) and reaction time on remaining voices of others, within blocks of trials (excluding trials of

other voices preceded by the voice of self). Here, the results were somewhat more complex. When this comparison was made for all subjects who committed false positive responses, faster reaction time on trials following false positive responses was not found, $t(30) < 1$. However, subjects were divided into two groups based on whether they were aware of committing false positive errors, as indicated by postexperimental self-reports. It was found that subjects who claimed not to be aware of making these errors were faster in their identifications following false positive responses, $t(14) = 2.25$, $p < .025$, one-tailed. Subjects who claimed to be aware of false positive errors were not faster on trials immediately following such responses than on remaining trials of voices of others. This pattern of results appears to buttress the validity of the unobtrusive measure of awareness. It also indicates that the lack of faster responding subsequent to false negative errors cannot be accounted for simply by the possible retarding effects of the commission of errors per se.

In summary, the results of both the obtrusive and unobtrusive measures of awareness converge in indicating that virtually all subjects who committed false negative responses were unaware of having made such errors. About half of the subjects who committed false positive errors were not cognizant of the fact.

The fourth criterion: The motivational aspect. In order to test the motivational account of misidentifications of voices, scores on our two measures of cognitive discrepancy were compared for subjects who committed two or more false positive errors and subjects who did not commit such errors. As expected, the groups differed on the Discrepancy scale taken from the EPI, $t(40) = 2.00$, $p < .05$, one-tailed, and on the Worry scale of the PEQ, $t(40) = 2.23$, $p < .025$, one-tailed. The group hypothesized to be actively seeking out self-confrontation (false positive responders) scored low on these predictors of the aversiveness of self-confrontation. Since this finding pertains only to one type of error, and in order to test the validity of the claim that both false negative and false positive errors can be instances of self-deception, SDQ scores were compared for subjects who had two or more misidentifications of either type and the remaining subjects. The results indicated that the former group had higher scores on this measure of individual differences in tendencies to engage in self-deception, $t(58) = 1.99$, $p < .05$, one-tailed.

These findings provide initial support for the motivational account of misidentifications of voices of self and others.

Consideration of Alternative Explanations

Independence of GSR and self-reports. The evidence presented in support of the first criterion can be interpreted as indicating that self-report identifications and identifications of voices as inferred from the GSR measure were relatively independent. In particular, when subjects misidentified voices by their self-reports, correct identifications could often be made on the basis of the GSR measure. This pattern of results is reminiscent of what early studies on subception were attempting to demonstrate (e.g., Lazarus & McCleary, 1951). In these studies, instances where correct identifications of stimuli could be attributed on the basis of GSR measures, while verbal reports were incorrect, were taken as evidence for subception. Eriksen (1956, 1958) criticized this interpretation of the findings by claiming that GSR and verbal reports were not independent response systems. He argued that GSR and verbal reports were substantially but imperfectly correlated and that the subception findings could be simply accounted for by the uncorrelated error term for GSR and verbal reports.

In order to examine whether the two response systems were independent, we first relied on Eriksen's method of computing the point-biserial correlation on mean GSR scores for trials of self when subjects were correct and incorrect in their self-reports. The same correlation was computed for trials of the other. Both correlations were not significant, $r_{pb}(25) = .03$ for voices of self, and $r_{pb}(30) = .15$ for voices of the other. These results indicate that whether subjects identified

voices as self or other had little influence on levels of GSR, and therefore, in this respect, the two response systems were independent.

Signal detection theory affords an opportunity to separate within response systems the effects of sensitivity and response bias in the identification of stimuli (Green & Swets, 1966). In the present context, such procedures would be particularly useful in establishing whether subjects who differ in the correctness of their self-reports do not differ in the sensitivity of their GSR in discriminating between voices of self and others. However, the small number of trials administered to each subject and the fact that the voice of self was presented to each subject on only five occasions preclude the application of standard signal detection methodology to these data, since separate measures of sensitivity and response bias can not be computed for each subject. Nonetheless, for descriptive purposes it might be useful to examine the sensitivity of GSR in groups of subjects pooled as a function of their self-report behavior.

In order to obtain confidence levels of the degree to which the GSR measure indicated judgments of self, percentages were computed for the GSR on each trial relative to the largest GSR within each block of trials. These percentages were categorized into 10% units. The number of trials of self and other voices falling within each unit was totaled for each subject. As in the analysis of results bearing on the first criterion, subjects were classified into the four groups on the basis of their self-report identifications. The data for subjects were pooled within groups to form essentially four types of perceivers.

Figure 4 presents receiver operating characteristic (ROC) curves for the four groups on the GSR measure. Cumulative probabilities of identifying the voice of another as self, $P(S/o)$, on the basis of GSR, are on the abscissa, and cumulative probabilities of identifying the voice of self as self, $P(S/s)$, are on the ordinate. A nonparametric measure of sensitivity, the area under the ROC curve, or $P(A)$ (McNicol, 1972), was computed for each group.[7] As can be seen in Figure 4, the sensitivity of the GSR was well above chance for all four groups of subjects. Subjects who did not commit any errors, $P(A) = .86$; subjects whose only errors were false negative responses, $P(A) = .82$; and subjects whose only errors were false positive responses, $P(A) = .85$, did not differ in sensitivity, whereas subjects who made both types of errors, $P(A) = .71$, showed lowered sensitivity in their GSR identifications of self and others. At least for the groups that identified voices in a manner congruent with the motivational account of misidentifications, these findings indicate that the self-report and GSR response systems were independent.

Effects of certainty of identifications. Another alternative account of the findings related to the first criterion might be stated as follows: Subjects have more difficulty and, therefore, are less certain when they identify their own voices as compared to voices of others. This greater uncertainty produces greater psychophysiological reactivity to voices of self. Furthermore, the differences in levels of GSR found for the four response categories may reflect differences in levels of certainty. In particular, the high level of GSR found when subjects commit false negative errors may indicate that subjects are particularly uncertain when such misidentifications are made. Of course, this account would have to posit that subjects are not uncertain when false positive errors are made.

---

[7] The most common measure of sensitivity derived from ROC curves is d'. However, to use d' it must be assumed that the distributions of signal (i.e., self) and of signal plus noise (i.e., other) are equal in variance and that both are normal—assumptions we are not entitled to make with the present data. The nonparametric measure of sensitivity, $P(A)$, does not require fulfillment of these assumptions. Furthermore, signal detection methodology usually enables one to separate the effects of sensitivity from the effects of response criteria or response bias on performance (Green & Swets, 1966). However, while the area under the ROC curve, or $P(A)$, is an acceptable nonparametric equivalent of d', there are no satisfactory nonparametric alternative measures of response criteria (McNicol, 1972). Therefore, response criteria effects were not examined in this analysis.

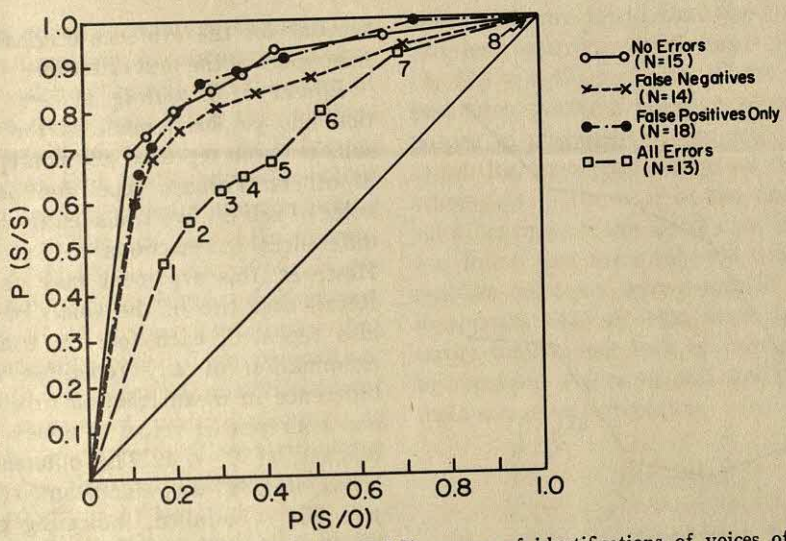


Figure 4. Receiver operating characteristic (ROC) curves of identifications of voices of self and others based on galvanic skin response (GSR) reactivity for groups committing no self-report errors, false negative errors, false positive errors, and false negative and false positive errors (all errors). [Points marked 1 correspond to the cumulative probabilities of hits, $P(S/s)$, and false alarms, $P(S/o)$, for GSR scores that were between 91% and 100% of the largest GSR within their respective block of trials. Points marked 2 correspond to the 81% to 90% interval; points marked 3 correspond to the 71% to 80% interval; points marked 5 correspond to the 41% to 50% interval. Because of low frequencies of occurrence, the data for the intervals 21% to 30% and 31% to 40% were combined and correspond to points marked 7. Likewise, points marked 8 correspond to the 0% to 20% interval.]

A three-factor (Self versus Other Voices × Sex × Trials) repeated measures analysis of variance on certainty judgments showed main effects for self versus other voices, $F(1, 58) = 14.49$, $p < .001$, and trials, $F(4, 232) = 7.92$, $p < .001$. Subjects were less certain of their identifications of the voice of self than of the voices of others, and certainty increased with trials (see Figure 5). Sex had no significant effect, $F(1, 58) < 1$, nor were there any significant interactions. These findings provide some support for the certainty account. There are a number of reasons, however, for rejecting this alternative explanation.

When certainty of identifications was included as a covariate in both analyses of GSR and reaction time, as a function of self versus other voices, sex, and trials, the patterns of results reported earlier were unaltered. In the analysis of the first criterion, the general finding was that levels of GSR did not differ for true positive and false negative responses on the one hand and for true negative and false positive responses on the other. For subjects who committed false negative errors, mean levels of certainty of identifications were compared for false negative and true positive responses. Subjects were neither more nor less certain when they made correct or incorrect identifications of their own voices, $t(25) < 1$. Therefore, the equivalence of levels of GSR despite differences in the correctness of self-reports for these responses cannot be attributed to differences in the certainty of identifications.

A similar comparison was made for levels of certainty on true negative and false positive responses for subjects who committed false positive errors. Here, it was found that when subjects correctly identified others, they were more certain than when they misidentified others, $t(30) = 8.12$, $p < .001$, two-tailed. Assuming that less certainty should result in greater psychophysiological arousal, this difference in certainty for the two response groups cannot account for their equivalence in low levels of GSR.

Finally, mean GSR for the four types of response might be compared at each level of certainty. Given that few subjects committed

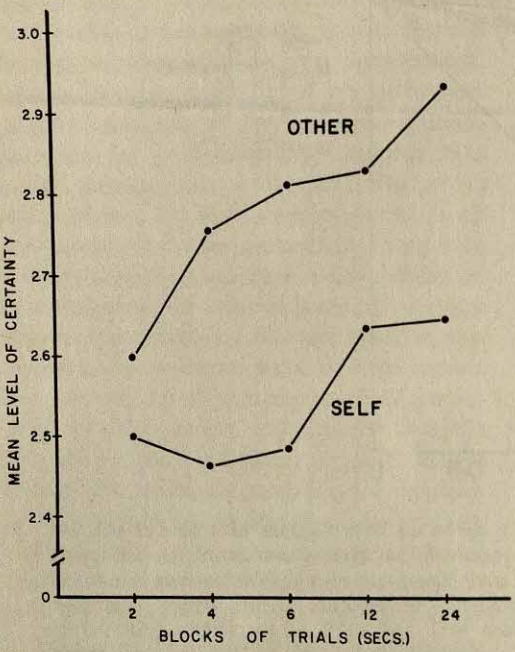


Figure 5. Certainty of identifications of voices of self and others as a function of blocks of trials. (Each block of trials increasingly differed in voice duration, as indicated on the abscissa.)

all four types of response and that it was impossible for any subject to make all four responses at each level of certainty, data had to be pooled across subjects. In Table 1, mean GSR is presented for the four types of self-report response at the three levels of certainty. The pooling of data across subjects, whereby subjects do not equally contribute to each cell of the table, precludes the use of standard statistical tests of significance of differences. However, inspection of Table 1 indicates greater GSR reactivity for true positive responses with the higher levels of certainty. There is no clear-cut relationship between GSR and certainty for false negative responses. For true negative responses, there is a trend for lower levels of certainty to be associated with greater levels of GSR. As in the case for true positive responses, higher levels of GSR for false positive responses were associated with higher levels of certainty. In sum, the trend of the results in Table 1, along with the previous analysis, disconfirms the alternative explanation that differences in the certainty of identifications account for the evidence originally presented in support of the first criterion.

Effects of repeating voices. No distinction has yet been made in the obtained results between repeated and nonrepeated voices of others. Perhaps repetition per se of the voice of self on five trials led to the significant differences in reactions to self and others. However, this argument may be discounted. Recall that two of the voices of others were also repeated, each for five trials, to allow examination of any repetition effects. The difference in mean reaction time on repeated and nonrepeated voices of others was not significant, $t(59) < 1$. The difference in mean levels of GSR was significant, $t(59) = 3.80$, $p < .001$, two-tailed, indicating greater habituation of GSR on repeated trials of other voices. Since this observed effect is opposite in direction to that which would account for the greater reactivity on trials of self, the combining of repeated and nonrepeated voices of others was justified.

Effects of previous exposure to playbacks of self. There may be a more parsimonious explanation of individual differences in rates of misidentifications of the self and others (rather than the motivational account offered here). People differ in their experience with tape playbacks of their own voices. The

Table 1
Mean GSR Reactivity Across Subjects for True Positive, False Negative, True Negative, and False Positive Responses for each of Three Levels of Certainty of Response

Response type	Certainty level	*n*	*M*	*SD*
True positive	High	48	1.18	.86
	Medium	25	.80	.70
	Low	13	.85	.54
False negative	High	14	.77	.80
	Medium	12	1.61	1.77
	Low	7	1.47	1.93
True negative	High	60	.53	.37
	Medium	48	.59	.37
	Low	26	.69	.61
False positive	High	11	.73	.60
	Medium	14	.86	.74
	Low	18	.47	.37

only previous study to examine the relationship between frequency of previous exposure to tape playbacks of the self and the rates of false negative responding found the relationship to be negative (Rousey & Holzman, 1967). However, that study, as well as other studies that have provided data concerning nonrecognition of the self (e.g., Holzman & Rousey, 1966; Holzman et al., 1966; Huntley, 1940; Olivos, 1967; Rousey & Holzman, 1967; Wolff, 1943), used procedures that tended to maximize rates of nonrecognition. This approach was deemed unadvisable for the present experiment, which attempted to minimize rates of nonrecognition attributable to factors other than self-deception. In this regard, the repetition of the trials of the self, the quality of the audio stimuli, the longer durations of voices, and the explicit instructions that the task was one of identification all served to minimize the difficulty of making self–other discriminations. The lowest rate of nonrecognition previously reported was in the Olivos (1967) experiment, where false negative responses constituted 45% of the responses to voices of self. In the present investigation, 18.7% of the trials of self involved false negative responses. The difference is maintained even when the percentage of false negatives on the first trial of self (31.7%) is compared with those of previous studies. Rates of false positive responding had not been reported in the literature. The rate of false positive responding in the present investigation was 3.8%. It can be concluded that the method employed in the present experiment succeeded in minimizing instances of nonrecognition.

To examine the association between exposure to playback of the self and rates of misidentifications, a one-way analysis of variance was performed for groups stratified for number of false negative responses (none, one, and two or more). This analysis indicated no association, $F(2, 56) < 1$. A similar analysis for groups stratified for number of false positive responses indicated that the groups did not differ significantly, $F(2, 57) = 2.74$, $.05 < p < .1$. However, when subjects who committed two or more false positive errors were compared to the rest of the sample, it was found that the former group had less experience with tape playbacks, $F(1, 57) = 5.47$, $p < .025$. These results suggest that overall frequency of previous exposure to playbacks of the self does not account for misidentifications of the self and others. As in the case of the analysis of results bearing on the third criterion, where it was found that some subjects who made false positive responses were aware of their errors, it appears that at least some false positive errors may reflect lack of previous exposure to recorded voices of self and are unlikely instances of self-deception.

Discussion

The results of Experiment 1 support the contention that there is a phenomenon that fits the concept of self-deception, since confirming evidence was found for fulfilling the four criteria for its ascription. When subjects misidentified the voices of self and others, they showed that at some level of processing correct identifications were made; their levels of GSR did not differ from those when they correctly identified voices, and therefore, they simultaneously held contradictory beliefs. Furthermore, subjects were not aware of misidentifying the voice of self and sometimes were not aware of incorrectly identifying voices of others. Finally, there is initial evidence that misidentifications of the voices of self and others are motivated. It was claimed that subjects who misidentify others as self are narcissistic in their self-regard and do not find self-confrontation aversive. In fact, they seek it out. These subjects scored low on measures that predict the aversiveness of self-confrontation. Subjects who misidentified either voices of self or others scored high on a measure of individual differences in tendencies to engage in self-deception.

Of the four criteria for ascribing self-deception, the results of the present study are least compelling in regard to satisfying the fourth criterion of motivation. The view that misidentifications of self and others are motivated requires further substantiation. The evidence presented so far is correlational in nature and does not fully address the question of whether

these misidentifications are purposeful, an issue implicit in attributions of motivation (Irwin, 1971). Furthermore, the results bearing on the fourth criterion pertain primarily to subjects who committed false positive errors. The claim that subjects who commit false negative errors do so because they find self-confrontation particularly aversive could not be examined. Additional evidence must be sought relating levels of cognitive discrepancy or self-esteem to misidentifications of both self and others in order to support the claim that these errors are motivated.

Experiment 2

Recent studies have shown that experimental manipulations of cognitive discrepancy influence selective exposure to the self. Duval, Wicklund, and Fine (cited in Duval & Wicklund, 1972, pp. 16–21) found that subjects who received prior negative false feedback about themselves departed from a room in which they were confronted with a mirror sooner than subjects who received prior positive false feedback or subjects who were not confronted with a mirror. Gibbons and Wicklund (1976) found that subjects whose self-esteem was enhanced by positive interaction with a female confederate subsequently spent more time listening to tapes of their own voices compared to subjects who experienced decreased self-esteem. Davis and Brock (1975) found that self-confronted subjects who received positive false feedback concerning a bogus test of creativity emitted more first person pronouns, when asked to guess English translations of foreign words, than subjects who received negative false feedback or subjects who were not self-confronted.

The results of these studies indicate that after experiences of negative feedback, self-esteem is lowered, and confrontation with the self becomes more aversive. On the other hand, positive feedback enhances self-esteem and makes self-confrontation less aversive. If the misidentifications subjects make when self-confronted are indeed motivated, one would expect subjects who have experienced failure or who have lowered self-esteem to demonstrate greater difficulty in making identifications of self. They should be slower in their reaction time on trials of self, they should be less certain in making such identifications, and, in particular, they should commit more false negative errors. Subjects who have experienced success should show less difficulty in identifying the self. Most importantly, they should commit a greater number of false positive errors.

Method

Subjects

The subjects were 60 University of Pennsylvania undergraduates (30 male, 30 female) who volunteered to participate in a study on intelligence and voice discrimination. All subjects had been born and lived continuously in the northeastern United States.

Materials and Procedure

When subjects reported for the study, they were informed that they would participate in two experiments, one of which required the recording of a speech sample. They were then asked to read the same paragraph that was used in Experiment 1, and their voices were recorded. Subjects were then greeted by another experimenter, who told them that they were to undergo an assessment of their intellectual capacities. They were informed that the purpose of that part of the study was to validate a short test designed to measure intelligence among college students. They were told that they would be presented with 15 multiple-choice synonym problems and that they must provide an answer to each problem within 30 sec of presentation. They were also informed that they would receive feedback concerning their performance on each question and on the test as a whole.

The experimenter then presented subjects with 15 synonym problems taken from the *Guilford-Zimmerman Aptitude Survey, Verbal Comprehension* (1953). Each problem was presented individually on an index card. The experimenter used a stopwatch to time subjects' responses. If a subject did not produce a response within 30 sec of presentation, the experimenter informed the subject that time had expired and requested an answer. Subjects were given feedback as to the correctness of each response, and when errors were made, subjects were told the correct answers. When the test was completed, subjects were told how many of the 15 problems they had answered correctly.

Subjects in the success condition were presented with the 15 easiest problems on the test, and subjects in the failure condition were presented with the 15 most difficult problems. Subjects were randomly assigned to the two conditions. However, final ac-

ceptance into the failure condition was contingent on subjects not succeeding on more than 10 of the 15 items. Similarly, final acceptance into the success group was contingent on correctly solving more than 10 of the problems. Five subjects who were initially assigned to the failure condition succeeded in solving more than 10 of the problems. The results of these subjects were dropped from the analyses reported below. None of the subjects initially assigned to the success group failed to reach the criterion for final inclusion in that group. In order to avoid any bias in assignment of subjects to groups, the results of four subjects, who were initially assigned to the success group and who had succeeded in solving all problems, were also dropped from the analyses. These four subjects were randomly chosen from all subjects in the success condition who had correctly solved the 15 problems. This procedure resulted in a final sample of 30 subjects in the failure group and 21 subjects in the success group.[8]

After completing the verbal problems, subjects were escorted to the experimental room, in which the voice identification task took place. Subjects were administered the Multiple Affect Adjective Checklist (Zuckerman & Lubin, 1965) and were then read the same instructions for the voice identification task as were used in Experiment 1, with the exception that subjects were not informed as to the number of voices that would appear on the tape. The tapes played to subjects contained 28 voices, each voice 4 sec in duration. The voice of the self appeared four times on each tape in Positions 6, 11, 19, and 24. A period of silence, ranging from 15 to 25 sec ($M =$ 20 sec) occurred before and after each voice. Male subjects were played tapes that contained male voices, and female subjects heard only voices of females. Each subject heard the same voices of others, and all voices began with the third sentence in the recorded paragraph. The voices of others were provided by 24 male and 24 female undergraduates at the University of Pennsylvania. All had been born and lived continuously in the northeastern United States. They did not differ in age from the subjects in this study.

After the tape was played, subjects completed a postexperimental questionnaire. Questions were designed to check the effect of the pretreatment manipulations. Respondents were also asked whether they had hearing difficulties, whether they had been aware of committing errors, what their reactions were to their own voices, and how frequently in the past they had been exposed to tape recordings of themselves.

The apparatus was identical to that used in Experiment 1, with the exception that GSR recordings were not taken.

Results

Manipulation Checks

The Multiple Affect Adjective Checklist (Zuckerman & Lubin, 1965) is scored on three dimensions: anxiety, depression, and hostility. As is seen in Figure 6, subjects in the failure condition were more anxious, $t(49) = 2.23$, $p < .025$, one-tailed; more depressed, $t(49) = 3.03$, $p < .005$, one-tailed; and more hostile, $t(49) = 2.03$, $p < .025$, one-tailed, than subjects in the success condition.

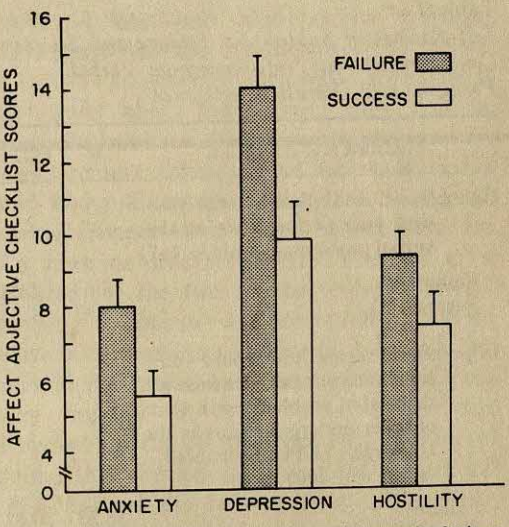


Figure 6. Effects of failure and success manipulations on Multiple Affect Adjective Checklist scores.

Subjects' evaluations of their performance on the verbal problems were assessed on the postexperimental questionnaire. As is shown in Table 2, subjects in the success condition were more pleased with their performance, $t(49) = 4.18$, $p < .001$, one-tailed, and rated their performance as comparing more favorably to that of other undergraduates, $t(49) = 3.92$, $p < .001$, one-tailed, than subjects in the failure condition. Ratings of degree of intelligence regardless of performance did not differ significantly between the two groups, $t(49) = 1.02$, *ns*.

---

[8] It could be argued that the five subjects initially assigned to the failure condition, by solving most of the difficult questions, had an experience of success and should have been reassigned to the success group. Were this to be done, there would be 30 subjects in each condition. With this assignment of subjects to conditions, the results are somewhat stronger than those to be reported below.

Table 2
Self-Ratings of Subjects in Failure and Success Conditions About Performance on Verbal Problems and Overall Intelligence

Measure and condition	*M*	*SD*
How pleased or displeased were you with your performance on the verbal problems? (scale = 1–7)		
Failure[a]	3.63	1.64
Success[b]	5.57	1.53
In percentile terms, how would you compare your performance on the verbal problems with that of other undergraduates at the University of Pennsylvania? (scale = 0–100)		
Failure[a]	58.33	20.83
Success[b]	81.62	19.89
How would you rate your overall intelligence compared to other undergraduates, regardless of your performance today? (scale = 0–100)		
Failure[a]	71.00	16.80
Success[b]	76.67	22.11

[a] $n = 30$.
[b] $n = 21$.

Effects of Success and Failure on Reaction Time and Certainty of Identifications

In line with the findings of Experiment 1, overall, subjects were slower in making identifications on trials of self than on trials of others. A three-factor (Success versus Failure × Self versus Other Voice Conditions × Trials) repeated measures analysis of variance on reaction time revealed a significant main effect for self versus other voice conditions, $F(1, 49) = 11.46$, $p = .002$. The main effect of order of trials was also significant, $F(3, 147) = 21.23$, $p < .001$. In Experiment 1, the order of trials was confounded with the duration of voices. The results of the present experiment indicated that when duration was held constant, subjects became faster in making identifications with increasing numbers of trials. The interaction between success versus failure and self versus other voice conditions was significant, $F(1, 49) = 4.24$, $p < .05$. As is seen in Figure 7, subjects in the failure condition, when compared to subjects in the success condition, were slower in making identifications on trials of self ($p < .05$). Subjects in the two groups did not differ in reaction time on trials of others. This finding indicates that the effects of failure did not generalize in slower reactions to all voices but resulted in retarded performance only on trials of self.

A similar analysis was performed on certainty of identifications. The main effects of self versus other voices, $F(1, 49) = 7.76$, $p < .01$, and order of trials, $F(3, 147) = 5.13$, $p < .005$, were significant. The interaction between success versus failure and self versus other voices did not reach significance, $F(1, 49) = 2.29$, $p > .05$. The results are illustrated in Figure 8. Figure 8 indicates that there was a trend for less certainty in identification on trials of self in the failure group and not in the success group.

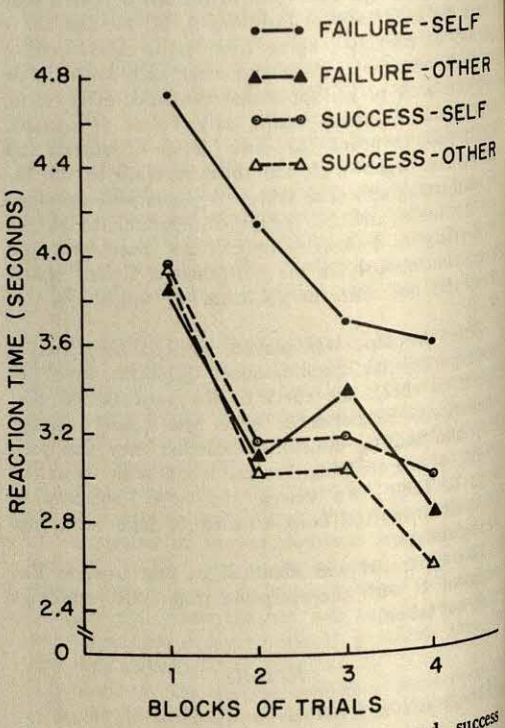


Figure 7. Reaction time of failure and success groups in identifying voices of self and others, as a function of block of trials.

Effects of Success and Failure on Errors of Identification

The rationale for Experiment 2 was based on the argument that errors made in identification of self and others are instances of self-deceptive acts and are therefore motivated. Accordingly, it was expected that subjects who had experienced failure would make more false negative errors, whereas subjects who had experienced success would commit more false positive errors. Eleven out of 30 subjects in the failure group had at least one failure to identify self (false negative error), while only 2 out of 21 subjects in the success group committed at least one false negative response. This effect was significant, $\chi^2(1) = 4.79$, $p < .05$. On the other hand, 13 subjects in the failure group made at least one erroneous judgment of self (false positive error), and 16 subjects in the success group made at least one false positive response. This effect was also significant, $\chi^2(1) = 5.44$, $p < .025$. In short, when people are made to feel good about themselves, they tend to "project" and see themselves in places where they are not. When people are made to feel bad about themselves they tend to "deny" seeing themselves in places where they are. To examine whether overall rates of error differed for the two groups, a degree-of-self-projection measure was computed by scoring each false negative response as -1, each correct identification as 0, and each false positive response as 1. Subjects' scores were summed for the 28 identifications, and it was found that the success group ($M = 1.19$, $SD = 1.68$) and the failure group ($M = -.07$, $SD = 1.81$) differed in types of errors committed, $t(49) = 2.47$, $p < .01$, one-tailed. Thus, the experimental pretreatment differentially influenced rates of types of misidentifications, thereby supporting the motivational account of these errors.[9]

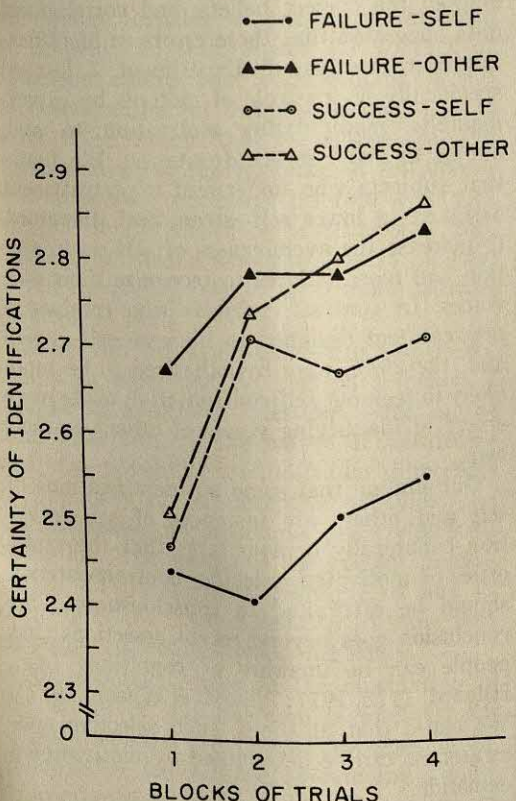


Figure 8. Certainty of failure and success groups in identifying voices of self and others, as a function of block of trials.

Reactions to Hearing the Voice of the Self

In the postexperimental questionnaire, subjects were asked to rate how much they enjoyed hearing their own voices and how unpleasant or pleasant they found their voices to be. Supporting the hypothesis that manipulations of cognitive discrepancy influence the aversiveness of self-confrontation, subjects in the failure condition ($M = 3.73$, $SD = 1.41$) tended to enjoy hearing the voice of self less than subjects in the success condition ($M = 4.33$, $SD = 1.04$), $t(49) = 1.63$, $p < .06$, one-tailed. The failure group ($M = 4.00$, $SD = 1.07$) also rated their voices as less pleasant than the success group ($M = 4.81$, $SD = 1.05$), $t(49) = 2.63$, $p < .01$, one-tailed.

[9] The obtrusive and unobtrusive measures of awareness of errors taken in Experiment 1 were also computed for Experiment 2. Replicating the results of Experiment 1, subjects were not aware of committing false negative errors, whereas some subjects were aware of having made false positive errors.

Discussion

We have argued that when people misidentify voices of self and others, they are engaging in self-deceptive behavior (Sackeim & Gur, 1978). In our view, part of the attribution of self-deception necessitates demonstrating that these errors are motivated. We claimed that individuals who are dissatisfied with themselves find confrontation with the self aversive. Identification of the self is more difficult for them, and they will avoid the aversive consequences of self-confrontation by failing to identify correctly the voice of self. On the other hand, we argued that individuals who hold themselves in high esteem do not find confrontation with the self aversive. Indeed, such individuals narcissistically seek out self-confrontation. Identification of the self is not difficult for them, and they demonstrate their preference for self-confrontation by identifying voices of others as self.

Previous studies have indicated that experimental manipulations of cognitive discrepancy or self-esteem influence selective exposure to the self (Duval & Wicklund, 1972; Gibbons & Wicklund, 1976). The principal measure used in these studies was the amount of time individuals engaged in exposure to the self. If our claim is valid that errors made in identification of the self and others are motivated, then we would expect manipulations of self-esteem to influence the difficulty of making identifications of the self and the rates of types of errors committed in identifying the self and others.

The results of Experiment 2 supported these predictions. Subjects who experienced failure were slower in making identifications of voice of self than subjects given pretreatments of success. The effects of the failure pretreatment did not generalize to identification of all voices, since the failure and success groups did not differ in the speed of their identifications of others. Furthermore, the two groups differed in the types of errors they committed. The failure group engaged in more false negative responding (misidentifications of self); the success group made more false positive errors (misidentifications of others). It could be argued that a number of seemingly "defensive" behaviors, such as self-serving biases in perception and memory, need not be determined by motivational factors (e.g., Greenwald, Note 3). However, the pattern of results presented here indicates that misidentifications of the self and others involve motivated distortions of reality.

General Discussion

This investigation was an attempt to demonstrate that a specific phenomenon, misidentification of the voices of self and others, is an instance of self-deception. In Experiment 1, we showed that when subjects made errors in identifying the voices of self and others, they provided evidence of simultaneously holding correct and incorrect beliefs as to the nature of the voices. We also found that subjects, for the most part, were not aware of holding the correct beliefs, and correlational data suggested that these errors in identification were motivated. Experiment 2 focused specifically on the role of motives by experimentally manipulating motivation to seek out or to avoid self-confrontation. We found that subjects who underwent a pretreatment designed to lower self-esteem and, therefore, to increase the aversiveness of self-confrontation had more failures to recognize their own voices. In contrast, subjects who received a pretreatment designed to increase self-esteem and, therefore, were hypothesized to be more likely to seek out self-confrontation made more errors of identifying voices of others as their own.

Our finding that some misidentifications of self and others are instances of self-deception collaterally demonstrates that the properties of motivated, selective nontransparency should be attributed to consciousness. This conclusion goes beyond recent assertions that people can be unaware of cognitions (e.g., Hilgard, 1976, 1977; Nisbett & Wilson, 1977). We claim that at times, such selective nonawareness can be determined by motivational demands.

A question that presents itself concerns the relationship of the concept of self-deception to the psychoanalytic notion of repression.

Certainly, the concept of repression also attributes a motivated, selective, nontransparent nature to consciousness. Since Freud (1914/1957) wrote that repression is "the cornerstone on which the whole structure of psychoanalysis rests" (p. 16), hundreds of psychological investigations have been interpreted as either propping up or tearing down this cornerstone. In a review of many of these investigations, Holmes (1974) concluded that there is no consistent experimental evidence supporting the existence of repression. However, it should be pointed out that the criteria for the ascription of self-deception may be necessary, but are certainly not sufficient, for the ascription of repression. The ascription of repression requires not only the claim that there is a motivated selective nonawareness of beliefs, it also requires the additional assertion that beliefs not subject to awareness are stored in an unconscious. The unconscious is a functionally independent control system, capable of purposeful influence on behavior. In this respect, the concept of repression entails that consciousness is not only nontransparent but also nonunitary. It is for these reasons that we believe that the demonstration of the existence of self-deception is logically necessary prior to a demonstration of the existence of repression and, further, that such a demonstration in the case of repression is likely to be a more arduous endeavor.

Given this initial evidence that self-deception is an experimentally real phenomenon, a number of additional questions come to the fore. They concern the nature of self-deception, the possibility of individual differences in frequency of self-deceptive behavior, and the processes and structures underlying self-deception. The results of the present investigation do not provide solutions to any of these questions, but in relation to some they do suggest directions for further study.

An important issue in the elucidation of the nature of self-deception is whether acts of self-deception are typical responses of individuals to, for instance, threatening stimuli, or whether they are specific and situation- or stimulus-bound. Our use in Experiment 1 of a paper-and-pencil questionnaire to assess individual differences in tendencies toward self-deception was predicated on the view that self-deception is not a stimulus-bound phenomenon but a generalized response set, or characteristic defense (Hilgard, 1949), the frequency of its use varying among people. In support of this view, we found that subjects who committed either false negative or false positive errors showed greater tendencies to deny psychologically threatening statements on the SDQ. Elsewhere (Sackeim & Gur, 1978, 1979), we have found that there are substantial negative correlations between SDQ scores and degree of self-reported psychopathology. The magnitude of the associations between SDQ and psychopathology measures is greater than that between standard lie scales and self-reported psychopathology. These findings support the contention of Meehl and Hathaway (1946) that "what is much more important, they (lie scales) are mainly directed at the sort of *conscious* falsehood which most writers have stressed, while ignoring the more subtle tendencies to self-deception which are probably of even greater importance in affecting scores" (p. 528). As a generalized response set, self-deception may influence behavior in contexts other than personality testing. For instance, Mischel (1974) argued that the "neurotic paradox" reflects a motivated attempt by individuals to keep specific ideas out of awareness and that all neurotic behavior is self-deceptive (cf. Abramson & Sackeim, 1977). The phenomena found in experiments on cognitive dissonance, self-serving biases in attribution, and studies of subjects' accounts of their behavior in obedience and conformity situations may be likewise interpreted in terms of self-deceptive acts. Since it may be possible to measure relative tendencies to employ self-deception, there is now an opportunity not only to examine individual differences in self-deception but also the influence of self-deception on broad ranges of behavior. As Jacques Rivière (cited in Fingarette, 1969) has suggested, "The discovery of a deceiving principle, a lying activity within us, can furnish an absolutely new view of all conscious life" (p. 1).

Reference Notes

1. Dickinson, W. H., & Ray, T. S. *Immediate effects of body image confrontation with chronic schizophrenic women.* Paper read at the meeting of the Oklahoma State Psychological Association, Oklahoma City, October 1965.
2. Sackeim, H. A. *Personality and the self-confrontation experience.* Unpublished manuscript, Oxford University, 1974.
3. Greenwald, A. G. *The totalitarian ego: Fabrication and revision of personal history.* Manuscript submitted for publication, 1978.

References

Abramson, L. Y., & Sackeim, H. A. A paradox in depression: Uncontrollability and self-blame. *Psychological Bulletin,* 1977, *84,* 838–851.

Alkire, A. A., & Brunse, A. J. Impact and possible causality from videotape feedback in marital therapy. *Journal of Consulting and Clinical Psychology,* 1974, *42,* 203–210.

Anastasi, A. *Psychological testing.* New York: Macmillan, 1961.

Boyd, H. S., & Sisney, V. V. Immediate self-image confrontation and changes in self-concept. *Journal of Consulting Psychology,* 1967, *31,* 291–294.

Bruner, J. S., & Postman, L. Perception, cognition, and behavior. *Journal of Personality,* 1949, *18,* 14–31.

Camus, A. *The fall.* New York: Vintage Books, 1956.

Canfield, J. W., & Gustafson, D. F. Self-deception. *Analysis,* 1962, *23,* 32–36.

Castaldo, V., & Holzman, P. S. The effects of hearing one's own voice on sleep mentations. *Journal of Nervous and Mental Disease,* 1967, *144,* 2–13.

Castaldo, V., & Holzman, P. S. The effect of hearing one's own voice on dream content: A replication. *Journal of Nervous and Mental Disease,* 1969, *148,* 74–82.

Davis, D., & Brock, T. C. Use of first person pronouns as a function of increased objective self-awareness and performance feedback. *Journal of Experimental Social Psychology,* 1975, *11,* 381–388.

Demos, R. Lying to oneself. *Journal of Philosophy,* 1960, *57,* 588–595.

Dixon, N. F. *Subliminal perception: The nature of a controversy.* London: McGraw-Hill, 1971.

Duval, S., & Wicklund, R. A. *A theory of objective self-awareness.* New York: Academic Press, 1972.

Edelberg, R. E. Electrical activity of the skin. In N. S. Greenfield & R. A. Sternback (Eds.), *Handbook of psychophysiology.* New York: Holt, Rinehart & Winston, 1972.

Erdelyi, M. H. A new look at the new look: Perceptual defense and vigilance. *Psychological Review,* 1974, *81,* 1–25.

Eriksen, C. W. Subception: Fact or artifact? *Psychological Review,* 1956, *63,* 74–80.

Eriksen, C. W. Unconscious processes. In J. R. Marshall (Ed.), *Nebraska Symposium on Motivation* (Vol. 6). Lincoln: University of Nebraska Press, 1958.

Eriksen, C. W., & Browne, C. T. An experimental and theoretical analysis of perceptual defense. *Journal of Abnormal and Social Psychology,* 1956, *52,* 224–230.

Eysenck, H. J., & Eysenck, S. B. G. *Eysenck Personality Inventory.* London: University of London Press, 1963.

Fingarette, H. *Self-deception.* London: Routledge & Kegan Paul, 1969.

Freud, S. [On the history of the psychoanalytic movement] In J. Strachey (Ed. and trans.), *The complete psychological works of Sigmund Freud* (Vol. 14). London: Hogarth, 1957. (Originally published, 1914.)

Gardiner, P. L. Error, faith and self-deception. *Proceedings of the Aristotelian Society,* 1970, *50,* 221–243.

Gibbons, F. X., & Wicklund, R. A. Selective exposure to self. *Journal of Research in Personality,* 1976, *10,* 98–106.

Gide, A. *The counterfeiters.* New York: Modern Library, 1955.

Green, D. M., & Swets, J. A. *Signal detection theory and psychophysics.* New York: Wiley, 1966.

Guilford-Zimmerman Aptitude Survey. Beverly Hills, Calif.: Sheridan Supply Co., 1953.

Gur, R. C., & Sackeim, H. A. Self-confrontation and psychotherapy: A reply to Sanborn, Pyke and Sanborn. *Psychotherapy: Theory, Research and Practice,* in press.

Hilgard, E. R. Human motives and the concept of the self. *American Psychologist,* 1949, *4,* 374–382.

Hilgard, E. R. Neodissociation theory of multiple cognitive control systems. In G. E. Schwartz & D. Shapiro (Eds.), *Consciousness and self-regulation, advances in research* (Vol. 1). New York: Plenum Press, 1976.

Hilgard, E. R. *Divided consciousness: Multiple controls in human thought and action.* New York: Wiley, 1977.

Holmes, D. S. Investigations of repression: Differential recall of material experimentally or naturally associated with ego threat. *Psychological Bulletin,* 1974, *81,* 632–653.

Holzman, P. S., & Rousey, C. The voice as a percept. *Journal of Personality and Social Psychology,* 1966, *4,* 79–86.

Holzman, P. S., Rousey, C., & Snyder, C. On listening to one's own voice: Effects on psychophysiological responses and free associations. *Journal of Personality and Social Psychology,* 1966, *4,* 432–441.

Howie, D. Perceptual defense. *Psychological Review,* 1952, *59,* 308–315.

Huntley, C. W. Judgments of self based upon records of expressive behavior. *Journal of Abnormal and Social Psychology,* 1940, *35,* 398–427.

Irwin, F. W. *Intentional behavior and motivation: A cognitive theory*. Philadelphia, Pa.: Lippincott, 1971.

Kuhn, T. S. *The structure of scientific revolutions*. Chicago: University of Chicago Press, 1962.

Lazarus, R. S., & McCleary, R. A. Autonomic discrimination without awareness: A study of subception. *Psychological Review*, 1951, *58*, 113–122.

Liebert, R. M., & Morris, L. W. Cognitive and emotional components of test anxiety: A distinction and some initial data. *Psychological Reports*, 1967, *20*, 975–978.

Lykken, D. T. The GSR in the detection of guilt. *Journal of Applied Psychology*, 1959, *43*, 385–388.

Lykken, D. T. Psychology and the lie detector industry. *American Psychologist*, 1974, 29, 725–738.

Martin, D. G., Hawryluk, G. A., & Guse, L. L. Experimental study of unconscious influences: Ultrasound as a stimulus. *Journal of Abnormal Psychology*, 1974, *83*, 589–608.

McNicol, D. *A primer of signal detection theory*. London: Allen & Unwin, 1972.

Meehl, P. E., & Hathaway, S. R. The K factor as a suppressor variable in the Minnesota Multiphasic Personality Inventory. *Journal of Applied Psychology*, 1946, *30*, 525–564.

Mischel, T. Understanding neurotic behavior: From "mechanism" to "intentionality." In T. Mischel (Ed.), *Understanding other persons*. Totowa, N.J.: Rowman & Littlefield, 1974.

Morris, L. W., & Liebert, R. M. Effects of anxiety on timed and untimed intelligence tests. *Journal of Consulting and Clinical Psychology*, 1969, *33*, 240–244.

Morris, L. W., & Liebert, R. M. Relationship of cognitive and emotional components of test anxiety to physiological arousal and academic performance. *Journal of Consulting and Clinical Psychology*, 1970, *35*, 332–337.

Murphy, G. Experiments in overcoming self-deception. *Psychophysiology*, 1970, *6*, 790–799.

Murphy, G. *Outgrowing self-deception*. New York: Basic Books, 1975.

Murray, H. A. Studies of stressful interpersonal disputations. *American Psychologist*, 1963, *18*, 28–36.

Nisbett, R. E., & Bellows, N. Verbal reports about causal influences on social judgments: Private access versus public theories. *Journal of Personality and Social Psychology*, 1977, *35*, 613–624.

Nisbett, R. E., & Wilson, T. C. Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, 1977, *84*, 231–259.

Olivos, G. Response delay, psychophysiologic activation and recognition of one's own voice. *Psychosomatic Medicine*, 1967, *29*, 433–440.

Orne, M. T., Thackray, R. I., & Paskewitz, D. A. On the detection of deception: A model for the study of the physiological effects of psychological stimuli. In N. S. Greenfield & R. A. Sternbach (Eds.), *Handbook of psychophysiology*. New York: Holt, Rinehart & Winston, 1972.

Penelhum, T. Pleasure and falsity. In S. Hampshire (Ed.), *Philosophy of mind*. New York: Harper & Row, 1966.

Rousey, C., & Holzman, P. S. Recognition of one's own voice. *Journal of Personality and Social Psychology*, 1967, *6*, 464–466.

Sackeim, H. A., & Gur, R. C. Self-deception, self-confrontation, and consciousness. In G. E. Schwartz & D. Shapiro (Eds.), *Consciousness and self-regulation, advances in research* (Vol. 2). New York: Plenum Press, 1978.

Sackeim, H. A., & Gur, R. C. Self-deception, other-deception, and self-reported psychopathology. *Journal of Consulting and Clinical Psychology*, 1979, *47*, 213–215.

Sackeim, H. A., Packer, I. K., & Gur, R. C. Hemisphericity, cognitive set, and susceptibility to subliminal perception. *Journal of Abnormal Psychology*, 1977, *86*, 624–630.

Sartre, J. P. [*Being and nothingness: An essay on phenomenological ontology*] (H. Barnes, Trans.). London: Methuen, 1958.

Siegler, F. A. Demos on lying to oneself. *Journal of Philosophy*, 1962, *59*, 469–475.

Storms, M. D. Videotape and the attribution process: Reversing actors' and observers' points of view. *Journal of Personality and Social Psychology*, 1973, *27*, 165–175.

Venables, P. H., & Martin, I. Skin resistance and skin potential. In P. H. Venables & I. Martin (Eds.), *A manual of psychophysiological methods*. Amsterdam: North-Holland, 1967.

Verwoerdt, A., Nowlin, J. B., & Agnello, S. A. A technique for studying effects of self-confrontation in cardiac patients. *Health Sciences TV Bulletin*, 1965, *2*, 1–6.

Wason, P. C., & Johnson-Laird, P. N. *Psychology of reasoning: Structure and content*. Cambridge, Mass.: Harvard University Press, 1972.

Wolff, W. *The expression of personality*. New York: Harper, 1943.

Wundt, W. *An introduction to psychology*. New York: Arno Press, 1973. (Originally published, 1912.)

Zuckerman, M., & Lubin, B. *Manual for the Multiple Affect Adjective Checklist*. San Diego, Calif.: Educational and Industrial Testing Service, 1965.

Received March 10, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 2, 170-179

Television Viewing and Fear of Victimization: Is the Relationship Causal?

Anthony N. Doob and Glenn E. Macdonald
University of Toronto
Toronto, Canada

Previous findings have suggested that people who watch a lot of television are more likely to fear their environment than are those who report being less frequent viewers of television. From this simple correlation, previous authors have suggested that television causes people to overestimate the amount of danger that exists in their own neighborhoods. The present study attempted to replicate this finding and to determine if the apparent effect was due to a previously uncontrolled factor: the actual incidence of crime in the neighborhood. Respondents to a door-to-door survey indicated their media usage and estimated the likelihood of their being a victim of violence. Neighborhoods were chosen so as to include a high- and a low-crime area in downtown Toronto and a high- and a low-crime area in Toronto's suburbs. Pooling across the four areas sampled, the previous findings were replicated. However, the average within-area correlation was insignificant, suggesting that when actual incidence of crime is controlled for, there is no overall relationship between television viewing and fear of being a victim of crime. A multiple regression analysis and a canonical correlation analysis confirmed these findings.

A variety of social problems have been attributed to television viewing. It is said that television makes people more violent, that it lowers the level of literacy in the population, and that it distorts the viewer's perception of the world. There is little denying that the picture of reality that comes into people's homes is not an accurate reflection of their own society. That we learn from television, as we learn from every other medium, seems intuitively plausible independent of research results. Gerbner and Gross (1974, 1976a, 1976b; Gerbner et al., 1976), however, have suggested something even more serious than simple learning effects: that people not only learn factual information, such as the proportion of people involved in law enforcement, but that they generalize from the information that they get from television. In particular, Gerbner and his associates show that those who watch a lot of television are more likely to feel that they might be involved in some kind of violence during a given week than do those who watch relatively little television. This same pattern of results shows up in a variety of questions having to do with the viewers' perceptions of various aspects of the society in which they live. As Gerbner et al. (1976) point out, "Their heightened sense of fear and mistrust is manifested in their typically more apprehensive responses

We wish to thank Julian Roberts for his help and useful suggestions at all phases of this work and Bob Gebotys for his help in analyzing the data. The Metropolitan Toronto Police were very open and helpful to us in providing the data necessary for choosing our experimental neighborhoods. The research was supported by the (Ontario) Royal Commission on Violence in the Communications Industry, and a version of this study is published in Volume V of that commission's report. We wish to thank Mr. C. K. Marchant, the Director of Research of that commission, for his help and encouragement throughout all phases of the work. The data analysis was supported, in part, by funds provided to the Centre of Criminology, University of Toronto, by the Ministry of the Solicitor General, Canada.

Requests for reprints should be sent to Anthony N. Doob, Department of Psychology, University of Toronto, Toronto, Ontario, Canada M5S 1A1.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3702-0170$00.75

to questions about their own personal safety, about crime and law enforcement, and about trust in people" (p. 9).

Obviously, heavy television viewing is not independent of other social factors. Gerbner and Gross (1976a) have found that "heavy viewing is part and parcel of a complex syndrome which also includes lower education, lower mobility, lower aspirations, higher anxieties, and other class, age, and sex-related characteristics" (p. 191).

Because of the problem of confounding variables, Gerbner has been careful to break down his data on various other characteristics of television viewers such as age, sex, educational level, news reading, news magazine reading, prime-time viewing, and viewing or nonviewing of TV news. The notable finding in all of these comparisons is that although there may be main effects of some of these other characteristics, in all cases, heavy viewers are more likely than light viewers to feel that they might be involved in some violence.

No list of possible confounding variables can be complete. The worry of any researcher doing correlational research and wishing to make a causal statement is that some other variable would, in fact, account for the effect apparently demonstrated. We felt that there is one quite plausible factor that might account for the correlation between viewing and fear of violence: People who watch a lot of television may have a greater fear of being victims of violent crimes because, in fact, they live in more violent neighborhoods.

The study that this explanation suggests, then, is quite obvious: A survey of the television viewing habits of people and their perception of being involved in violence should be performed in both high- and low-crime neighborhoods. Pooling across neighborhoods, we should be able to replicate Gerbner's and his associates' findings; within neighborhoods, however, the effect should be substantially reduced or eliminated.

Method

For purposes of efficient distribution of resources, the Metropolitan Toronto Police have divided Toronto into approximately 210 patrol areas. The size of these patrol areas varies not only as a function of the resident population but also as a function of the number of calls of all types that the police receive in the area: Busy areas thus tend to be smaller in terms of the size of the population served and in terms of geographic area. The police identified for us the 10 patrol areas with the highest number of reported assaults and woundings and the 14 areas with the lowest number of reported assaults and woundings for the 7-month period ending 2 months before the beginning of the survey. From these data, four geographic areas, approximately equal in size, were chosen. Two (one within the city of Toronto, the other suburban) were high in reported crime; two (one city, one suburban) were low in reported crime. It is difficult to estimate the exact rates of crime for the four areas. However, very rough estimates would suggest that the rates of assaults and woundings per 100,000 resident population for the 7-month period for the four designated areas would be the following: high-crime city, 614; low-crime city, 8; high-crime suburb, 195; low-crime suburb, 6. It must be emphasized that these are very rough figures: The low-crime areas each had only two reported assaults (and no woundings) for the entire 7-month period; hence the estimates are bound to be unstable. There were eight patrol areas constituting the high-crime city area; one patrol area was sampled for the high-crime suburban area and two each for the low-crime areas.

Obviously, the four areas differ considerably on a large number of social variables other than reported crime rates. The high-crime city area contains a portion of the downtown commercial/entertainment district of the city, the largest block of public housing in the metropolitan area, and much of the poorest portion of the population. The low-crime city area is largely expensive, single, detached houses and is one of the more exclusive residential areas. The high-crime suburban area contains a high concentration of low-rise public housing and is generally fairly poor. The low-crime suburban area is mostly single, detached, middle-class housing.

Random households were chosen within each of these areas. Interviewers, employed by a commercial survey company, did a door-to-door survey. The person who answered the door was asked to list all of the people over 18 years of age living in the household. One of these people was then chosen at random by the interviewer. If this person could not be interviewed at that time or at some mutually acceptable time, the interviewer went on to the next randomly chosen household, and the procedure was repeated. The effect of this selection procedure was an oversampling of women (70.5%) and, presumably, a general oversampling of those who spend much of the time at home. Although this effect would be unfortunate if one were interested in estimating population values for the measures that were taken, it was less relevant in our study, where we were interested in the relationship between television viewing and fear of criminal victimization.

Table 1
Mean Fear-of-Crime Factor Scores for Each of the Sampled Areas

Area	City		Suburb	
	M	*n*	*M*	*n*
High crime	.28	83	.15	69
Low crime	−.34	71	−.13	77

Note. The higher the number, the more fear.

Respondents were first asked to indicate those programs that they had watched during the previous week. According to our interviewers, very few people had any difficulty in doing this. They were then asked to complete a 37-item fixed-alternatives questionnaire. This questionnaire consisted of six questions dealing directly with the person's estimate of his or her own likelihood of being a victim of a crime; four questions dealing with estimates of the likelihood of particular groups of people being victims; four questions dealing with the perception of crime in general being a problem and there being a need for more police personnel; two questions dealing with the necessity to arm oneself; eight questions of a factual nature dealing with crime; three questions dealing with society's response to crime; four questions dealing with the respondents' view of Toronto with respect to crime; three questions dealing with the respondents' prediction of their response to a request for help; and three questions dealing with media usage. The whole interview took approximately 45 minutes on the average.

For purposes of analyzing the types of television, we decided to use the number of programs watched as an index of total viewing. In addition, programs were coded by a research associate into violent and nonviolent types before the tabulation of the other data. It should be pointed out that this last measure is, necessarily, somewhat subjective. However, as will be seen, this turns out not to be a serious problem in understanding the results.

Results

In order to reduce the number of measures to a somewhat workable number, a factor analysis was performed [1] on the 34 opinion questions. Using a varimax rotation, only one factor accounted for a substantial amount of the common variance. The percentages of the common variance accounted for by the first 4 of the 11 factors were 35.9%, 12.5%, 10.8%, and 8.4%. The questions that loaded highest on the first factor are shown in Table 4; they were the six questions related to the respondents' estimates of their own chance of victimization, two of the questions dealing with the chances of victimization of particular groups, and one of the questions dealing with crime as a general problem. Generally speaking, it seems fair to label this factor "fear of crime." Each of the next three factors had substantial loadings from only one or two questions.

As one would expect, the residents of the four areas differed significantly on their overall fear of crime. The average factor scores for the four areas are shown in Table 1. Analysis of variance on the factor scores revealed a main effect for high-/low-crime area that was highly significant, $F(1, 296) = 17.79$, $< .01$. Neither the city/suburb effect nor the interaction was significant. It is clear, then, that people who live in high-crime areas, are, in fact, more afraid.

The four areas sampled also differed on their exposure to the various media. Table 2 presents these data. Overall, people in high-crime areas watched more television and, generally speaking, tended to watch more violent television. Although there were interactions between the two factors on these two measures, for the purposes of this article, these interactions are not very important. As one might expect, since the areas differed on so many dimensions, there were also effects on self-report of exposure to radio news: People living in low-crime areas tended to report listening to radio news more frequently. Furthermore, the reported frequency of newspaper reading was higher in low-crime areas and in the city.

Gerbner and his associates (Gerbner & Gross, 1974, 1976a, 1976b; Gerbner et al., 1976) do not directly present measures of association between the total amount of tele-

[1] The input for the factor analysis consisted only of those 300 respondents (of the total of 408) who answered every question. Most of the other 108 respondents failed to answer only a few of the questions. The proportion of complete questionnaires varied somewhat from area to area. The numbers of complete/total questionnaires are as follows: high-crime city, 83/119, or 70%; low-crime city, 71/118, or 60%; high-crime suburb, 69/85, or 81%; low-crime suburb, 77/86, or 90%.

Table 2
Media Usage for the Four Areas

Medium	High-crime area		Low-crime area		*F* value		
	City (119)	Suburbs (85)	City (118)	Suburbs (86)	High/low crime	City/ suburb	Interaction
Total TV	36.25	31.71	18.89	25.03	25.21**	<1	4.98*
TV violence	6.97	3.73	2.11	3.33	23.25**	2.72	18.11**
TV news	3.07	2.99	3.72	3.74	2.83	<1	<1
Radio news	5.07	4.96	5.44	5.37	9.79**	<1	<1
Newspaper reading	4.78	4.58	5.26	4.80	6.89**	6.01*	<1

Note. For TV viewing, numbers refer to mean number of programs watched during the previous week. The other measures are mean values on a scale where 1 equals never and 6 equals daily. *n*s are in parentheses.
* $p < .05$.
** $p < .01$.

vision viewed by their respondents (in response to the question "How many hours a day do you usually watch television? Please include morning, afternoon, and evening") and their fear of being a victim of a violent crime (in response to the question "During any given week, what are your chances of being involved in some type of violence—about a 50–50 chance, about a 1-in-10 chance, or about a 1-in-100 chance?"). However, estimating from the data that are presented in the various reports, we calculated a phi coefficient of .13 and a contingency coefficient of the same value.

Looking at our data, then, we calculated the (Pearson) correlation between our fear-of-crime factor scores and our various measures of media usage. These correlations are presented in the first column of Table 3. It is quite clear that the basic effect is much the same as that found by Gerbner and his associates: Across the four areas, those who watched the most television (or violent television) tended to be those who were the most afraid. However, the effect *within* area is not quite so simple: Although the effect would appear to hold in the high-crime area of the city, it tended to disappear for the other areas. Indeed, the average correlations (last column of Table 3) indicate that there is essentially no relationship between media usage and fear of crime when the effect of neighborhood is removed. We have suggested that the artifact that created the first two correlations in the first column might be labeled "actual incidence of crime." However, in terms of the focus of this article (the relationship of media usage to fear of crime), the

Table 3
Correlations Between Media Usage and Fear-of-Crime Factor Scores for all Subjects (Pooled), for Each of the Four Areas, and for the Average of the Four Areas

Medium	Pooled across all areas	High crime		Low crime		Average correlation
		City (83)	Suburb (69)	City (71)	Suburb (77)	
Total TV	.18**	.24*	.16	.06	−.09	.09
TV violence	.18**	.22*	−.03	.14	−.04	.07
TV news	.05	.14	−.04	.05	.06	.05
Radio news	.05	.18	−.09	−.02	.21	.07
Newspaper reading	−.07	−.20*	−.14	.09	.15	−.03

Note. Positive correlations indicate more fear associated with higher media usage. *n*s are in parentheses.
* $p < .05$.
** $p < .001$.

Table 4
Fear-of-Crime Questions and the Correlations Between Responses to Each Question and Total TV Viewing and TV Violence for the Four Areas Pooled and the Average of the Four Areas Calculated Individually

Question	Total TV		TV violence		High TV viewing associated with
	Pooled	Average within area	Pooled	Average within area	
1. To what extent are crimes of violence a serious problem in your neighborhood? (399)	.07	−.02	.16*	0	Serious problem
2. What do you think the chances are that if you were to walk alone at night on the residential streets of your neighborhood each night for a month that you would be the victim of a serious crime? (391)	.18*	.10*	.19*	.05	High probability (1 in 10)
3. If a child were to play alone in a park each day for a month, what do you think the chances are that he would be the victim of a violent crime? (382)	.12*	.02	.22*	.12*	High probability (1 in 10)
4. If you were to walk by yourself in a park close to your home each night for a month, what do you think the chances are that you would be the victim of a serious crime? (391)	.10*	.02	.14*	.04	High probability (1 in 10)
5. What do you think the chances are that an unaccompanied woman would be the victim of a violent crime late at night in a Toronto subway station? (389)	.10*	.04	.09	.04	High probability (1 in 10)
6. What do you think the chances are that you, one of your family, or one of your close friends might be the victim of an assault during the next year? (385)	.13*	.02	.12*	−.02	High probability (1 in 10)
7. How likely do you think it is that you or one of your close friends would have their house broken into during the next year? (405)	−.04	−.07	.01	−.01	Extremely unlikely
8. Do you ever decide not to walk alone at night because you are afraid of being the victim of a violent crime? (402)	.05	.06	−.04	−.04	Very often
9. Is there any area around your home (i.e., within a mile) where you would be afraid to walk alone at night? (403)	.12*	.05	.02	−.06	Yes

Note. *n*s of respondents for each question are in parentheses.
* $p < .05$.

name given to this variable is unimportant: When the effect of neighborhood is removed, the "effect" of television is reduced to almost nothing.

Clearly, however, the artifact (whatever it is called) measures in only the crudest way the amount of crime that a person is exposed to. For example, a person living in one part of what we have labeled as a high-crime section of the city might, in fact, be quite safe: The crimes might well be in a different section of that patrol area. However crude the measure might be, the size of the correlations does drop dramatically.

It should be pointed out that correlations are responsive to effects other than the strength of the relationship between two variables. In particular, as McNemar (1962) points out, "the magnitude of the correlation coefficient varies with the degree of heterogeneity (with respect to the traits being correlated) of the sample" (p. 144). Given that we have divided our overall sample (into the four areas) in a manner that clearly relates to both of the variables (see Tables 1 and 2), this curtailment of the variance could be a problem. It turns out, however, not to be a serious problem in this case. The ratios of the standard deviations of the "curtailed" distribution (average of the four areas' standard deviations) to the uncurtailed distribution (standard deviation for the four areas pooled) are .965, .917, and .844 for the factor scores, total TV viewing scores, and TV violence scores, respectively. McNemar (1962) indicates that "formulas for 'correcting' for double curtailment are not too satisfactory" (p. 145). However, correcting for the curtailment of the range for the most curtailed distribution (TV violence) would only raise the average correlation between fear and amount of violent TV watched (averaged across the four areas) from .07 to .09.

The data look very much the same when analyzed question by question. The nine questions with the highest weight on the first factor of the factor analysis are shown in Table 4. In addition, the overall correlations for all subjects pooled across areas are shown in the first column (for each question with total TV viewing) and the third column (for the responses to each question and violent TV). All of the significant correlations are in the direction consistent with the Gerbner (Gerbner & Gross, 1974, 1976a, 1976b; Gerbner et al., 1976) findings (i.e., more TV associated with higher likelihood of victimization, etc.). Generally speaking, it is clear that the correlations tend to decrease substantially in size when they are run within the four areas and then pooled.

Table 5
Stepwise Multiple Regression Summary[a]

Variable	R when entered[b]	F when entered[c]	F in final equation[d]
High/low crime	.234	17.234**	16.704**
City/suburb	.235	.186	.051
Interaction: Crime × City/Suburb	.251	2.335	2.322
Sex	.350	19.858**	21.176**
Age	.376	6.605*	5.206*
Total TV	.385	2.363	.058
TV violence	.391	1.703	1.537
Radio news	.401	2.637	3.002
Newspaper reading	.403	.389	.389

[a] Variables entered in the order indicated.
[b] R achieved with this and all variables above it included.
[c] Equivalent to a test of the significance of the partial correlation between this variable and fear of crime with all variables listed above it partialed out.
[d] Equivalent to a test of the null hypothesis that the beta for this measure in the final equation involving all nine variables is zero.
* $p < .05$ ($df = 1$ and ≥ 290 for all Fs).
** $p < .01$.

An alternative method of analyzing these data is in a stepwise multiple regression analysis using the fear-of-crime factor scores as the criterion and various other social and media exposure data as predictors. In order to control for neighborhood, this was entered first into the regression equation (coded as three variables: high/low crime, city/suburb, and their interaction). Next, two subject characteristics, sex and age, were entered, since both of them related to the fear-of-crime measure. (Not surprisingly, women and older people reported higher levels of fear than did men and younger people.) After these more "basic" variables had been entered, total TV viewing and TV violence were entered. Finally, the frequency of listen-

Table 6
Standardized Canonical Variate Coefficients

Variable	High score indicates	Variate 1	Variate 2	Variate 3
Outcome set				
Question 1	No problem	.536	−.102	.339
Question 2	Low chance	.064	−.108	.266
Question 3	Low chance	.365	−.012	−.773
Question 4	Low chance	.107	−.175	.283
Question 5	Low chance	−.159	−.059	−.806
Question 6	Low chance	.394	.216	.541
Question 7	Unlikely	−.420	.131	.131
Question 8	Never	−.501	−.451	.417
Question 9	No	.028	−.577	−.281
Predictor set				
Total TV	Much	.079	.088	−.063
TV violence	Much	−.136	−.017	.335
Radio news	Little	−.177	−.194	.430
Newspaper reading	Little	−.014	−.168	.032
High/low crime area	High	−.609	.472	−.178
City/surburb	City	−.279	−.145	−.845
Interaction: Crime × Location	High suburb/ low city	.465	−.283	−.236
Age	Older	.010	.163	.096
Sex	Female	.229	.889	−.110
Canonical correlation		.608	.468	.305
p value[a]		<.001	<.001	<.002

[a] Using Wilks' lambda. Using the method of the greatest characteristic root, the third pair of canonical variates is not significant. For a discussion of this problem, see Harris (1976).

ing to radio news and newspaper reading were entered into the equation.

The results are shown in Table 5. It is clear that after the subject characteristics had been entered, the media questions had no significant predictive value. Most relevant to the Gerbner results is of course the lack of importance of total TV viewing when it first entered the equation.

Finally, a canonical correlation analysis was done, using the nine fear-of-crime questions (see Table 4) as the criterion set and the same nine variables as in the multiple regression analysis (see Table 5) as the predictor set. Three significant canonical correlations were found. The variates associated with these correlations are shown in Table 6.

The first pair of canonical variates suggests that those who do not see crimes of violence as a problem in their neighborhood (Question 1), who do not think that a child playing alone in a park is in danger (Question 3), and who do not think that they themselves are likely to be victims of an assault (Question 6), but who are afraid that their houses will be broken into (Question 7) and who do not walk alone at night (Question 8) tend to be females living in low-crime (city) areas.

The second pair of canonical variates appears to indicate that people who have areas near them that they will not walk in at night (Question 9) and who fear walking alone at night (Question 8), but who do not think that they will be victims of a violent crime (Question 6), tend to be females living in high-crime (city) areas who listen to a lot of radio news.

The third set of variates suggests that those who think that unaccompanied female subway riders (Question 5) and children playing alone in parks (Question 3) are vulnerable to attacks, but who themselves do not feel vulnerable (Question 6) and do not worry

about walking alone at night (Question 8), since their neighborhoods are safe (Question 1), tend to be suburban (low-crime area) residents who watch a lot of violent TV and do not listen to radio news.

The total amount of television watched did not seem to be important, and the amount of violent television watched entered into the interpretation only in the third canonical variate. Even in this case, it appears that the amount of violent TV watched related positively to the perceived vulnerability of particular groups (female subway riders and children playing alone in parks) but negatively to the perceived likelihood of the respondents themselves being victims of violent crime.

In summary, then, looking at these data from three somewhat different points of view, it appears that the amount of television watched did not relate to the amount of fear a person felt about being a victim of crime when other, more basic variables were taken into account.

Other Findings

As indicated earlier, we asked 25 other questions. Most of these questions related, directly or indirectly, to the respondents' views of the nature and frequency of crime or violence around them. In 14 of the questions, there was a significant relationship (pooled or calculated individually and then averaged) between TV viewing and the response to the question. The questions and the relationship of each question to TV viewing are shown in Table 7. What is noteworthy about these correlations is that there is generally not a substantial drop when the correlations are computed within area and then averaged (column 2 of Table 7). Thus, it appears that the relationship between total TV viewing and responses to these questions is not mediated by the area in which the respondent resides. These were the only other questions that correlated with TV viewing, and, at least from our point of view, they are qualitatively different from those that were large contributors to the "fear index" (see Table 4). Because it is not of central interest to this study, we did not look at other possible factors that might account for the correlations that we have presented.

Discussion

A number of things are reasonably clear from these data. First of all, the basic findings of Gerbner and his associates (Gerbner & Gross, 1974, 1976a, 1976b; Gerbner et al., 1976) are replicable: People who watch a lot of television are more likely to indicate fear of their environment. It is equally clear, however, that this relationship disappears when attempts are made to control for other variables, including the actual incidence of crime in the neighborhood. Thus, it would appear that television itself is not likely to be a direct cause of people's fear of being victims of crime.

Although clearly at the level of speculation, it is interesting to note that Gerbner's (Gerbner & Gross, 1974) own data on this issue were collected by telephone interview in four cities: Philadelphia, Chicago, Los Angeles, and Dallas. One can assume that there exists in these cities some variability in the dangerousness of different neighborhoods. Since households for that survey were selected randomly from telephone books, it seems reasonable to expect that neighborhoods differing in actual dangerousness would be included from each city. This variation could be sufficient to produce the apparently small correlation that Gerbner found. More interesting, however, is the possibility that for some unspecifiable reason, the relationship only holds in high-crime areas, or in high-crime cities in particular. As shown in Table 3, we, too, got significant correlations within the high-crime area of the city of Toronto. One possible admittedly post hoc explanation for this result is that television violence in the form of police shows and so forth deals mostly with high-crime city neighborhoods. It is possible that people outside of such areas do not feel that the violence on television has any relevance for them; hence, there is no relationship between the amount of television watched and the perception of the likelihood of being a victim.

Table 7
Questions Associated Significantly With TV Viewing: Correlations Between Responses to Each Question and Total TV and TV Violence for the Four Areas Pooled and the Average of the Four Areas Calculated Individually

Question	Total TV		TV violence		High TV viewing associated with
	Pooled	Average within area	Pooled	Average within area	
10. Would you imagine that you would be more likely to be seriously harmed by someone you knew previously or by a complete stranger? (400)	.12*	.09*	.04	.06	Previously known
12. How dangerous do you think it is for a female driver of a car to pick up a male hitchhiker who is a stranger? (404)	.09*	.06	.06	.06	Dangerous
13. Do you think that it would be a good idea to spend more money on police patrols of your area of the city? (403)	.11*	.06	.12*	.08	Definitely yes
17. Do you think that it is useful for people to keep firearms in their homes to protect themselves? (405)	.31*	.20*	.25*	.11*	Definitely yes
20. Should women carry a weapon such as a knife to protect themselves against sexual assault? (405)	.18*	.17*	.19*	.18*	Definitely yes
21. Some people have suggested that one way to reduce the incidence of violent crime is to encourage people to stay away from areas thought to be high in crime. Do you think that this is a good way of dealing with the problem of crime? (403)	.09*	.07	.09*	.06	Definitely yes
22. What proportion of murders in Toronto do you think are committed by people who could be classified as mentally ill? (381)	.11*	.10*	.08*	.09*	High proportion
23. Approximately what proportion of assaults in Toronto are directed against members of racial minorities (i.e., nonwhites) by whites? (359)	.10*	.12*	.09*	.11*	High proportion
24. What proportion of serious assaults in Toronto do you think are carried out by nonwhites? (364)	.08	.11*	.08	.12*	High proportion
25. How many murders do you think took place in metropolitan Toronto during 1975? (372)	.17*	.16*	.16*	.13*	Large number

Table 7 (*continued*)

Question	Total TV		TV violence		High TV viewing associated with
	Pooled	Average within area	Pooled	Average within area	
26. During the last 5 years, how many people do you think were murdered in the TTC subway? (384)	.15*	.12*	.09*	.08*	Large number
31. If you were walking alone on a residential street at night and someone asked you for directions, would you stop and give him the directions? (404)	.10*	.11*	0	0	Definitely not
32. If a person were to have an epileptic seizure on the street in front of you, how likely do you think most people would be to help? (405)	.10*	.07	.08	0	Very likely
33. If, in the middle of the night, a stranger knocked on your door and asked to use your telephone to call someone to help him start his car that had apparently stalled on your street, which of the following would you be most likely to do? (404)	.11*	.12*	.05	.03	Not help

Note. *ns* of respondents for each question are in parentheses.
* $p < .05$.

The second general point that should be made about our data is that although the correlation between TV viewing and fear dropped off when neighborhood was used as a controlling factor, this same factor did not eliminate the relationship between TV viewing and other factors (see Table 7). It is possible that the questions listed in Table 7 are, in fact, related to television viewing because they deal with matters of a more factual nature than the questions having to do with the person's own level of fear. Thus, television may well act as a source of information with regard to questions of fact, whereas it does not change people's views of how afraid they should be.

References

Gerbner, G., & Gross, L. *Trends in network television drama and viewer conceptions of social reality, 1967–1973. Violence profile number 6*. Washington, D.C.: Educational Resources Information Center, 1974.

Gerbner, G., & Gross, L. Living with television: The violence profile. *Journal of Communication*, 1976, *26*, 172–199. (a)

Gerbner, G., & Gross, L. The scary world of TV's heavy viewer. *Psychology Today*, April 1976, pp. 41–45; 89. (b)

Gerbner, G., Gross, L., Eleey, M. F., Fox, S., Jackson-Beeck, M., & Signorielli, N. *Trends in network television drama and viewer conceptions of social reality, 1967–1975. Violence profile number 7*. Philadelphia, Pa.: The Annenberg School of Communications, University of Pennsylvania, 1976.

Harris, R. J. The invalidity of partitioned-U tests in canonical correlation and multivariate analysis of variance. *Multivariate Behavioral Research*, 1976, *11*, 353–366.

McNemar, Q. *Psychological statistics* (3rd ed.). New York: Wiley, 1962.

Received February 24, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 2, 180–194

Retrieval Selectivity in Memory-Based Impression Judgments

John H. Lingle
Livingston College, Rutgers—
The State University

Thomas M. Ostrom
Ohio State University

Judgments about others are often based on memory for information about the persons being judged. Three studies are reported that use decision time to determine what information subjects selectively recall when they make memory-based person judgments. Each study employed a sequential judgment paradigm in which a subject first made an impression judgment about a person on one dimension while stimulus information was continuously available. Immediately thereafter, the subject made a second judgment about the same person on a different dimension without the stimulus information being available. It was concluded that subjects' memory-based judgments were based on memory for their first impression judgments combined with a selective memory search for negative stimulus information.

In recent years a great deal of impression formation research has been directed towards describing how people integrate multiple information items when they make a single person-impression judgment. Such research has dealt almost exclusively with stimulus-based judgments. These are judgments that derive from descriptive information that the experimenter provides immediately prior to, or simultaneously with, the judgment that is made. In most day-to-day situations, however, the judgments that people make about others are memory-based. That is, rather than being derived from a set of presented factual information items, the judgments are based on a selective sampling of information from a cognitive representation of the person that the perceiver has built up and stored in memory. In order to understand the majority of interpersonal judgments that people make, it is necessary to understand how (and what kind of) information is sampled from memory when memory-based interpersonal decisions are made. To examine this question, this article reports three experiments that investigated the degree to which an initial stimulus-based judgment that a person makes about a stranger influences a subsequent memory-based decision made about the same individual.

Effects of a Stimulus-Based Judgment on Subsequent Memory-Based Decisions

It is possible that the information that a person selectively samples from memory and considers when making a memory-based judgment will be similar in composition to the information available when an initial stimulus-based judgment is made, especially when both judgments occur close together in time. That is, in both cases a person may simply review the factual stimulus information (either as it is presented or from memory) and integrate it into a single response. Some amount of forgetting of the stimulus information might occur in the case of the memory-based judgment, but this would not represent an important difference in the composition of the

Portions of these experiments were reported at the annual meetings of the American Psychological Association in Chicago, September 1975, and in Washington, D.C., September 1976. The research was supported by National Science Foundation Grant GS-38604. For their help in analyzing the data and preparing the manuscript we thank David Ronis, Janet Dukerich, and Barbara Turpin.

Requests for reprints should be sent to John H. Lingle, Department of Psychology, Livingston College, Rutgers—The State University, New Brunswick, New Jersey 08903.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3702-0180$00.75

sampled information as long as the forgetting was unselective.

Alternately, there are several ways in which an initial judgment might affect the cognitive representation a person forms about another individual as well as the information likely to be sampled from memory when a later memory-based decision is made. One way would be to make available from memory new thoughts or beliefs (apart from the items of stimulus information) that could be recalled and used as the basis of a decision. Such potentially available cognitions would include memory for the initial judgment as well as any person characteristics normally associated with that judgment. It has been shown, for example, that people are sometimes able to remember their own cognitive responses better than the stimulus information on which such responses are based (Greenwald, 1968). Furthermore, research exists demonstrating that people rely on abstracted inferences and affective impressions in addition to factual information in decision tasks (cf. Posner & Snyder, 1974). Thus, if a person were to decide upon meeting someone that the new acquaintance would probably be a good lawyer, she or he might later conclude that the same person would make a good judge, simply because, being a "good lawyer type," the acquaintance would likely be intelligent and judicious. Such a judgment process would be different from, and highly dependent on, the earlier stimulus-based judgment (i.e., this person would make a good lawyer).

A second way in which an initial judgment might alter the thoughts that are processed when a later memory-based judgment is made would be to influence the likelihood that different subsets of the original stimulus information would later be recalled. There are at least two reasons why, following an initial judgment, stimulus information relevant to that judgment should be remembered better than irrelevant information. First, it has been shown that the more deeply an item of information is processed when it is first encountered, the better it will later be remembered (Craik & Lockhart, 1972; Craik & Tulving, 1975). When a person makes an initial stimulus-based judgment about someone, it seems likely that she or he will pay greatest attention to, process more deeply, and remember better information that is relevant to making that judgment. Second, if an initial judgment is later remembered, it could easily serve as a cuing mechanism and facilitate the recall of information that had been instrumental in reaching the judgment (cf. Tulving & Thomson, 1973; Tulving & Watkins, 1975). Both processes, then, would likely result in judgment-relevant information being remembered better than judgment-irrelevant information when memory-based person judgments are made.

Availability Versus Retrieval Selectivity

In considering how an initial stimulus-based judgment influences a subsequent memory-based judgment, it is necessary to distinguish between the entire set of thoughts a person might have about another and the cognitions that are selectively sampled in the process of reaching a specific decision about that person. For example, research by Geva (1977), Lingle and Ostrom (in press), and Lingle, Geva, Ostrom, Leippe, and Baumgardner (in press) has provided evidence that an initial stimulus-based judgment influences how subjects cognitively represent a described person, as measured by memory for the stimulus traits and freely generated thoughts used to describe the stimulus person. However, such descriptions and recall measures do not identify how an initial judgment affects the set of cognitions that is sampled and actually used when a later memory-based judgment is made. To study this problem, the three experiments reported in this article used decision time as a basis for making inferences about the set of cognitions subjects rely on when they make memory-based person judgments. Experiment 1 was designed to determine whether decision time is affected by the degree of similarity between a first (stimulus-based) and second (memory-based) judgment. Experiments 2 and 3 used decision time to test three alternative processing models capable of accounting for the results of Experiment 1.

Experiment 1

If an initial stimulus-based judgment influences the manner in which a person cognitively represents another, thoughts that are recalled when a later memory-based decision is required should be more relevant to a similar subsequent judgment than to a dissimilar one. As a result, subjects should be able to make similar second judgments more quickly than dissimilar ones.

Experiment 1 tested the hypothesis that an initial judgment influences the set of cognitions on which a later judgment is based by determining whether the similarity between a first and second judgment influenced how quickly the second judgment could be made. If the first judgment has no systematic effect on the cognitive representation, then decision time for the second judgment should not be affected by the similarity between the two judgments. To test the hypothesis, a within-subjects design was used in which 12 subjects made two occupational judgments about 12 different stimulus persons. Subjects made each of their second judgments without being re-exposed to the stimulus traits presented when they made their first occupational judgment. For some of the trials, the initially judged occupation was similar to the second occupation, whereas on other trials, the initial occupation was dissimilar to the second.

Method

Subjects. Subjects in the experiment included eight male and four female undergraduate students from Ohio State University who participated in partial fulfillment of an introductory psychology course requirement. Subjects were randomly assigned to one of 12 different stimulus presentation orders used in counterbalancing the design.

Procedure. Upon arriving, each subject was seated in a desk chair beside a slide projector. A toggle switch was mounted on the arm of the chair that could be moved to the right (labeled "good") or to the left (labeled "bad"). Subjects were told that as part of a study on job counseling they would be asked to role play a job placement counselor and make a series of decisions concerning the suitability of hypothetical individuals for different occupations. It was explained that at the beginning of each trial, an initial occupation would be projected on the wall (approximately 2.5 m in front of the subject) followed by a slide containing four traits describing a stimulus person. The subject was to consider the suitability of the person for the job previously shown. After the traits had been displayed for a fixed period of time, the initial occupation was again presented, and the subject was asked to register his or her decision by moving the toggle switch to the good or bad position. The stimulus traits were displayed for a fixed 20 sec in order to minimize the possibility that subjects would rush through their judgments without considering all of the stimulus information. Pilot testing indicated that 20 sec was sufficient time for subjects to read the four traits, form an initial impression, and decide whether they thought the person would be successful in the designated occupation.

After a subject registered his or her initial occupational rating, a second occupation was shown, and the subject was asked to judge the suitability of the same stimulus person for this second profession. The subject was told to register the decision for this second judgment as soon as he or she was able to make it. Finally, a blank slide was presented to indicate the end of the trial. The process was repeated with a new set of traits and occupations until each subject had made a pair of occupational judgments for 12 different stimulus persons. On each trial, the blank slide and initial occupation were each displayed for 5 sec; the slide with four traits appeared for 20 sec. The two decision slides followed one another immediately, each being changed 3 sec after the subject had indicated his or her judgment. Two practice trials (one with a similar and one with a dissimilar pair of judgments) were repeated until both were completed without a mistake, in order to ensure that subjects understood the procedure. At the end of the experiment, probes for suspicion were conducted, and subjects were debriefed as to the purpose of the research.

Second decision time was measured from the moment at which the slide change mechanism of a Kodak Carousel 800H projector was activated, displaying the stimulus slide, to the point at which the subject moved the response switch. Decision time was measured by a Hewlett-Packard 12.5-MHz electronic counter and was automatically recorded.

Design and stimulus materials. Twelve groups of four occupations were generated. Each group contained two pairs of similar occupations, each of which was dissimilar from the occupations of the other pair (e.g., store clerk/salesman – lawyer/judge). Three independent judges demonstrated 100% agreement in judging each pair of similar occupations to be more similar than any possible dissimilar pair of occupations within the same occupational group (according to the criterion that "the occupations would require individuals with similar characteristics"). Each group of four occupations was paired with four randomly selected descriptive traits chosen from Anderson's (1968) trait-adjective list.

A completely within-subjects design was used in which each of the 12 subjects judged one pair of occupations from each of the 12 occupation groups.

The design was counterbalanced so that across the 12 subjects, each occupation appeared twice with each of the three occupations in its group, one time as the initial judgment and one time as the second judgment. Consequently, over the 12 subjects, each occupation within an occupational group was used as an initial judgment three times, as a second dissimilar judgment two times, and as a second similar judgment one time. The order of presentation of similar and dissimilar occupational pairs was counterbalanced across the 12 subjects by means of a Latin square (12 subjects being the minimum number of subjects that allowed complete counterbalancing of occupations across subjects).

Results and Discussion

During the debriefing, none of the subjects was able to verbalize any part of the experimental hypothesis, nor did any subject indicate an awareness that decision time was the principal dependent variable.

A test of the experimental hypothesis depended on an examination of subjects' second decision times. To review the alternative possibilities, if subjects unselectively review the stimulus information both when they make an initial stimulus-based judgment and when they make a subsequent memory-based judgment, the similarity of the first judgment to the second should not affect how quickly the later decision is made. On the other hand, to the extent that subjects base a second judgment on a set of remembered cognitions that is influenced by an initial judgment, they should make second similar judgments more quickly than second dissimilar judgments. Consistent with this latter alternative, mean decision time was significantly shorter for similar occupational judgments than for dissimilar judgments (Ms = 5.2 versus 6.3 sec, respectively), $t_{dep}(11) = 2.78$, $p < .02$. When examined individually, 10 out of the 12 subjects spent less time on the average in making similar as opposed to dissimilar second judgments. This was significant by the sign test ($p < .04$).

As might be expected, the subjects' second decisions matched their first decisions more often when they made similar (as opposed to dissimilar) second judgments (81% versus 63%, respectively). It was conceivable, therefore, that the differences in second judgment times resulted merely from subjects consistently making two matching responses more quickly than two nonmatching responses, not from the difference in relevance of their remembered impressions to their second judgments.

To examine this alternative explanation, the point-biserial correlation between second decision time and whether a second decision matched its corresponding first decision was computed for both the similar and dissimilar second judgment conditions. In the similar conditions, there was a small positive correlation ($r = .17$) such that subjects showed a slight tendency to make faster decisions when they made two matching judgments. However, in the dissimilar judgment conditions, the correlation was slightly negative ($r = -.15$). The correlation across both experimental conditions was $-.04$. Thus, the pattern of correlations did not support the explanation that the decision-time differences resulted from subjects persistently making matching second responses more quickly than nonmatching ones.

The difference in judgment times also did not appear to result from subjects making a disproportionate number of negative responses in either of the two conditions. The percentage of negative responses was almost identical (54% for similar versus 57% for dissimilar). Furthermore, the point-biserial correlation between second-decision valence and response time was negligible (the correlations for the dissimilar, similar, and combined conditions were .00, $-.03$, and .01, respectively).

Although the results from Experiment 1 are inconsistent with the thesis that subjects based their second judgments on an unselected review of the presented stimulus traits, they do not discriminate between any of the several alternative information-processing models capable of accounting for the differences in similar and dissimilar judgment times. Experiments 2 and 3 were designed for this purpose.

Experiment 2

In the introduction, three distinct categories of remembered cognitions on which a

subject might base a second judgment about a person were identified: (a) traits originally used to describe the person, (b) a previous judgment the subject had made about the person, and (c) characteristics associated with an earlier judgment. Several models of how these three types of cognitions are used singly, and in combination with each other, would be capable of accounting for the differences in judgment times obtained in Experiment 1. In this section, three such models are described. Each model makes a different prediction about how the pattern of second judgment times observed in Experiment 1 would be affected by the number of traits (set size) used to describe the stimulus person. Experiment 2 manipulated both similarity and set size in order to test which of the models best describes how subjects process information when they make second occupational judgments. As with any such test, the three models (and their differing predictions) rest on several assumptions that are briefly identified prior to our discussing the models themselves.

The first such assumption is that the amount of time subjects require to review a set of information items in memory increases as the number of items in the set increases. This assumption is empirically well supported and has been shown to hold true in a wide range of cognitive processing tasks including set inclusion tasks (Sternberg, 1969), person-trait identification tasks (Posner & Snyder, 1974), and paragraph comprehension tasks (Kintsch, 1974).

A second assumption is that the amount of time subjects require to make a judgmental response in this task, once they have reviewed the set of cognitions on which the occupational judgment will be based, is independent of the number of cognitions that are reviewed. While this assumption is less well supported empirically than the first, there is evidence that in at least some judgment tasks, response time is independent of the amount of information that is reviewed in memory to make the judgment (cf. Shanteau & McClelland, 1975).

A final assumption of Experiment 2 (and Experiment 3) is that the amount of time required to make a judgment is a direct function of the relevance of a subject's cognitive representation to the judgment. It is assumed that the less relevant a set of remembered cognitions is to a judgment, the more time a subject will have to spend in reviewing his or her cognitive representation in order either to find judgment-relevant cognitions or to generate judgment-relevant inferences.

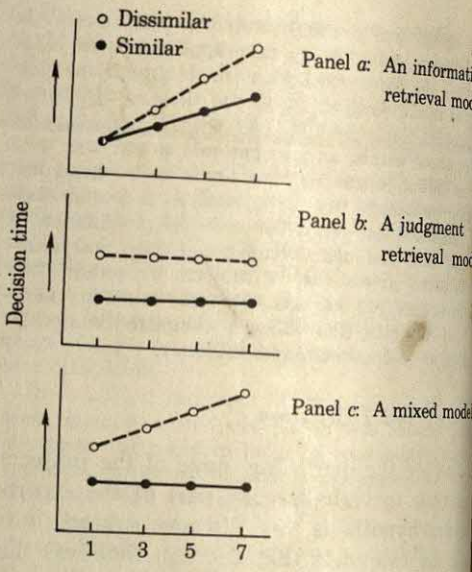


Figure 1. Predicted second judgment times for three different processing models when judgment similarity and set size are both varied.

With these assumptions as background, the three processing models tested in Experiment 2 can be described. The patterns of decision times predicted by each model when judgment similarity and trait set size are varied are displayed in Panels *a*, *b*, and *c* of Figure 1.

Three Models of Memory-Based Impression Judgments

An information-retrieval model. One model capable of accounting for the results of Experiment 1 posits that when subjects make their second occupational judgments, they serially review from memory the stimulus traits until they feel they have considered a sufficiently representative proportion of the available factual information to reach a decision. The major effect of the first judgment is to increase the probability that traits relevant

to the first decision will more readily be recalled to mind as compared with traits irrelevant to that decision. Since the amount of time required to review a set of items in memory is assumed to increase as set size increases, subjects in both the similar and dissimilar conditions should take longer to make their judgment as the amount of information in the person description increases. However, something else should also occur as set size increases. Since some percentage of the traits used to describe each stimulus person is relevant to the first judgment (e.g., 50%), the absolute number of relevant traits—and the time required to recall a representative number of these traits—will increase with set size. When subjects are making a similar second occupational judgment, these traits, being relevant to the second judgment as well, should be sufficient to make a decision. However, when making a dissimilar second judgment, subjects should not find these traits particularly helpful, since the dissimilar occupation was selected to require dissimilar abilities. In the dissimilar condition, then, subjects should have to recall the remaining traits in the set. Since the number of remaining traits will also increase with set size, subjects will have to spend additional time remembering and reviewing them. As a result, the average *difference* in decision time for similar and dissimilar second judgments should increase as set size increases. This pattern of results is depicted in Panel *a* of Figure 1.

A judgment-retrieval model. In the introduction, it was argued that an initial decision and its associated characteristics may serve as the basis for subsequent judgments. The second model focuses on retrieval of the initial judgment rather than on retrieval of the information or events on which the first impression was based. If subjects tend to base a second judgment on an initial judgment (and its associated characteristics), the number of cognitions considered in reaching a second decision should be independent of the number of stimulus traits initially used to describe a person. As a result, neither similar nor dissimilar judgment times would be expected to increase with set size. However, because the initial judgment and its related attributes would be less relevant to the second judgment in the dissimilar (as opposed to the similar) condition, subjects would be expected to spend additional time reaching a decision in the former situation. The pattern of decision times that would be expected is pictured in Panel *b* of Figure 1.

A mixed model. People are flexible information processors capable of using myriad decision strategies. It is possible that the difference in decision times in Experiment 1 came about because subjects used different decision strategies in making the two types of judgments. When subjects are faced with two very similar judgments, they may tend to base their second decision on their first, without considering the original stimulus information. However, when confronted by a dissimilar second judgment, subjects may find their first judgment insufficient grounds for the second and begin to review from memory the previously presented stimulus traits. In such a case, decision time for dissimilar judgments would increase with set size, whereas decision time for similar judgments would not. The pattern of response times would then be expected to parallel Panel *c* of Figure 1.

To test which of the three models best accounts for the similarity effect of Experiment 1, two sequential replications of Experiment 2 were conducted. In Replication 1 homogeneous trait sets, comprised of either all positive or all negative traits, were used to describe the stimulus persons. In Replication 2, heterogeneous trait sets, consisting of both positive and negative characteristics, were employed. The purpose of the two replications was to extend the generalizability of the results to more than one configuration of person information. Conceivably, the decision strategy adopted by a subject is affected by the degree of heterogeneity in the information set. For example, people may be more prone to engage in information retrieval than judgment retrieval with heterogeneous (as opposed to homogeneous) stimulus ensembles.

Both replications in this study used a within-subjects design in which subjects made pairs of occupational judgments about 12 different stimulus persons. As in Experiment

1, subjects made their second judgments without access to the stimulus traits. Half of these second judgments were similar to a first judgment and half dissimilar; an equal number of the 12 stimulus persons each subject evaluated were described by two, four, and six traits.

Method

Subjects. Subjects were 48 undergraduates from Ohio State University who participated in the experiment in partial fulfillment of an introductory psychology course requirement. In Replication 1, 11 male and 13 female subjects participated; in Replication 2, the subjects were 14 males and 10 females. As in Experiment 1, subjects were randomly assigned to one of the counterbalancing presentation orders.

Procedure. The experimental procedure and instructions to the subjects were identical to Experiment 1. Subjects were first shown an initial occupation, followed by a set of stimulus traits that was displayed for 20 sec. Thereafter, they were asked to register their decision concerning the described person's suitability for the initially displayed occupation, as well as for a second occupation that was either similar or dissimilar to the first. As in Experiment 1, subjects were given a fixed amount of time to study the traits and reach their initial occupational decision, in order to maximize the probability that they would take time to consider all of the stimulus traits.

In Experiment 2, the number of practice trials was increased to three in order to include stimulus persons described by each of the three set sizes (two, four, and six traits). Subjects made similar occupational judgments in one of the practice trials and dissimilar judgments in the other two.

Design and stimulus materials. Twelve groups of three occupations were selected from the occupations used in Experiment 1. Each of these triads consisted of one occupation arbitrarily selected to be used as a second occupational judgment and two to be used for the first judgment, one similar and one dissimilar to the second occupation.

In each replication, 24 subjects made two sequential judgments about 12 stimulus persons. The second judgment was preceded by a similar first occupational judgment for half of these stimulus persons and by a dissimilar first occupational judgment for the other half. Likewise, one third of the stimulus persons judged were described by two traits, one third by four traits, and one third by six traits, thus providing two within-subjects replications of the 2×3 design (Type of First Judgment $\times$ Set Size). All subjects judged the 12 second occupations in an identical order. Each subject was matched with another who saw identical sets of stimulus materials, except that if one subject judged the similar first occupation in a triad, the second subject judged the dissimilar occupation. These subject pairs then formed the unit of counterbalancing in the rest of the design.

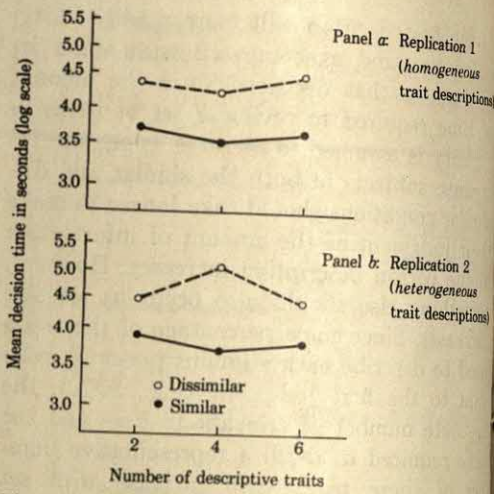


Figure 2. Mean second occupational judgment times from Experiment 2 as a function of judgment-pair similarity and set size.

For each of the two sequential replications, four lists of 12 traits were selected from Anderson's (1968) trait-adjective list. In Replication 1, two of the lists were selected from the positive half of the scale and two from the negative half of the scale, providing evaluatively homogeneous sets of traits. In Replication 2, traits for all four lists were selected from the middle three fifths of the scale, providing heterogeneous descriptions containing both positive and negative adjectives.

Counterbalancing of the traits across judgment similarity and set size for the 12 subject pairs in each replication was achieved by the method of cyclical replications. Traits within each list were randomly ordered from 1 to 12. The traits in each list were then sorted into the three experimental set sizes in 12 unique ways by simply moving the first trait to the end of the list and grouping the remaining traits in order into sets of two, four, and six. For example, (1, 2), (3, 4, 5, 6), and (7, 8, 9, 10, 11, 12) was the first group of sets; (2, 3), (4, 5, 6, 7), and (8, 9, 10, 11, 12, 1) was the second, and so forth. Three person descriptions were therefore obtained from each of the four lists to provide the 12 trait sets needed for each subject. The 144 stimulus persons generated in this manner were counterbalanced so that each trait appeared equally often for each level of set size and similarity. This counterbalancing scheme assured that none of the traits was presented to a subject more than, or less than, once.

Results and Discussion

Replication 1. As in Experiment 1, a test of the three models depended on a comparison

of subjects' second, memory-based judgment times. Natural log transformations of these response times were performed prior to the data analyses in order to eliminate a strong correlation between the cell means and variances (cf. Winer, 1971, p. 400). Analyses of subjects' untransformed judgment times were also conducted and produced results similar to the analyses of transformed scores. None of the statistical tests for either Experiment 2 or Experiment 3 that are appropriate for discriminating between the several models differed in significance as a function of whether the raw or transformed scores were used in the analysis. Subjects' retransformed mean decision-time scores for Replication 1 are presented in Panel *a* of Figure 2. The scores were analyzed using the multivariate analysis of variance approach for within-subjects designs as discussed by Poor (1973).[1]

The first replication involved homogeneous trait descriptions. The analysis of subjects' transformed second judgment times yielded a significant effect for similarity, $F(1, 23) = 14.7$, $p < .001$, replicating the principal finding of Experiment 1. However, no significant main effect or interaction for set size emerged (both Fs < 1).

Replication 2. Subjects' scores from Replication 2 (heterogeneous trait descriptions) were analyzed in an identical manner to the scores from Replication 1. The retransformed means are displayed in Panel *b* of Figure 2. The analysis again produced a significant main effect for the similarity of the second judgment, $F(1, 23) = 30.8$, $p < .001$, and a nonsignificant main effect for set size ($F < 1$). The interaction term approached significance, $F(2, 22) = 2.88$, $p < .10$. This borderline interaction was due entirely to the quadratic component of the set-size factor, $F(1, 23) = 6.02$, $p < .03$. The interaction term for the linear trend produced an F of less than 1.

While the results of both replications were most consistent with a judgment-retrieval model, two anomalies emerged in the pattern of results when heterogeneous, as opposed to homogeneous, trait sets were employed. First, in Replication 2 a quadratic interaction was obtained, raising the possibility that a more complicated version of the mixed model might be required to explain the manner in which subjects processed heterogeneous information sets. Second, subjects took slightly longer to make a second decision when heterogeneous, as opposed to homogeneous, trait sets were used to describe the stimulus person (Ms = 4.22 versus 3.99, respectively). A clear interpretation of the source of these differences was impossible, since the two replications were conducted during different months in the academic quarter, using different subject populations. A third experiment, to be reported below, was therefore conducted in which heterogeneity of the trait sets was manipulated as a separate within-subjects factor.

First judgments. As discussed earlier, the test of the three models involved the assumption that the time required to review items in memory increases as set size increases. In Experiment 2, subjects' first judgments were based on sets of differing sizes. However, the design of this (and the following) study precluded using first-decision times to verify this assumption. As previously described, in both studies subjects were first shown the occupation and then given a fixed amount of time to consider the traits, regardless of set size. Subjects were instructed to reach their judgment during the time the stimulus items were exposed and to report that decision when the traits were removed and the first decision reshown. Because there was no reason for subjects to review from memory the traits when reporting their first judgment, decision time for first judgments was not recorded.

Experiment 3

Experiment 3 was similar to Experiment 2 except for two changes. First, a wider range

[1] This approach analyzes within-subjects factors by defining a set of new variables that incorporates the elements of the within-subjects effects. An analysis of variance (sometimes univariate, sometimes multivariate) is then performed on the transformed variables. The principal advantage of the multivariate approach is that it is based on a less restrictive set of assumptions than the traditional within-subjects analysis. In the present report, boldfaced F values are multivariate values using Wilks's Lambda criterion.

of set sizes was used in order to insure that the findings of Experiment 2 were not the result of a restricted range. The set sizes used were one, three, five, and seven traits. The inclusion of a condition in which the stimulus person was described by a single trait also provided an additional test between the models. Because Experiment 3 was designed in such a way that no systematic difference existed in the relevance of the single trait descriptions to subjects' second judgments, the information-retrieval model (in which subjects rely solely on remembered stimulus traits) would not predict any difference in decision times for similar and dissimilar judgments. On the other hand, the judgment-retrieval model would predict a longer decision time for subjects' dissimilar second judgments even with only one trait, since a subject's first judgment would differ in relevance to his or her second in the two conditions (see Figure 1).

A second change incorporated into Experiment 3 was that the heterogeneity of the person descriptions was included as a separate within-subjects factor. This was done because of the unexpected quadratic interaction with heterogeneous sets in Experiment 2. In Experiment 3, each subject judged an equal number of stimulus persons who were described by a homogeneous trait set (either all positive or all negative) and a heterogeneous trait set (including both positive and negative traits).

Method

Subjects. Subjects included 11 male and 21 female undergraduates from Ohio State University who participated in partial fulfillment of a course requirement. These subjects were randomly assigned to the 32 stimulus presentation orders used to counterbalance the design.

Procedure. The experimental procedure and instructions to the subjects were similar to those used in Experiments 1 and 2. The only difference was that in this case, subjects were required to complete four practice trials, which included stimulus persons described by each of the four set sizes. For two of these practice trials, subjects judged pairs of similar occupations; for the other two, they judged dissimilar pairs.

Design and stimulus materials. The basic design and stimulus materials of Experiment 3 were similar to Experiment 2. However, because four set sizes were employed, the counterbalancing scheme required that 32 subjects (16 pairs) be used, each of whom made pairs of occupational judgments about 16 different stimulus persons. In order to include composition of trait sets as a within-subjects factor, 8 of the 16 stimulus persons judged by each subject were described by traits selected randomly from the complete range of Anderson's (1968) list, thus producing heterogeneous descriptions. Of the remaining eight descriptions seen by each subject, half were comprised of traits selected from the positive portion of Anderson's list and half from the negative portion, thus providing evaluatively homogeneous descriptions. As in Experiment 2, the counterbalancing of traits across similarity and set size was achieved by cyclical replications. In the case of Experiment 3, each list consisted of 16 traits so that set sizes of one, three, five and seven could be generated (i.e., $1 + 3 + 5 + 7 = 16$).

Results and Discussion

Effects of trait heterogeneity. One purpose of Experiment 3 was to determine if the differences in subjects' pattern of judgment times for heterogeneous and homogeneous trait sets observed in Experiment 2 were reliable. The first question investigated was whether the quadratic interaction between judgment similarity and set size was replicated for the heterogeneous trait sets. This was examined in several ways. First, the design of Experiment 3 permitted a statistical test of whether the trait-set composition (homogeneous versus heterogeneous) affected the shape of the Similarity × Set Size interaction. This test (the three-way interaction between judgment similarity, set size, and trait-set composition) proved to be nonsignificant, $F(3, 29) = 1.96$, $p < .20$. An additional analysis was performed on the heterogeneous data alone, and neither the overall interaction between similarity and set size, $F(3, 29) = 1.78$, $p < .20$, nor its quadratic component, $F(1, 31) = 1.02$, $p < .40$, was significant.

Finally, subjects' mean judgment times for heterogeneous trait sets were plotted and visually examined. This examination showed that there was no tendency for the similar and dissimilar judgment times to reconverge between Set Sizes 5 and 7 as they had reconverged between Set Sizes 4 and 6 in Experiment 2. In sum, there was no evidence that the quadratic interaction found in Ex-

periment 2 (Replication 2) was replicated in Experiment 3.

The second question examined was whether the tendency found in Experiment 2 for subjects to take longer with heterogeneous than with homogeneous trait sets was replicable. This was not the case. The main effect for trait-set heterogeneity on subjects' second judgment times was nonsignificant ($F < 1$).

Test of the three models. Because trait-set heterogeneity had no detectable effect on subjects' second judgment times, this factor is not discussed further. In order to test the three models, the data were analyzed in a similar manner to the data of Experiment 2. Subjects' retransformed response times, combined across trait-set replications, are presented in Figure 3. As in both Experiments 1 and 2, the judgment similarity main effect was significant, $F(1, 31) = 31.9$, $p < .001$. Neither the set-size effects nor the Set Size × Judgment Similarity interaction reached significance, $F(3, 29) = 2.27$, $p < .15$, and $F < 1$, respectively. Thus, even with an increased range of set sizes, the pattern of results obtained in Experiment 2 was replicated and best supported the judgment-retrieval model.

This model was further supported (and the information-retrieval model infirmed) by the fact that even when subjects made judgments on the basis of only one trait, 75% of them took longer on the average to make a second dissimilar judgment than a similar one (sign test, $p < .004$). It therefore would appear that even subjects in the single-trait condition were recalling cognitions that differed in relevance to their second occupational judgments.

Combining Experiments 2 and 3. Although no statistical statement can be made about the veracity of the null hypothesis regarding the absence of a set-size main effect or interaction with similarity, it is possible to aggregate the findings across experiments in a way that allows all the relevant available data to be brought to bear on a particular hypothesis. Winer (1971, p. 49) outlines a procedure whereby individual significance tests from a series of experiments may be combined to test a single hypothesis. If different samples

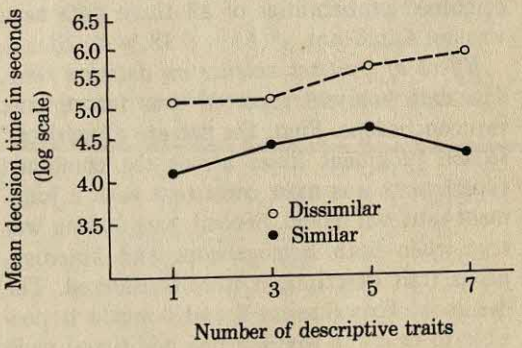


Figure 3. Mean second occupational judgment times from Experiment 3 as a function of judgment-pair similarity and set size.

are employed in each experiment, an overall probability statement can be computed by aggregating the natural logarithms of the independent probability values and comparing this aggregation with the chi-square distribution. In the present case, such a test can be made even more sensitive by considering just the linear component of the set-size effect, since it is the component that differentiates between the three models being considered.

The first test performed on the combined experimental data was a test of the linear interaction between judgment similarity and set size. Such an interaction is predicted by both the information-retrieval and mixed-judgment models, but not by the judgment-retrieval model. The F values of this linear interaction for all three data sets (cf. Figures 2 and 3) were less than 1, and the significance test based on the combined probabilities was also nonsignificant, $\chi^2(6) = 3.46$, $p < .50$.

The judgment-retrieval model predicts not only the absence of an interaction between judgment type and set size but also the absence of a linear increase in decision time across set size. A test of this hypothesis was conducted by collapsing across judgment type and analyzing the linear trend of decision times over set size. Whereas for both replications of Experiment 2 this analysis yielded Fs of less than 1, the corresponding analysis of Experiment 3 yielded a significant result, $F(1, 31) = 6.25$, $p < .02$. However, a test of the hypothesis that subjects' judgment times increased with set size, based on the

combined probabilities of all three data sets, was not significant, $\chi^2(6) = 9.38$, $p < .20$.

Effect of trait-set valence on decision time. The data analyses reported thus far support two conclusions. First, the pattern of subjects' second judgment times across the combined experiments was most consistent with a judgment-retrieval model. Second, this finding was true when both homogeneous and heterogeneous trait descriptions were considered. The design of Experiments 2 and 3 made it possible to examine whether one additional variable, trait-description valence, might limit the generalizability of the experimental results. This was possible because in both Experiments 2 and 3, to avoid a valence expectancy set, half of the homogeneous trait descriptions were constructed of positive traits and half were constructed of negative traits.

Two separate analyses indicated that the absence of an interaction between judgment similarity and set size held for both positive and negative trait sets. First, the three-way interaction between trait-set valence and the other two variables was not significant: The analysis of the linear component of the three-way interaction, based on the combined probabilities from Experiments 2 and 3, yielded a value of $\chi^2(4) = 7.81$, $p < .10$. Second, the two-way interactions between judgment similarity and the set-size linear components were nonsignificant when subjects' judgment times for positive and negative trait sets were analyzed independently: The analyses of this interaction, based on the combined probabilities from the two experiments, yielded values of $\chi^2(4) = 4.72$, $p < .50$, for the positive trait sets, and $\chi^2(4) = 4.68$, $p < .50$, for the negative trait sets.

The strength of the similarity effect also was not significantly affected by trait valence. For both types of traits, subjects took longer to make second dissimilar, as opposed to similar, judgments. This result was indicated by the absence of a significant interaction between trait valence and judgment similarity; the analysis based on the appropriate combined probabilities yielded a value of $\chi^2(4) = 2.19$, $p < .50$.

Finally, the question of whether trait valence in any way modified the finding that decision time is independent of set size was tested by analyzing the linear interaction between trait-set valence and set size. In contrast to the other analyses, the linear interaction between trait-set valence and set size was significant for both Experiments 2 and 3, $F(1, 11) = 8.20$, $p < .02$, and $F(1, 15) = 7.56$, $p < .02$, respectively,[2] as well as significant for the combined analysis, $\chi^2(4) = 16.80$, $p < .005$.

To determine the nature of this interaction, separate analyses of the linear trends of judgment times across set size for the positive and negative trait sets were conducted. These analyses indicated that when subjects considered negative trait sets, the linear component of their second judgment times did not vary significantly across set size, $F(1, 11) = 1.14$, $p < .31$, for Experiment 2; $F(1, 15) = 1.38$, $p < .26$, for Experiment 3; $\chi^2(4) = 5.06$, $p < .25$, for the combined probabilities. On the other hand, when subjects considered trait sets that consisted solely of positive traits, there was a significant linear increase in their second judgment times as the number of traits used to describe the stimulus person increased, $F(1, 11) = 4.99$, $p < .05$, for Experiment 2; $F(1, 15) = 10.87$, $p < .005$, for Experiment 3; $\chi^2(4) = 16.70$, $p < .005$, for the combined probabilities. This effect is shown in Table 1, where subjects' retransformed mean decision times for positive and negative trait sets from Experiments 2 and 3 are displayed. As can be seen, there was a tendency for subjects' judgment times to decrease as set size increased for negative trait sets. This tendency approached in magnitude the increase in judgment time over set size for positive traits. The linear trend was sig-

---

[2] For design efficiency, both Experiments 2 and 3 contained a confounding between trait-set valence and judgment similarity. However, this confounding did not exist if pairs of subjects were considered as the unit of analysis and the three factors (judgment similarity, trait valence, and set size) were considered as within-(pairs of) subjects factors. It is for this reason that the degrees of freedom for the present analyses are halved. Such an analysis is possible using a multivariate approach, since variance-covariance homogeneity is not a required assumption (see Poor, 1973, p. 206).

Table 1
Mean Second Occupational Judgment Times as a Function of Trait-Set Valence and Set Size: Experiments 2 and 3

Valence	No. of descriptive traits		
	Experiment 2		
	2	4	6
Positive	3.87	4.11	4.42
Negative	4.28	3.64	3.77
	Experiment 3		
	1	3	5	7
Positive	4.28	4.39	5.32	5.09
Negative	5.33	4.94	4.65	4.58

Note. Response times are subjects' mean log scores transformed into seconds. For Experiment 2, each mean is computed from 48 responses (24 subjects providing two responses each); for Experiment 3, each mean represents 32 responses (32 subjects providing one response each).

nificant for positive but not for negative traits because the mean square error of the linear trend analysis for negative traits was more than three times as large as the mean square error for positive traits in both experiments. This implies substantially greater individual differences in the effect of set size for negative information than for positive information.

The finding that subjects' second judgment times increased more with set size when they considered homogeneously positive (compared with negative) trait sets would not be predicted by any of the three processing models. One possible explanation of this unexpected finding is that subjects, in addition to recalling their initial judgment, spent some time searching their memory for stimulus information that might disqualify a person from being successful at the occupation. For negative homogeneous sets, the ease (or speed) with which a subject could recall a sufficient number of relevant negative information items to disqualify a person from being successful in a particular occupation could be expected to increase as set size increased. However, when considering sets of homogeneously positive traits, as set size increased, subjects would have to search an increasing number of information items before proportionally sampling the stimulus items for possible disqualifying attributes.

If such a memory search strategy did account for the obtained interaction between negative and positive information sets and set size, it would appear that it supplemented (rather than replaced) a judgment-retrieval process. The absence of an interaction between trait valence and judgment similarity (along with the absence of a three-way interaction between these two factors and set size) indicates that subjects consistently took longer to make dissimilar, as compared with similar, second judgments when the stimulus persons were described by both positive and negative trait sets.

Availability of stimulus information in memory. The primary concern of all three studies was to uncover the kinds of information people selectively review in memory prior to making a person judgment. As noted in the introduction, this concern is distinct from assessing the kinds of stimulus information that can potentially be remembered by a person. Although the data of Experiments 2 and 3 suggest that people do not actively recall more stimulus items as a function of set size while making a judgment (at least not for homogeneously negative or heterogeneous stimulus sets), this should not be taken to mean that people are incapable of remembering more stimulus items when a person description contains more items of information. In Experiment 3, after subjects had made

their last occupational judgment, they were unexpectedly asked to list as many of the traits describing the final stimulus person as they could remember. Subjects were able to recall an increasing number of traits as the number of traits in the stimulus set increased. The mean number of accurately recalled traits, when subjects had seen descriptions containing one, three, five, or seven traits was 1.0, 2.6, 2.9, and 3.2, respectively, $F(3, 12) = 3.34$, $p < .06$. Thus, subjects were capable of recalling more traits as set size increased, even though there was little evidence that they were doing so while making their judgments.[3] Although the information-retrieval model did not adequately account for the pattern of response times found in the present decision setting, this model may well be appropriate in other contexts. The recall data show that people can, if required, recall more information items as set size increases in this type of sequential judgment task.

General Discussion

Summary and Conclusions

The research reported in this article was undertaken to investigate the selective retrieval process subjects employ when making memory-based judgments. The first study found that an initial stimulus-based judgment made about a person influenced how quickly subjects made a subsequent memory-based judgment about the same individual. The fact that a similar second decision was made more quickly than a dissimilar second decision indicated that subjects' memory-based judgments were not based solely on an unselective review of the stimulus information they had received.

Experiments 2 and 3 compared three different decision models that could account for the judgment similarity effect discovered in Experiment 1. Two stimulus-information-retrieval models were described that accounted for the similarity effect by postulating that selective sampling of the stimulus traits occurred when subjects made their second judgment. Neither of these models provided an adequate account of the decision-time differences. In fact, the most striking outcome of Experiments 2 and 3 was how impervious the similarity effect was to manipulations of the stimulus information. The magnitude of the similarity effect was undiminished by trait-set composition (heterogeneous versus homogeneous), trait-set valence (positive versus negative), and set size (one through seven items of person information). Even when subjects demonstrated 100% recall ability (as in Experiment 3, where only one stimulus trait was presented), 75% of them took longer to reach a second dissimilar (as opposed to similar) decision. The persistence and stability of the similarity effect proved most consistent with a judgment-retrieval model, which postulates that subjects base their second decisions on memory for their initial judgment and its associated characteristics.

Although stimulus-information-retrieval processes were not responsible for the judgment similarity effect, it cannot be concluded that such processes were totally absent in the present judgment task. There was evidence that subjects' judgment processes included a selective search for negative stimulus information in addition to recall of their initial judgment. When subjects were presented with a homogeneously positive set of descriptive traits, their mean judgment times increased as set size increased. However, when only negative traits were presented, there was a nonsignificant tendency for decision time to decrease as set size increased.

While subjects in the present decision task appeared to rely on their first impression judgments in combination with a memory search for disqualifying attributes, there is little doubt that their cognitive processing strategy for making memory-based decisions would be different in other settings. For example, had there been a substantial period of time be-

---

[3] The only other significant result from the analysis of the memory data was a significant effect for judgment type, $F(1, 12) = 4.84$, $p < .05$. Subjects in the dissimilar judgment condition recalled on the average more traits than subjects from the similar condition (Ms = 2.73 versus 2.08, respectively). However, this effect must be regarded with caution, since no such difference was found in a comparable person-impression sequential judgment task (cf. Lingle & Ostrom, in press).

tween the first impression judgment and the memory-based judgment, it would be much less likely that subjects could have remembered any specific item of descriptive information. Under such circumstances, they might have resorted solely to a judgment-retrieval process. Alternatively, a person might depend much more heavily on an information-retrieval process if there were sufficient incentive. This might be the case, for example, if the subjects were rewarded for a correct decision, were required to justify their second judgment, or perceived that the experimenter's hypothesis concerned memory for the traits.

Methodological Note

To understand the way in which people draw on memory when they make impression judgments, social psychologists will need to develop and adopt methodologies appropriate for investigating cognitive information processing. A number of investigators (cf. Anderson, 1977; Anderson & Hastie, 1974) have shown how decision time may be used to infer the structure of people's cognitive representations of others. The present studies demonstrate the way in which decision time may also be used to make inferences about how people draw on their memory in a person judgment task.

Although decision time represents a powerful, relatively unobtrusive measure of cognitive processing, there are problems associated with its use. Decision-time data are difficult to interpret without precise processing models that make clear alternative predictions. Such models invariably rest on multiple assumptions whose number tends to increase as the complexity of the decision being studied increases. Consequently, decision time is much more suitable for eliminating processing models than it is for providing definitive evidence that any single model is a true representation of the way in which people process information. Thus, the present studies provide relatively strong evidence that neither the information-retrieval model nor the mixed model, as described in this article, accurately reflects subjects' judgment process in this task. Other versions of these models, however, might account for the data better. For example, it may have been that in the present judgment context, subjects felt time-limited, as opposed to relevance-limited, in their memory search. That is, the lack of relationship between set size and decision time may have resulted from subjects tending to sample cognitions from memory for a fixed period of time rather than sampling a proportional number of cognitions. Although the interaction between trait valence and set size suggests that subjects were not tightly time-bound in their processing strategy, the possibility of this type of limitation cannot be ruled out. Consequently, the present support for judgment-retrieval processes must be regarded as tentative.

The existence of alternative explanations for the present findings reflects the fact that model building is an iterative process in which some models are eliminated while others are expanded and refined to conform with empirical regularities that have emerged from previous testing. We hope that the present studies represent only a first step in an evolutionary process that will result in the development of more complete models of the ways in which people selectively draw on memory when they make memory-based impression judgments.

References

Anderson, J. R. Memory for information about individuals. *Memory & Cognition,* 1977, *5,* 430–442.

Anderson, J. R., & Hastie, R. Individuation and reference in memory: Proper names and definite descriptions. *Cognitive Psychology,* 1974, *6,* 495–514.

Anderson, N. H. Likableness ratings of 555 personality words. *Journal of Personality and Social Psychology,* 1968, *9,* 272–279.

Craik, F. I. M., & Lockhart, R. S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior,* 1972, *11,* 671–684.

Craik, F. I. M., & Tulving, E. Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General,* 1975, *104,* 268–294.

Geva, N. *The role of memory in person perception.* Unpublished doctoral dissertation, Ohio State University, 1977.

Greenwald, A. G. Cognitive learning, cognitive response to persuasion, and attitude change. In A. Greenwald, T. Brock, & T. Ostrom (Eds.), *Psychological foundations of attitudes.* New York: Academic Press, 1968.

Kintsch, W. *The representation of meaning in memory*. Hillsdale, N.J.: Erlbaum, 1974.

Lingle, J. H., Geva, N., Ostrom, T. M., Leippe, M. R., & Baumgardner, M. H. Thematic effects of person judgments on impression organization. *Journal of Personality and Social Psychology*, in press.

Lingle, J. H., & Ostrom, T. M. Principles of memory and cognition in attitude formation. In R. Petty, T. Ostrom, & T. Brock (Eds.), *Cognitive responses to persuasion*. Hillsdale, N.J.: Erlbaum, in press.

Poor, D. S. Analysis of variance for repeated measures designs: Two approaches. *Psychological Bulletin*, 1973, *80*, 204–209.

Posner, M. I., & Snyder, C. R. R. Attention and cognitive control. In R. Solso (Ed.), *Information processing and cognition*. Hillsdale, N.J.: Erlbaum, 1974.

Shanteau, J., & McClelland, G. H. Mental search processes in problem solving. *Memory & Cognition*, 1975, *3*, 627–634.

Sternberg, S. Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 1969, *57*, 421–457.

Tulving, E., & Thomson, D. M. Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 1973, *80*, 352–373.

Tulving, E., & Watkins, M. J. Structure of memory traces. *Psychological Review*, 1975, *82*, 261–275.

Winer, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1971.

Received September 27, 1977

Notice on Author Alterations

Effective January 1979, the procedure for billing authors for alterations will change. Authors of articles in APA journals are customarily charged for alterations on proofs that result from (a) authors' errors or omissions in the manuscript and (b) changes that result from authors' failure to review the edited manuscript or the proof. Until now, alteration charges have been determined by the number of lines affected by changes requested and have been based on 1974 printing rates. Beginning with the proofs for the first issues of 1979, authors' alterations will be determined not only by the number of lines affected but also by the number of figures and pages affected and will be based on current printers' rates.

This change in alteration billing is being made so that alteration charges will more closely reflect the actual cost of changes requested and may mean larger alteration bills for some authors. Authors are reminded that alteration charges apply only when changes are made on proofs and that *there is no charge* for changes made at the manuscript stage.

Journal of Personality and Social Psychology
1979, Vol. 37, No. 2, 195–202

Personality Correlates of Susceptibility to Biasing Information

Elisha Y. Babad
School of Education
Hebrew University of Jerusalem
Jerusalem, Israel

Subjects were asked to score two drawings, one allegedly made by a high-status child and the other by a disadvantaged child. The difference between the scores that each subject attributed to the two drawings (compared to the objective base level) was the subject's bias score. High-bias subjects (with bias scores one standard deviation or more above the overall mean bias) described themselves as more reasonable than no-bias subjects (bias scores of zero), and as less emotionally extreme in the directions of toughness, tenderness, and level of enthusiasm. This pattern was interpreted as reflecting the need system suggestive of the dogmatic personality. In a second study, no differences were found between the extreme-bias groups in responses to self-report questionnaires and the Embedded Figures Test, and in political views. However, high-bias subjects responded *more extremely* than no-bias subjects to direct questions pertaining to their political ideology, independent of its content and direction, in line with the conception of dogmatism as a style of thinking independent of ideological content.

Experimenter bias effects, expectancy effects, and self-fulfilling prophecies of various types have been repeatedly demonstrated in a variety of populations, settings, content areas, and types of experiments (Rosenthal, 1966, 1969, 1971). In such studies, changes in dependent variables (mostly task performances) are attributed to the influence of biasing information, which is presumably task irrelevant. However, expectancy effects are *not* universally demonstrated in every study and every replication, and even when the overall effects are statistically significant, there are great variations in susceptibility to biasing information (Rosenthal & Rubin, in press).

I wish to thank my friends Daniel J. Davis for his helpful support, Bruce T. Oppenheimer for his continued participation and suggestions, and Chaim Schlezinger for his patient assistance in data analysis.

Requests for reprints should be sent to Elisha Y. Babad, School of Education, Hebrew University of Jerusalem, Jerusalem, Israel.

Variations in the influence of biasing information are probably caused by various factors, including design and procedure differences, nuances of delivery, and personality differences among subjects. This study focused on the latter factor, and an attempt was made to identify systematic personality correlates of susceptibility to biasing information.

McFall and Schenkein (1970) divided their subjects into extreme groups on several variables (need for achievement, field dependence). Then they divided each extreme group into two experimental groups and administered Rosenthal's (1966) photo-rating task to these groups, feeding one experimental group with "positive" expectancy and the other with "negative" expectancy. Laszlo and Rosenthal (1970) used a similar design, with dogmatism scores as the criterion for dividing the groups.

In a different design, each subject receives both positive and negative expectancy information, and the resulting difference between the subject's two performances serves as an

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3702-0195$00.75

index of susceptibility to bias. This design seems to be more powerful than the McFall and Schenkein design, since *each subject's* influenceability can be directly assessed. On the other hand, the disadvantage of this design is that the credibility of the cover story and of the biasing information (enough of a problem even when subjects receive only one type of biasing information) may be seriously hampered, limiting the range of experimental possibilities. Laszlo and Rosenthal (1970) used the first design and later created a measure of individual influenceability by computing the difference between each subject's mean photo rating and the mean photo rating obtained by the same experimenter from all his subjects contacted under the opposite condition of expectancy. The experimental situation chosen for the present study—test scoring—makes it possible to provide each subject with positive *and* negative expectancy information without losing credibility. In this design, subjects are asked to score mental tests of two "examinees," one allegedly of high status and the other of low status.

This type of study measures *suggestibility*, namely, the influence of the information on the performance of the receiver. Another possible type of study measures *communicability*, that is, the extent to which the receiver of the biased information influences the behavior of other people (Finn, 1972). Bias in scoring (e.g., Babad, Mann, & Mar-Hayim, 1975) exemplifies suggestibility effects, whereas teacher expectancy effects (Rosenthal & Jacobson, 1968) and photo-rating bias (Rosenthal, 1966) are examples of communicability effects. Suggestibility is a necessary, but not sufficient, component for communicability effects. It is interesting to note that McFall and Schenkein (1970) used *tape recordings* of experimenters in Rosenthal's photo-rating task, thus turning a communicability study into a suggestibility study.

In the present project, the Goodenough–Harris Draw-A-Person (DAP) Test (Harris, 1963) was used to determine subjects' susceptibility to biasing information. The advantages of the DAP test are its simplicity, the minimal experience it requires, and the short time it takes to learn the scoring system. While the first study in this report was exploratory, the second study was designed to test hypotheses formulated on the basis of the results of the first study.

Study 1

Method

Overview of the Study

Introductory psychology students were taught the scoring procedures of the Draw-A-Person Test and were then asked to score two drawings. The drawings were actually duplicated from the test manual, but the students were provided with falsified information alleging that one drawing was made by a disadvantaged, low-class child and the other, by a high-status, advantaged child. Later, the students completed a self-report personality inventory.

Subjects

Subjects were 133 students in introductory psychology classes at the School of Education of the Hebrew University of Jerusalem. All subjects were freshmen.

Procedure

The experiment was conducted in one (2-hour) session, the purpose of which (the students were told) was "to familiarize you with simple instruments of psychological evaluation via direct, first-hand experience." The instructor presented the Draw-A-Person Test and distributed a sheet summarizing the 73 scoring categories. The procedures for scoring were then taught, and the subjects practiced (a) by scoring a drawing on the board and (b) by scoring their neighbors' drawings. A practice sheet was then distributed, and the students were told that a reliability study would be conducted. The sheet was divided into two columns, each presenting a drawing of a man and information about the "child" who allegedly drew that picture. The information consisted of name, address, school, number of siblings, and parents' occupations and education. One child was described as coming from a high-status background (with a European name, Rubinstein), whereas the other child was depicted as a disadvantaged child of Moroccan origin (named Amsalem).

The drawings were actually duplicated from the DAP manual, so that their objective manual scores were initially known. To increase the credibility of the cover story, the drawing selected for the high-status child (Harris, 1963, p. 268, Figure 64) had a somewhat higher manual score than the drawing attributed to the low-status child (Harris, 1963, p. 271, Figure 72). Therefore, a difference of 3 points in favor of the high-status child represented "no bias."

(To double check, we asked at a later time five graduate students to score these drawings without adding the expectancy information. Their scores confirmed the 3-point difference between the two drawings.)

The instructor explained to the students that these drawings were photocopied from records collected in a recent study conducted in various Jerusalem schools. The instructor commented that "one can immediately see the difference between the drawings of Rubinstein and Amsalem" and proceeded to urge the students to be objective in their scoring, reminding them that the reliability of their scoring would be examined.

After the completion of the scoring, the Adjective Check List (ACL; Gough & Heilbrun, 1965) was distributed. The ACL consists of 300 adjectives, and subjects are asked to check those adjectives that describe them as they are (i.e., "actual-self" image). The ACL was chosen because of its nonpathological focus, the short time it takes to administer, and the availability of a Hebrew version.

Results

Overall Bias

The mean bias score (i.e., the difference between the scores of the drawings attributed to Rubinstein and Amsalem) of the 133 students was 6.47, the standard deviation was 5.9, and the bias scores ranged from −6 to 26 points. Considering a difference of 3 points as "zero bias," the t test yielded a highly significant result ($t = 6.77$, $p < .001$) This finding is another demonstration of bias in scoring (see Babad et al., 1975); for the purpose of this analysis of susceptibility to biasing information as a personality dimension, the wide range of individual differences in the bias scores is quite satisfactory.

Correlational Analysis and Construction of Personality Scales

The analysis of the ACL responses presented a problem. Gough and Heilbrun (1965) developed over two dozen scales for scoring the ACL, but these scales had never been standardized or empirically validated on Israeli subjects. Since numerous adjectives had changed considerably in the process of translation, it was inappropriate to use the American scales on the Israeli subjects. Some adjectives have a very special meaning or double meaning in American English (e.g., "queer"), other adjectives cannot be adequately translated for lack of an appropriate word in Hebrew (e.g., "zany"), and cultural nuances are not readily reproducible in Hebrew.

Being unable to use Gough and Heilbrun's scales (and lacking an alternative nonpathologically oriented test in Hebrew), we were compelled to construct new scales from the available data. Factor analysis and other multivariate methods were ruled out because of the small number of subjects (133) relative to the number of variables (300). The most viable alternative was to use the subjects' bias scores as reference and to construct ad hoc scales according to the correlations of the ACL adjectives with that score. By chance alone, 15 of the 300 correlations between the ACL adjectives and the bias score should have reached the .05 level of significance, and 30 correlations should have reached the .10 level of significance. The results showed 34 correlations significant at the .05 level and 57 at the .10 level, both of which exceed chance level.

Of the 34 correlations reaching the .05 level of significance, 28 (or 82%) were negative, indicating that the high-bias subjects tended to check these adjectives *less* frequently than the low-bias subjects. To further examine this finding, we computed the correlations between the bias score and (a) the total number of adjectives checked, (b) the number of favorable adjectives checked, and (c) the number of unfavorable adjectives checked. (This analysis seemed more legitimate than the use of Gough and Heilbrun's (1965) specific scales, since it dealt with *overall* numbers of adjectives.) Only the last correlation was statistically significant ($r = -.17$, $p < .05$), indicating that the more biased subjects tended to check unfavorable adjectives *less* frequently in describing themselves. Gough and Heilbrun (1965) state that the low-scoring person on the number of unfavorable adjectives "is more placid, more obliging, more mannerly, more tactful, and probably less intelligent" (p. 6).

Four distinct clusters emerged in the scrutiny of the adjectives significantly related

Table 1
Mean Scores of High-Bias and No-Bias Groups on the Four Ad Hoc Scales (Study 1)

Group	Scale			
	Reasonableness (13 items)	Over-toughness (38 items)	Over-tenderness (31 items)	Over-enthusiasm (20 items)
High bias (n = 25)	6.28	4.68	5.12	3.72
No bias (n = 34)	4.97	7.29	6.74	5.59
t test	1.87*	2.25*	1.80*	2.18*

* $p < .05$.

to the bias score. About two thirds of the "significant" adjectives fitted in these clusters, whereas no meaningful pattern could be detected connecting the remaining adjectives. On the basis of these clusters, four ad hoc scales were defined: (a) *Reasonableness.* In this cluster, subjects described themselves as very reasonable, adhering to values of sobriety and realism. Typical adjectives are *rational, reasonable, civilized, mature, relaxed,* and *discreet.* (The adjectives in this cluster were *positively* related to bias.) (b) *Over-toughness.* This factor emphasizes extremely forceful and egocentric interaction with, and control of, other people. Typical adjectives are *aggressive, autocratic, cruel, rude, dominant, hardheaded,* and *hostile.* (The adjectives in this cluster—as well as in the next two clusters—were *negatively* related to bias.) (c) *Over-tenderness.* This factor also consists of a rather extreme description, but the emphasis is on lack of forcefulness, dependence on other people, and an overly "soft" position. Typical adjectives are *softhearted, timid, defensive, dependent,* and *sensitive.* (d) *Over-enthusiasm.* This factor relates to a behavioral style of a particular nature, characterized by excess energy and activity. Typical adjectives are *enthusiastic, blustery, praising, energetic, spunky,* and *uninhibited.*

The definitions of these four factors and the 300 ACL adjectives were submitted to three experienced psychologists, who were asked to mark all adjectives fitting these definitions. The three psychologists were unaware of the correlations between the adjectives and the bias score. All adjectives appearing in the lists of all three judges were included in the new ad hoc scales. The four new scales consisted altogether of 102 (out of 300) ACL adjectives, the majority of which (66) had *not* been significantly related to the bias score; only 23 were significant at the .05 level.

Extreme Group Comparison

To examine the relationships between the ad hoc scales and the bias score, an extreme-group design was chosen, comparing high-bias to no-bias subjects. All subjects whose bias scores were one standard deviation or more above the overall mean of the bias scores were included in the high-bias group (n = 25). All subjects whose bias scores were at the "zero bias" plus or minus one (i.e., granting Rubinstein between 2 and 4 points more than Amsalem) were included in the no-bias group (n = 34). Scores on the four ad hoc scales were computed for the subjects in these extreme groups, and t tests between the scores of the high-bias and no-bias groups were calculated. All t tests yielded statistically significant ($p < .05$) differences; the pattern of these results is presented in Table 1. The high-bias subjects described themselves as more reasonable than the no-bias subjects, while the latter group was higher on all three "over" scales—over-toughness, over-tenderness, and over-enthusiasm.

Discussion

This pattern is extremely interesting. The subjects who showed more susceptibility to biasing information, and whose scoring of the

drawings was less objective, chose to describe themselves as more reasonable, rational, and objective and as less emotionally extreme. On the other hand, the subjects whose scoring was more objective and unaffected by the biasing information described themselves as less reasonable and as more extreme in several emotional directions. To interpret these results, it seems quite clearly indicated that the self-descriptions should *not* be viewed as reflecting real, "actual-self" attributes but as expressions of a need system and an "ideal-self" image. Thus, the high-bias person *needs* to see himself or herself as more reasonable and must *deny* the existence of any emotional extreme in his or her personality. The no-bias subjects are not necessarily less reasonable, tougher, or more tender than the high bias subjects, but they can more readily admit the existence of these extremes in their personality and do not need to hold on so strongly to the rational facade.

This image of the high-bias person seems familiar from the literature on authoritarianism and dogmatism: rigid adherence to and overemphasis of conventional values (being OK and reasonable); rejection of emotional extremes and of self-awareness that may threaten one's adjustment; and narrow range of consciousness. Extreme strength or weakness must be rejected as "not me" if one's attempt to be over-reasonable is to succeed.

In conclusion, the only personality correlate of susceptibility to biasing information inferred from the results of Study 1 is dogmatism. This is in line with the findings reported by McFall and Schenkein (1970) and by Laszlo and Rosenthal (1970). Moreover, the absence of findings on other personality correlates of susceptibility to bias in this study is also consonant with other reports (Laszlo & Rosenthal, 1970; Rosenthal, 1966).

Although it appeals to common sense that susceptibility to bias is a reflection of dogmatism, this conclusion in the present context is speculative and highly tentative. Consequently, Study 2 was designed to focus directly on various manifestations of dogmatism and their relations to susceptibility to biasing information. The instruments used in Study 2 to assess dogmatism included specific self-report questionnaires, a performance task of field dependence, and questions concerning political ideology.

Study 2

Method

Overview of the Study

Bias in scoring was measured in Study 2 as in Study 1. Immediately thereafter, the Embedded Figures Test (Oltman, Raskin, Witkin, & Karp, 1971) was administered to measure field dependence, followed by a self-report questionnaire with scales measuring dogmatism, locus of control, defensiveness, and impulsiveness. A week later, the subjects responded to several direct questions concerning their political views.

Subjects

A new sample was selected for Study 2, consisting of 179 introductory psychology students at the School of Education of the Hebrew University of Jerusalem. All subjects were freshmen, and the study was conducted at the beginning of the school year.

Procedure

The DAP experiment was conducted in replication of Study 1. The only exception was that the drawing attributed to the low-status child (Amsalem) was replaced by another drawing duplicated from the DAP manual. The new drawing (Harris, 1963, Figure 57, p. 264) had a *higher* manual score than the drawing attributed to the high-status child, but in superficial examination it seemed to be of poorer quality than the other drawing. This change was due to the investigator's eagerness to maximize the bias effect and increase the spread of the bias scores. In retrospect, it seems that a simple counterbalancing of stimuli would have been much more appropriate, and this change is regrettable. Contrary to the manual scores, the five judges who scored the two drawings without the expectancy information and without access to the scoring of these drawings in the DAP manual awarded the new drawing 5 points *less* than Rubinstein's drawing. Thus, the investigator's overeagerness resulted in a "boomerang effect." The manual scores were therefore discarded, and a difference of 5 points in favor of the high-status child was defined as "zero bias."

Immediately following the scoring of the two drawings, a short (16-item) version of the Embedded Figures Test (Oltman et al., 1971) was administered, with a time limit of 4 min. The students then filled out a self-report questionnaire containing the short version of Rokeach's Dogmatism Scale (Schulze, 1973); a locus of control questionnaire (Rotter,

1966); a defensiveness questionnaire (from Babad, 1974); the Impulsiveness Scale (Barratt, 1965); and several filler items.

In the next class session, the students were asked to provide ratings (on 5-point scales) to the following questions: (a) What is your political ideology?—from 1 (left) to 5 (right); (b) What is the *intensity* of your political views?—from 1 (very low) to 5 (very high); and (c) What is your position regarding "greater Israel"? (i.e., a dovish or a hawkish position concerning the possibility of returning territories on the West Bank of the Jordan River to the Arabs—from 1 (against—dovish stand) to 5 (for—hawkish stand).

Results and Discussion

Overall Bias

The mean bias score of the 179 subjects was 6.20, the standard deviation was 5.33, and the bias scores ranged from −5 to 23 points. Considering a difference of 5 points as "zero bias" (in accordance with the stricter scoring of the five judges and disregarding the manual scores), the t test yielded a significant result ($t = 3.00$, $p < .05$). This finding is yet another demonstration of bias in scoring. In addition, it constitutes a replication of the DAP method of affecting expectancy bias.

Extreme Group Comparison

As in Study 1, the analysis of personality correlates was carried out using an extreme-group design, and the same principles were employed to define the high-bias and no-bias groups. All subjects whose bias scores were one standard deviation or more above the overall mean of the bias scores were included in the high-bias group ($n = 20$), while all subjects whose bias score was 5 ("zero bias") constituted the no-bias group ($n = 18$).

The self-report questionnaires were analyzed in several ways. The responses of all 179 subjects to each of the scales (dogmatism, locus of control, defensiveness, and impulsiveness) were factor analyzed, and new factor scores were computed. *None* of the t tests (of total scales *and* of new factor scores) yielded a significant difference between the high-bias and no-bias groups. Thus, no correlates of susceptibility to biasing information were identified in these self-report questionnaires.

No significant difference was found between the scores of the high-bias and no-bias groups on the Embedded Figures Test. This finding is contrary to McFall and Schenkein's (1970) report, where a significant expectancy effect on Rosenthal's (1966) photo-rating task was found for groups of subjects defined as field-dependent according to their Embedded Figures Test performance.

Considering the three direct political questions (political ideology, intensity of political views, and position about returning territories to the Arabs),[1] no significant differences were found between the high-bias and no-bias groups, although a slight (insignificant) tendency of the high-bias subjects towards the right-wing and hawkish positions was detected. Following Rokeach's (1960) argument concerning "the authoritarian of the left," the subjects' responses to these questions were rescored according to their *level of extremity*, disregarding the direction of their content. Thus, ratings of 1 and 5 were scored as extreme (1), and 2 and 4 were scored as moderate (2), and a rating of 3 was scored as neutral (3). *The* t *tests of the extremity scores yielded a statistically significant* ($p < .05$) *difference, the high-bias subjects consistently taking a more extreme position in reacting to all three questions.* Thus, while the groups did not differ in the content direction of their political views, the high-bias subjects differed from the no-bias subjects in that they consistently responded more extremely to these questions.

In light of this finding, the responses to all self-report questionnaires were rescored according to their level of extremity, and the data were reanalyzed. No additional differences were found. Thus, the difference between the bias groups in the extremity of responding was limited to the direct political questions.

General Discussion

The findings of the two studies can be summarized as follows: (a) Bias in scoring was demonstrated and replicated. (b) High-bias subjects described themselves as more

[1] This study was conducted *prior* to President Sadat's peace initiative in November 1977.

reasonable and rational than no-bias subjects and as less emotionally extreme in the directions of toughness, tenderness, and level of enthusiasm. (c) High-bias subjects responded more extremely than no-bias subjects to questions pertaining to their political views. On the other hand, no differences were found between the high-bias and no-bias groups in (a) responses to self-report scales (and subscales) measuring dogmatism, locus of control, defensiveness, and impulsiveness; (b) performance on the Embedded Figures Test; and (c) the content and direction of the subjects' political ideology. These findings are interpreted as being consistent with theoretical treatments of dogmatism as a style of thinking independent of ideology, even though they did not correlate significantly with the Dogmatism Scale.

The Embedded Figures Test and the rod-and-frame test had been constructed as performance measures of field dependence. Their relationship with the authoritarianism–dogmatism personality syndrome (as measured by self-report questionnaires) has been debated for over 20 years, with no clear resolution (see, for example, Ohnmacht, 1968; Rudin & Stagner, 1958; Witkin & Moore, Note 1). As to the relationship between performance on the Embedded Figures Test and susceptibility to biasing information, here again the reports are in conflict. McFall and Schenkein (1970) found a significant expectancy effect for field-dependent subjects but not for field-independent subjects, whereas in the present investigation high-bias subjects did not differ from no-bias subjects in their Embedded Figures Test performance. Moreover, the correlation between the Embedded Figures Test and the Dogmatism Scale on all 179 subjects in Study 2 did not reach statistical significance.

Another issue that has received lively debate over the years is whether authoritarianism is related to conservative and right-wing political ideology or whether this personality syndrome is independent of the *content* of the political ideology. Rokeach (1960) wrote quite convincingly about the "authoritarian of the left" (although his most famous piece of evidence was based on a sample of only 13 British communists), and the construction of the Dogmatism (D) Scale as a substitute for the original authoritarianism scale, the F scale, (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950) was motivated (in part) by this conviction. More recently, the superiority of the D scale over the F scale with regard to left-wing ideology was put in serious doubt by Thompson and Michel (1972), who found that *both* F and D scales were significantly related to political conservatism and Christian traditionalism. There is a need for research directed at identifying measures that more adequately reflect cognitive styles theoretically related to authoritarianism and dogmatism.

The findings of Study 2 have some bearing on this issue. The tendency of the high-bias subjects toward the right-wing ideology was slight and insignificant, whereas their excess extremity in political view independent of its contentual direction was clear and statistically significant. Thus, if bias in scoring can be considered as one facet of dogmatism, Rokeach's argument that the trait of dogmatism is largely independent of the specific content of the political ideology is supported by these findings.

Conclusions

Bias in scoring was demonstrated and replicated.

It is indicated by the data that susceptibility to biasing information is related to a dogmatic style of thinking, but is independent of the content and direction of ideology.

The utilization of self-report personality questionnaires and scales with college students is highly questionable and problematic: There are grounds to suspect that *ideal* self-image responses are provided when *actual* self-image descriptions are asked for and that specific questionnaires and scales loaded with social desirability may be too transparent for college students, resulting in falsified responses.

Reference Note

1. Witkin, H. A., & Moore, C. A. *Cognitive style and the teaching–learning process.* Paper presented at the annual convention of the American Educational Research Association, Chicago, April 1974.

References

Adorno, T., Frenkel-Brunswik, E., Levinson, D., & Sanford, N. *The authoritarian personality*. New York: Harper, 1950.

Babad, E. Y. A multi-method approach to the assessment of humor: A critical look at humor tests. *Journal of Personality*, 1974, *42*, 618–631.

Babad, E. Y., Mann, M., & Mar-Hayim, M. Bias in scoring the WISC subtests. *Journal of Consulting and Clinical Psychology*, 1975, *43*, 268.

Barratt, E. S. Factor analysis of some psychometric measures of impulsiveness and anxiety. *Psychological Reports*, 1965, *16*, 547–554.

Finn, J. D. Expectations and the educational environment. *Review of Educational Research*, 1972, *42*, 389–409.

Gough, H. G., & Heilbrun, A. B. *The adjective check list manual*. Palo Alto, Calif.: Consulting Psychologists Press, 1965.

Harris, D. B. *Goodenough–Harris drawing test manual*. New York: Harcourt, Brace & World, 1963.

Laszlo, J. P., & Rosenthal, R. Subject dogmatism, experimenter status, and experimenter expectancy effects. *Personality*, 1970, *1*, 11–23.

McFall, R. M., & Schenkein, D. Experimenter expectancy effects, need for achievement and field dependence. *Journal of Experimental Research in Personality*, 1970, *4*, 122–128.

Ohnmacht, F. W. Factorial invariance of the teacher characteristics schedule and measures of two cognitive styles. *Journal of Psychology*, 1968, *69*, 193–199.

Oltman, P. K., Raskin, E., Witkin, H. A., & Karp, S. A. *The group embedded figures test*. Palo Alto, Calif.: Consulting Psychologists Press, 1971.

Rokeach, M. *The open and closed mind*. New York: Basic Books, 1960.

Rosenthal, R. *Experimenter effects in behavioral research*. New York: Appleton-Century-Crofts, 1966.

Rosenthal, R. Interpersonal expectations: Effects of the experimenter's hypothesis. In R. Rosenthal & R. L. Rosnow (Eds.), *Artifact in behavioral research*. New York: Academic Press, 1969.

Rosenthal, R. Teacher expectations and their effects upon children. In G. S. Lesser (Ed.), *Psychology and educational practice*. Glenview, Ill.: Scott, Foresman, 1971.

Rosenthal, R., & Jacobson, L. *Pygmalion in the classroom*. New York: Holt, Rinehart & Winston, 1968.

Rosenthal, R., & Rubin, D. B. Interpersonal expectancy effects: The first 345 studies. *The behavioral and brain studies* (Vol. 1). Cambridge, England: Cambridge University Press, in press.

Rotter, J. B. Generalized expectations for internal versus external control of reinforcement. *Psychological Monographs*, 1966, *80*(1, Whole No. 609).

Rudin, S. A., & Stagner, R. Figure–ground phenomena in the perception of physical and social stimuli. *Journal of Psychology*, 1958, *45*, 213–225.

Schulze, R. H. K. A shortened version of the Rokeach dogmatism scale. In J. P. Robinson & P. R. Shaver (Eds.), *Measures of social psychological attitudes* (Rev. ed.). Ann Arbor: Institute for Social Research, University of Michigan, 1973.

Thompson, R. C., & Michel, J. B. Measuring authoritarianism: A comparison of the F and D scales. *Journal of Personality*, 1972, *40*, 180–190.

Received December 5, 1977

ournal of Personality and Social Psychology
979, Vol. 37, No. 2, 203–220

Prediction and Perception of Social Motives

Judith E. Maki
University of California, Santa Barbara

Warren B. Thorngate
University of Alberta
Edmonton, Canada

Charles G. McClintock
University of California, Santa Barbara

In two experiments, subjects were required to observe and predict the behavior of a hypothetical "chooser" who made choices for him- or herself and for a hypothetical other in a series of decomposed games. The preference for outcomes, or social motivational orientation, of the chooser was preprogrammed and varied across conditions. Results of the experiments confirmed that subjects were more readily able to detect the outcome preferences of choosers who made choices according to individualistic or competitive choice rules than of choosers who behaved in a prosocial or negatively self-interested manner. Furthermore, the prediction data obtained from Experiment 2 revealed that subjects tended to perceive choosers' own gain as an important component of most of the choosers' secondary motivation. Evidence from subjects' ratings of the choosers' personality attributes and estimates of the relative weights the choosers attached to their own and the other's gain (Experiment 2) indicated that subjects formed distinctive impressions of the choosers despite differences in predictive accuracy across conditions. A third experiment, performed to investigate the relationship between predictive accuracy and the mathematical complexity of the choosers' various choice rules, found no evidence that mathematical complexity influenced subjects' performance on the prediction task.

The study of social exchange is concerned with situations where the outcomes of two or more individuals are dependent upon their mutual behaviors. Recent literature dealing with social exchange has focused on determining the values that individuals seek to maximize in outcome interdependence situations. McClintock (1972b), for example, has proposed that the value of outcomes to an individual may be defined by that person not only in terms of what he or she receives for self but also in terms of what others receive. Experiments by Messick and McClintock (1968), McClintock, Messick, Kuhlman, and Campos (1973), and Kuhlman and Marshello (1975) confirm that people show individual differences in their tendencies to maximize their own outcomes, joint outcomes, and the relative difference between their own and others' outcomes.

These preferences for outcomes, or social motives, have been formally characterized in a geometric model by Griesinger and Livingston (1973). The Griesinger and Livingston representation takes the form of a two-dimensional space where outcomes or payoffs to self are ordered along the horizontal axis, and outcomes to the other are ordered along the vertical axis. An actor's preference for certain distributions of points in this space is represented as a vector of infinite length extending from the origin. Thus, an individual-

The research reported here was partially supported by a Canada Council Doctoral Fellowship to Judith E. Maki, Canada Council Grant #S71-1044 to Warren B. Thorngate, and a University of California Faculty Research Grant and National Science Foundation Grant #77-03862 to Charles G. McClintock. The authors are indebted to Elizabeth Bigman, Melissa Bubar, and Beverly Davidson, who assisted in running the subjects.

Requests for reprints should be sent to Charles G. McClintock, Department of Psychology, University of California, Santa Barbara, California 93106.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3702-0203$00.75

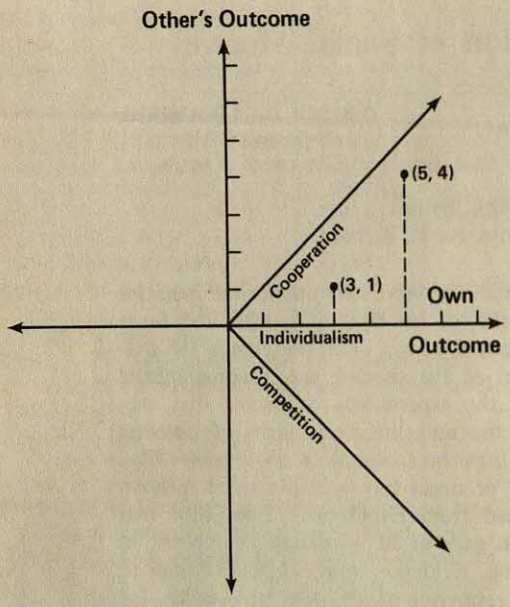


Figure 1. The Griesinger and Livingston vector representation of social motives. (Numbers in parentheses are examples of distributions of points to self and other.)

istic motive, or preference to maximize one's own gain, can be represented as a vector extending in the positive direction along the horizontal axis. The locations of the vectors representing cooperation (maximization of joint gain) and the competition (maximization of relative gain) are 45° removed from the horizontal in the counterclockwise and clockwise directions, respectively (see Figure 1).

Preference orderings for points in this outcome space are defined in terms of the magnitudes of their perpendicular projections on the actor's motivational vector. The greater the projection of a point, the more preferred is the associated outcome. For example, an individualistic actor might be presented with a choice between a point that affords self 5 units and the other 4 units and a point that affords self 3 units and the other one unit. As Figure 1 indicates, the point 5,4 is preferred to the point 3,1 because of its greater projection on the actor's motivational vector: Receiving 5 for self is better than receiving 4. However, if the actor is competitively oriented, 3,1 is preferred to 5,4 because it affords the actor greater relative advantage over the other (2 units as opposed to 1). The point 5,4 is preferred by a cooperative actor because it leads to the higher joint gain.

The previously cited work of McClintock, Messick, and their colleagues has led to the refinement of a method for assessing the social motivational orientations of experimental subjects utilizing a form of payoff representation known as the decomposed game. The decomposed game procedure involves simultaneously presenting both members of a pair of subjects with a series of two- or multichoice decision situations. Each of the choice alternatives in these situations affords a certain payoff to the subject and a certain payoff to another dyad member. Thus, each choice alternative in a decomposed game represents a point in self/other outcome space. Subjects are required to select their preferred alternative on each trial in ignorance of each other's behavior. Classification of subjects' motivational orientations is performed by comparing their patterns of choices in the decomposed games with those which would have been produced by actors displaying perfect consistency with the individualistic, cooperative, and competitive vectors.

This technique has two major strengths as a motive assessment procedure. First, while maintaining the essentials of outcome interdependence between subjects, the decomposed game minimizes the possibility that strategic considerations will influence the subjects' selections. Although each subject's choice affects both own and other's outcomes, neither subject can utilize his or her choices in an instrumental way to influence the other's choices. Second, the procedure allows the investigator to confront the subject with combinations of choices reflecting different outcome structures that are dominant in respect to the motives under consideration.[1]

[1] The dominance structure of a game is determined by examining which motives or values are maximized by particular choices. Suppose that an actor is presented with a decomposed game offering a choice between A, which affords the actor 50 units and the other 10 units, and B, which affords the actor 40 units and the other 30 units. Choice A will maximize both the actor's own gain and his or her relative gain. Choice B will maximize joint gain of the actor

The "Naive Psychology" of Social Motives

Although researchers have examined individual differences in motives and how they relate to subjects' observed laboratory game play (e.g., Kuhlman & Marshello, 1975), little empirical attention has been paid to the question of whether and how subjects, as "naive psychologists" (Heider, 1958), perceive the outcome preferences of others. Whenever outcome interdependence and the possibility of social influence exist, an actor intent on maximizing preferred outcomes must simultaneously take into consideration both own and other's goals, as well as the relative power of self and other (Thibaut & Kelley, 1959). Thus, in most settings of outcome interdependence, accurate information about the other's preferred outcomes should be instrumental in guiding the actor's attempts both to approximate directly preferred outcomes and to influence strategically the other's behavior.

Perception of motives in the Prisoner's Dilemma Game (PDG). One of the few investigations relevant to the topic of the naive perception of social motives was performed by Kelley and Stahelski (1970). They assessed the motivational orientations of subjects as cooperative or competitive by asking them about the goals they wished to maximize in the PDG. Subjects were then paired with other subjects who had expressed cooperative or competitive intent. Periodically, subjects' impressions of their partners were gathered during the course of PDG play. The results indicated that a competitive orientation on the part of the other was more readily perceived than a cooperative orientation. Further, although cooperative subjects were able to detect their partner's goals regardless of whether these goals were cooperative or competitive, competitive subjects often misjudged cooperative partners' goals to be competitive. Finally, when asked about the propensity of others to select the two PDG choices, the competitors indicated that the vast majority of people would be similar to themselves and opt for the competitive goal, whereas cooperators felt that some people would be like themselves and others would choose the competitive alternative.

Expectations concerning the goals of others in the decomposed game. The findings of a recent experiment conducted by Kuhlman and Wimberley (1976), however, suggest that the Kelley and Stahelski study may indicate more about the motivational structure of the PDG than about the beliefs that cooperators and competitors have concerning the relative predominance of these motivational orientations in the population. Kuhlman and Wimberly classified subjects as cooperative, competitive, or individualistic on the basis of their choices in a series of three-choice decomposed games. Since the desire to maximize relative gain and the desire to maximize one's own gain represent the two primary motivational reasons for "competition" in the PDG (McClintock, 1972a), the Kuhlman and Wimberley competitors and individualists correspond to a partition of Kelley and Stahelski's PDG competitors.

After making a series of decomposed game choices, subjects in the Kuhlman and Wimberley experiment were asked to estimate what proportion of other persons would make each of the three choices in four three-choice decomposed games having various dominance structures in regard to joint, relative, and own gain maximization. Also, subjects estimated the proportions of others who would make the cooperative and competitive choices in the PDG. Subjects then played 30 trials of a PDG against a uniformly cooperative or competitive other.

The results obtained on choice behavior in the PDG and subjects' estimates of others' choices in the PDG paralleled those of Kelley and Stahelski (1970). However, estimates of population choice proportions in the decom-

---

and the other. An example of a decomposed game with a different dominance structure is a situation in which the A choice affords 50 units for self and 20 units for the other and the B choice affords 40 units for self and 0 units for the other. Choice A maximizes (is the dominant choice with respect to) one's own and joint gain, whereas B maximizes relative gain. In the first example, the decomposed game has a similar motivational structure to the Prisoner's Dilemma Game (PDG); the latter example most closely resembles the Maximizing Difference Game (MDG; McClintock & McNeel, 1966).

posed games revealed different expectations about the preferences of others. Results from a game in which joint, own, and relative gain were each maximized by separate choices were particularly interesting. In this situation, cooperative subjects thought that the joint-gain maximizing choice would be the most popular followed by the own-gain choice and then the relative-gain choice. Competitive subjects' ordering of others' preferences was just the reverse of that of the cooperative subjects: relative gain followed by own and then joint gain. Individualistic subjects felt that the own-gain maximizing choice would be preferred by most and that the relative-gain maximizing choice would be slightly more popular than the joint-gain choice.

When considered together with the estimates for others' choices in the remaining decomposed games, these data suggest that each person may view her or his own motive as the modal orientation in the population. Recent work by Ross (1977) provides additional support for this interpretation. He observes that individuals who lacked good baseline data about typical behavior in a situation resorted to personal criteria to make predictions about others. In other words, one's own behavioral preferences are assumed to be more normative than nonpreferred tendencies.

Detecting the motives of others in the decomposed game. The previous research concerns how an actor's own motivational dispositions influence his or her attributions of motives to others. None of this research has asked the prior, perhaps more fundamental, questions of whether and how rapidly actors can make accurate attributions to various major dispositions, given that these dispositions are accurately and consistently represented in the behavior of another. Kelley (Note 1) terms such dispositional acts as motivational moves with "action-immediate consequences," that is, behaviors for which an observer can make a direct attribution concerning an actor's goal in regard to the actor's own and another's outcomes.

The major purpose of the studies to be reported was to determine whether and how rapidly actors can make consistent and accurate attributions regarding different classes of social motives. Subjects in Experiments 1 and 2 were required to observe and predict the behavior of a hypothetical person called "the chooser," who made choices consistent with a particular motivational vector across a series of decomposed games. The use of the decomposed game format made it possible to represent various motivational choice vectors in a relatively systematic and unambiguous way. Further, the hypothetical person's choices were made in a context where he or she had complete control over own and some other person's outcomes. Hence, the observer or subject had no reason to anticipate that the hypothetical person was making choices for strategic reasons, that is, to influence the other to behave differently at some subsequent point in time. In effect, then, the hypothetical person was portrayed as making motivational moves with action-immediate consequences. And it was the observer's task to detect the nature of these motivational moves.

A second purpose of the first two experiments was to explore subjects' impressions of actors behaving according to different motivational rules. The overwhelming majority of impression-formation studies have focused on impressions of individuals described by a series of personality trait adjectives. There is a paucity of information, however, about the dispositional information that overt behaviors convey. The lack of such information is unfortunate, since behaviors are theorized to play a central role in the attribution of dispositions (Jones & Davis, 1965). Dispositional attributions, if they deal with stable characteristics of an actor's orientation toward self and others, should in turn be significant predictors of an observer's reactions to the actor in outcome interdependence situations.

Of particular interest to the present investigation is whether choice behavior consistent with a particular motivational vector results in impressions that distinguish the actor from other actors who have different outcome preferences. To find out in what ways impressions might differ, subjects in Experiments 1 and 2 rated the chooser on bi

polar adjective scales immediately following their observations and predictions of the chooser's choices.

Finally, since each motivational vector is defined by a different set of mathematical operations in regard to one's own and the other's outcomes (e.g., maximization of relative gain implies the algebraic operation of subtracting other's outcomes from own outcomes, whereas joint gain implies adding own and other's outcomes), a third study was conducted to determine whether the rate of detecting a particular vector was related to its mathematical as well as its social properties.

Experiment 1

The major hypothesis of Experiment 1 was that subjects would experience greater difficulty predicting the choices made by a chooser who placed equal or greater value on other's outcomes (i.e., behaved prosocially) than a chooser who placed higher value on own rather than other's outcomes (i.e., behaved in a self-centered manner). This finding was anticipated on the basis of previous studies that have demonstrated the pervasiveness of self-centered choice behavior in laboratory outcome interdependence situations (e.g., McClintock & McNeel, 1966; McClintock & Moskowitz, 1976).

In order to get some idea of the generality of the predictions of the hypothesis, a range of motives was examined. In addition to the motivational vectors of cooperation, individualism, and competition, the prosocial motive, altruism (maximize other's gain), was included in the design. Three other conditions corresponded to vectors falling midway between pairs of these "pure" motives (i.e., altruism/cooperation, cooperation/individualism, and individualism/competition; see Figure 2). Thus, in three out of seven conditions, namely, altruism, altruism/cooperation, and cooperation, the chooser met the criterion of demonstrating as much or greater concern with other's outcomes as with own outcomes. Poorer predictive accuracy was expected in these conditions than in the remaining conditions, where the chooser valued own outcomes more highly than other's.

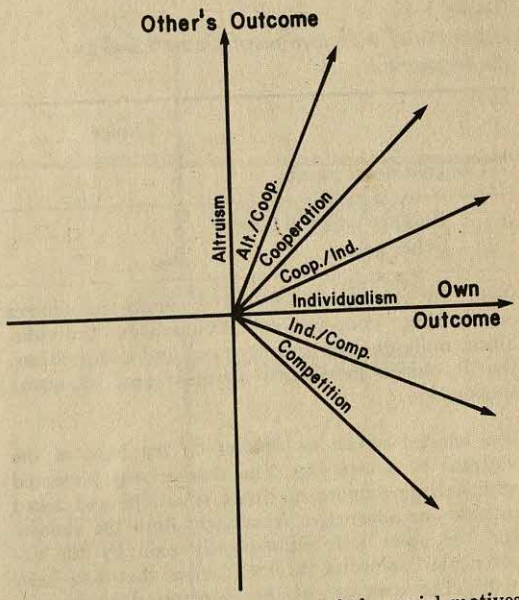


Figure 2. Vector representation of the social motives studied in Experiment 1.

Method

Subjects. The subjects were 70 introductory psychology students (29 males, 41 females) at the University of Alberta, Canada. All subjects received course credit for their participation. Subjects were randomly assigned to observe and predict a chooser having one of the seven social motives described previously.

Stimuli and apparatus. The stimulus materials consisted of two-choice decomposed games. Decomposed game situations were presented, responses were made, and feedback was given at a standard teletype connected to a PDP8/I computer. Payoffs to the chooser and the other in every decomposed game trial were randomly selected by a computer subroutine from the digits 0 through 9. The correct choice on each trial was determined according to which of the randomly generated chooser–other alternatives had the largest projection on the chooser's motivational vector. It is important to note that the correct choice often satisfied both the particular chooser's motive and one or more of the other motives also being investigated in the study. How many other motives and which other motives were also satisfied by the chooser's choice varied from trial and trial. An example of a decomposed game as it appeared to subjects is shown in Table 1.

Procedure. After subjects were brought to the experimental room, they were informed that the experiment was concerned with how well one could predict another person's behavior and the kinds of impressions formed of the person. Subjects were told that during the previous year, pairs of students had been brought to the lab. One member of each pair

Table 1
Example of a Decomposed Game Used in Experiment 1

Designation of points	Choice A	Choice B
Chooser gets	5	1
Other gets	3	6

Note. In this example, the A choice maximizes cooperation, cooperation/individualism, individualism, individualism/competition, and competition; the B choice maximizes altruism and altruism/cooperation.

was selected to act as chooser on the basis of the outcome of a coin flip. The chooser was presented with a large number of choice situations and asked to pick one alternative from each. Both the chooser and the other were subsequently paid by the experimenter according to the choices that had been made. The structure of the decomposed games was then explained, and subjects were told that during the experiment their task would be to predict, as best they could, what the chooser chose in each situation.

Three practice trials were provided (without feedback) to acquaint the subjects with the manner in which the situations would be be presented on the teletype and the manner in which the subjects were to enter their responses. Following the practice trials, subjects predicted the chooser for a total of 40 trials. On the first 10 of these trials, no feedback was provided about the correctness of predictions. After the 10th trial, feedback was introduced. The pace of the experiment was controlled by the subject's rate of responding.

Upon completion of the last trial, subjects evaluated the personality attributes of the chooser on 30 9-point bipolar adjective scales: young–old; unCanadian–Canadian; bad–good; active–passive; weak–strong; cooperative–competitive; rational–irrational; peaceful–warlike; conservative–liberal; immoral–moral; religious–unreligious; rich–poor; stable–unstable; untrustworthy–trustworthy; honest–dishonest; impolite–polite; selfish–unselfish; cruel–gentle; calm–agitated; wise–stupid; kind–unkind; usual–unusual; unpredictable–predictable; serious–humorous; unfriendly–friendly; sociable–solitary; timid–courageous; masculine–feminine; insincere–sincere; and tolerant–intolerant. These scales were selected both on the basis of general interest and on their potential ability to discriminate among the choosers along psychologically meaningful dimensions, as suggested by the factor-analytic work of Osgood, Suci, and Tannenbaum (1957) and Kuusinen (1969).

Results

The average number of times subjects correctly predicted the chooser in each block of 10 trials is shown in Figure 3. The figure pro
vides support for the hypothesis: Subject
were better at predicting individualistic
through-competitive choosers than coopera
tive-through-altruistic choosers. Overall, th
former choosers were predicted accurately o
about 83% of the trials, whereas for the lat
ter choosers, the accuracy rate was only 66%

Two separate analyses of variance were
performed after submitting the proportion o
correct predictions for each subject on each
trial block to an arc sine transformation. In
the first analysis, performance on the first
block of 10 trials when no feedback was
given to subjects was compared with mean
performance on the subsequent feedback trials.
Significant main effects were obtained for
chooser motive, $F(1, 63) = 6.10$, $p < .001$,
and feedback, $F(1, 63) = 21.09$, $p < .001$, as
well as a significant Motive × Feedback in-
teraction, $F(6, 63) = 3.16$, $p < .01$. An ex-
amination of Figure 2 indicates that while
accuracy increased in the competitive, altru-
istic, and individualistic/competitive chooser
conditions with the introduction of feedback,
the accuracy for cooperative and altruistic/
cooperative choosers declined slightly. The
second analysis, performed on the three
blocks of trials in which subjects received
feedback, once again yielded a significant
main effect for motive, $F(6, 63) = 7.92$, $p <$
.001. There was also a significant effect for
trial blocks, $F(2, 126) = 13.93$, $p < .001$, in-
dicating that subjects improved their per-
formances with practice. The Motive × Trial
Blocks interaction, however, was not signifi-
cant, $F(12, 126) < 1$. An a priori contrast
performed on the overall accuracy achieved
on feedback trials between those who ob-
served altruistic, altruistic/cooperative, and
cooperative choosers versus cooperative/indi-
vidualistic, individualistic, individualistic/
competitive, and competitive choosers indi-
cated a statistically reliable difference between
these two groups, $t(63) = 6.66$, $p < .001$.

To determine the extent to which observers
formed similar or dissimilar impressions of
the choosers whom they observed, the ob-
tained personality attribute ratings were sub-
mitted to a discriminant function analysis
(Tatsuoka, 1970). Discriminant function

analysis is a multivariate statistical technique that finds the set of uncorrelated, weighted linear combinations of predictor variables that maximally discriminates among individuals classified according to some grouping variable. In this study, groups consisted of chooser-motive treatment conditions.

Of the discriminant functions that resulted from this analysis, only one reached statistical significance, Wilks lambda (180 = .0116, $p < .02$. This function accounted for 47.53% of the discriminatory power of the battery of derived functions. The locations of group means on the function or "centroids" are plotted in Figure 4.

Figure 4 indicates that subjects' ratings were able to discriminate among the choosers along a dimension roughly corresponding to social motivational orientation. While the ordering of centroids on this dimension is not isomorphic with the degree of the various choosers' prosocial intent, subjects' ratings of choosers discriminated between those with prosocial and those with self-centered outcome preferences. The larger discriminant coefficients obtained from the analysis indicated that higher ratings of agitation, stability, selfishness, unfriendliness, badness, masculinity, and—oddly enough—politeness tended to be associated with self-centered classification by this function.

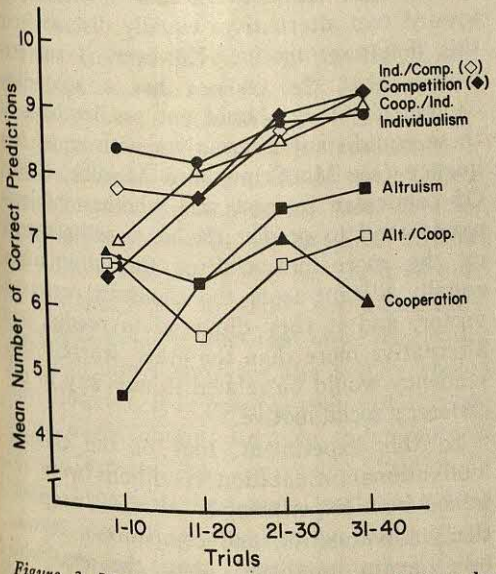


Figure 3. Mean number of correct predictions by chooser motive condition and trial blocks, Experiment 1.

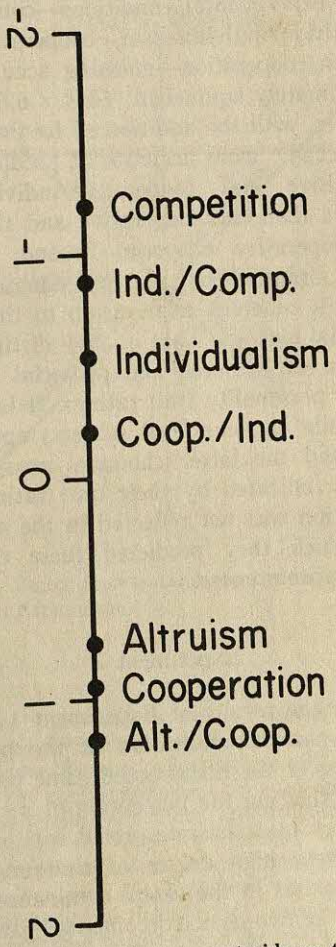


Figure 4. Location of group centroids on the significant discriminant function, Experiment 1.

Discussion

Two important findings emerged from this experiment. First, subjects were not uniformly accurate at predicting the outcome preferences of actors. In a third-party prediction task, the social motivational orientations of self-centered choosers were more readily detected than those of prosocial choosers. This finding is particularly interesting if considered in the light of subjects' performance on the initial no-feedback trials, which amounted to guessing the choosers' preferences. In four

of the experimental conditions—competition, competition/individualism, cooperation, and altruism/cooperation—guessing accuracy was approximately equivalent, (Ms = 6.6 to 7.0). However, with the addition of feedback, subjects became more accurate at predicting the competitive and cooperative/individualistic choosers than the cooperative and the altruistic/cooperative choosers. Second, subjects formed impressions of the choosers that showed a coherent relationship to the choosers' social motives. That is, they distinguished between self-centered and prosocial choosers in their personality trait ratings. It is curious that while at some level subjects apparently recognized the latter choosers' prosocial intent, as indicated by their trait ratings, this recognition was not reflected in the accuracy with which they predicted these choosers' outcome preferences.

Experiment 2

While the results of Experiment 1 are encouraging as an indication of the potential usefulness of the decomposed game technique for investigating the perception of social motives, they must be interpreted with caution. A relatively high degree of confounding of motives exists in the choice-dominance structures of the two-choice decomposed game. Depending upon the particular configuration of chooser's and other's payoffs in the situations employed as stimuli in Experiment 1, anywhere from one to all seven of the motivational vectors that were being investigated could have been represented by the same decomposed game choice. The most precise diagnostic information about the chooser's motive (from a mathematical standpoint) should have been communicated in decomposed games in which the alternative selected by the chooser was not the same as the alternative preferred by choosers with any of the other motivational orientations under study. However, this special situation could only have occurred in Experiment 1 when the chooser's motive was altruism or competition. Even under this circumstance, the situation would be expected to occur relatively infrequently because of the random nature of the computer-generated decomposed game entries. Thus, it is unclear to what extent subjects in every condition of Experiment 1 received equivalent amounts of information and whether it was in a sufficiently unconfounded form to permit them to accurately locate the chooser's motivational vector.

The purpose of Experiment 2 was to extend and replicate the findings of the previous study. In this experiment, the decomposed game observation and prediction paradigm of Experiment 1 was used but modified in order that the degree and nature of the confounding of motives could be precisely specified and controlled across conditions. Like the previous study, the issues of whether choice patterns consistent with different motivational vectors are perceived by subjects with differential accuracy and whether variations in social motivational orientation influence impression formation were of primary interest.

In addition, however, Experiment 2 was also concerned with subjects' reactions in situations where an actor's dominant choice vector is not represented among the decomposed game choices, but the two immediately adjacent choices are represented. If an actor is assumed to be motivated by only a single vector, then the actor should be indifferent toward two alternatives equally distant from this dominant motive. However, if subjects believe that the chooser has a secondary choice rule, they should not predict both of these equidistant alternatives with equal frequency (see MacCrimmon & Messick, 1976). Of particular interest was whether subjects would tend to predict the more self-centered or the more prosocial of the alternatives equally distant from the chooser's preferred vector, and if they did tend to predict one alternative more than the other, whether this tendency would be related in any way to the chooser's social motive.

In this experiment, four of the chooser motivational orientation conditions were retained from Experiment 1—altruism, cooperation, individualism, and competition. It was once again predicted that choosers who adopted an individualistic or competitive goal would be more accurately predicted than

choosers who made choices consistent with an altruistic or cooperative goal. As well as the vectors just noted, four motivational vectors not previously investigated were included in the design of Experiment 2—aggression (minimize other's gain), martyrdom (maximize other's relative gain), masochism (minimize own gain), and sadomasochism (minimize joint gain). The inclusion of these additional vectors permitted a more general test of the hypothesis that self-interested motives are the most accurately perceived. Figure 5 shows the locations of these vectors in own–other outcome space.

Although seldom observed in laboratory gaming situations, aggression undoubtedly plays a significant role in many real-life outcome interdependence situations. Studies conducted by Davis (1972) suggest that an aggressive preference on the part of an actor is often interpreted as a symptom of a competitive nature. Aggression differs fundamentally from martyrdom, masochism, and sadomasochism in that one's own gain receives greater weight (zero) than the other's gain (negative) in determining outcome preference. Thus, it was anticipated that subjects would experience relatively little difficulty in predicting an aggressive chooser and that accuracy would resemble that observed for other self-centered motives.

The vectors of martyrdom, masochism, and sadomasochism are rare not only in laboratory games but also in real life. Despite their abnormal status among social motives, subjects' relative abilities to detect them are of theoretical interest. To the extent that an actor repeatedly behaves in an atypical fashion (i.e., displays a preference for negative over positive outcomes to self), an expectancy hypothesis would predict that observers would have difficulty identifying the actor's choice rule. However, it is possible that just the opposite might obtain. Because these motives violate expectations concerning "rational" choice behavior, they may tend to arouse greater interest or attention on the part of observers and, hence, may potentially be detected more readily than those motives which affords the self greater positive outcomes.

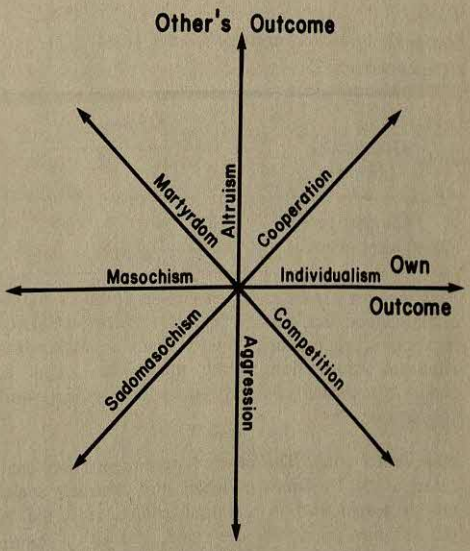


Figure 5. Vector representation of the social motives studied in Experiment 2.

Method

Subjects. The subjects were 89 female introductory psychology students at the University of California, Santa Barbara, who participated in the experiment in partial fulfillment of a course requirement. The number of subjects who watched choosers with each of the eight motivational orientations varied between 10 and 12. An additional 12 subjects were run but were discarded from the analysis for reasons of equipment failure (10), experimenter error (1), or failure to understand the instructions (1).

Stimuli. Subjects alternately observed the chooser make choices then predicted the chooser's behavior, without feedback, in a series of four-choice decomposed games (see Table 2). An observation/prediction trial block consisted of a total of 8 observation trials followed by 12 prediction trials. This observation/prediction sequence was repeated five times in the course of the experiment. Although the observation/prediction format differed from that used in Experiment 1, no differences in the direction of the findings were expected.

The eight decomposed games presented during the observation trials contained one choice that was dominant in terms of the chooser's motivational orientation. Games were constructed in a manner such that this dominant alternative maximized *only* the chooser's motive in 25% of the games. In the remaining 75%, this dominant choice not only maximized the chooser's motive but also two neighboring motives (motives falling within 90° of the chooser's vector).

Motivational structures of 8 of the 12 decomposed games presented during the prediction trials were the same as the 8 games presented during the

Table 2
Example of a Decomposed Game Used in Experiment 2

Designation of points	Choice A	B	C	D
Chooser gets	2	4	1	6
Other gets	1	0	3	1

Note. In this example, the A choice uniquely maximizes sadomasochism; the B choice uniquely maximizes aggression; the C choice maximizes altruism, martyrdom, and masochism; the D choice maximizes competition, individualism, and cooperation.

observation trials. The other 4 prediction trials consisted of (a) 2 games in which the chooser's social motive would lead to a unique choice, as it did in the 8 games previously described, and (b) 2 games in which there was no unique choice consistent with the chooser's motive, but rather the chooser's motive was equally satisfied by two outcomes representing maximal projections on the motivational vectors immediately adjacent to the chooser's vector. Inclusion of the latter games in the prediction trials provided a means for examining subjects' beliefs about the choosers' possible secondary motives.

Payoffs on all games were randomly transformed so that no two situations looked identical to subjects. Thus, accurate predictions could not be made by rote memorization of the observation trials.

Procedure. From one to four subjects participated in each experimental session. Upon their arrival, subjects were placed in individual cubicles. Each of these cubicles was equipped with a TV monitor and a response box with four letters labeled A through D. The instructions, presented on videotape, were read by a male graduate student.

Subjects were instructed that the purpose of the experiment was to develop a test of "social intelligence" that would measure a person's relative ability to understand and predict the behavior of another person. It was explained that they would be required to first observe and then predict the behavior of a person called "the chooser" in some simple choice tasks in which the chooser would make choices for self and for a person called "the other." No information about the other was provided, aside from the fact that the other was affected by the chooser's behavior. In an attempt to motivate subjects to be accurate, a reward of $25 was to be given to the subject in the study who made the most correct predictions. In addition, each subject was told that she would receive group and individual performance feedback. All of these promises to subjects were ultimately kept in a debriefing letter.

A sample decomposed game was then presented and the nature of the payoffs explained. Subjects were told that the payoffs were in cents and that they were to assume that these cents came from a bowl with an unlimited number of pennies. When the chooser made a choice that resulted in a positive payoff to self and or other, pennies were presumably extracted from the bowl. When either the chooser or the other got a negative payoff as a result of the chooser's choice, the party or parties who received the negative payoff returned the appropriate number of cents to the bowl. The reference to a bowl with an unlimited number of pennies was introduced to reduce the likelihood that subjects would interpret the situation as zero-sum (Rapoport, 1966).

During the first part of the experiment, blocks of observation and prediction trials were clearly segregated by segments of blank videotape and appropriate warning messages: "In five seconds the observation (prediction) trials will begin." On each observation trial, subjects first saw the decomposed game, and then, after a brief pause, the letter of the chooser's choice was displayed along with the decomposed game choices. Prediction trials were similar to observation trials except that the word *prediction?* appeared beneath the decomposed game instead of the letter of the chooser's selected alternative. No feedback was provided. Subjects were allowed approximately 5 sec in which to indicate their predictions by pressing the appropriate button. All responses were recorded by an Esterline-Angus event recorder.

Following the observation and prediction trials, subjects were given a questionnaire similar to that used in Experiment 1, which asked them to evaluate the chooser's personality attributes on 22 9-point bipolar adjective scales. These included 16 scales that were selected from among the items used in Experiment 1 because the mean ratings received by the choosers in the previous experiment exhibited a range of at least 2.5 scale points: calm–agitated; weak–strong; rational–irrational; religious–unreligious; impolite–polite; bad–good; usual–unusual; insincere–sincere; active–passive; cooperative–competitive; peaceful–warlike; stable–unstable; selfish–unselfish; kind–unkind; masculine–feminine; and

Table 3
Mean Percentage of Correct Predictions, by Chooser Motive Condition, in Experiment 2

Condition	% correct predictions
Individualism	78.8
Cooperation	64.1
Altruism	58.8
Martyrdom	41.1
Masochism	51.1
Sadomasochism	28.8
Aggression	65.0
Competition	74.7
Chance	29.2

unfriendly–friendly. Six new scales (punishes self–punishes other, hateful–loving, unfair–fair, ungenerous–generous, unconfident–confident, just–unjust) were added.

Subjects also estimated the weights that the chooser placed on own gain and other's gain on separate 9-point scales, which ranged from 1, "Chooser wanted to avoid giving cents to self (other)," to 9, "Chooser wanted to give as many cents to self (other) as possible."

Results

The overall mean percentages of correct predictions made by subjects in each condition are shown in Table 3. Figure 6 shows the mean number of correct predictions plotted over prediction trial blocks. For analysis of variance (ANOVA) purposes, those prediction trials for which there were two possible correct responses were omitted. (The results for these trials will be discussed later.) For ease of computation, subjects were randomly eliminated from conditions in which the number of observers exceeded 10. In addition, prior to performing the ANOVA an arc sine transformation was once again performed on the proportion of correct responses made by each subject on each block of prediction trials.

The results of the analysis revealed significant main effects for chooser motivational orientation, $F(7, 72) = 5.955$, $p < .001$, and trial blocks, $F(4, 288) = 8.076$, $p < .001$. As Table 3 and Figure 6 indicate, individualism and competition proved to be the easiest motives for subjects to detect, followed by aggression, cooperation, and altruism. Masochism, martyrdom, and sadomasochism posed the most difficulty, with subjects exposed to the sadomasochistic chooser performing at no better than the chance level. While there was a general trend toward improvement over the course of the chooser observation and prediction phase, as indicated by the significant trial blocks effect, the rate of improvement in accuracy did not differ appreciably across conditions: The Motive × Trial Blocks interaction did not reach statistical significance, $F(28, 288) = 1.102$.

Like the results of the overall ANOVA, a planned comparison designed to replicate the comparison performed on the data of Experiment 1 supported the original findings. The predictive accuracy of those who watched a

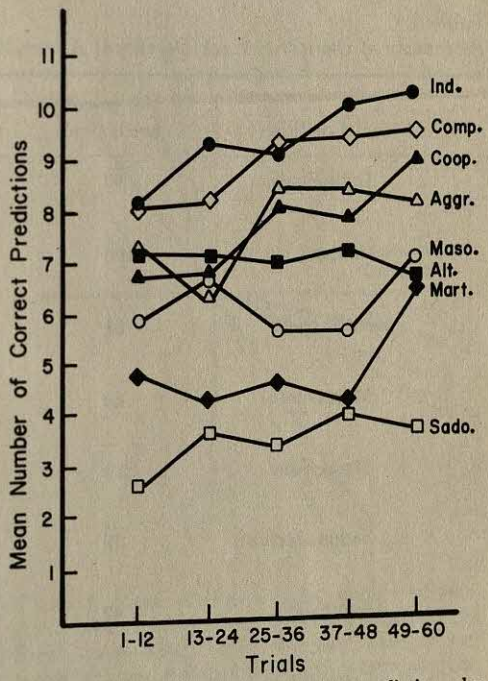


Figure 6. Mean number of correct predictions by chooser motive condition and trial blocks, Experiment 2. (Ind. = individualism; Comp. = competition; Coop. = cooperation; Aggr. = aggression; Maso. = masochism; Alt. = altruism; Mart. = martyrdom; Sado. = sadomasochism.)

cooperative or altruistic chooser was significantly poorer than those who watched an individualistic or competitive chooser, $t(72) = 2.43$, $p < .02$. A second planned comparison, between conditions in which the chooser assumed a choice rule in the motivationally "normal" range (i.e., altruism, cooperation, individualism, competition, aggression) versus conditions in which the chooser's social motive might be considered abnormal (i.e., martyrdom, masochism, and sadomasochism), demonstrated the existence of a significant difference in favor of higher accuracy on the part of subjects who watched the former types of choosers, $t(72) = 5.30$, $p < .001$.

Examination of subjects' biases concerning the choosers' possible secondary motives was performed by first counting the number of times that subjects made correct predictions on the total of 10 trials in which the decomposed game situation gave the chooser two dominant choice alternatives, then determining how often each of these two choices was

Table 4
Percentage of Choices of Each Dominant Alternative in "Chooser Indifferent" Prediction Trials

Condition	No. correct predictions	Choice	Z value
Individualism	99	76% competition 24% cooperation	5.20**
Cooperation	106	73% individualism 27% altruism	4.74**
Altruism	81	63% cooperation 37% martyrdom	2.34*
Martyrdom	63	58% altruism 42% masochism	1.25
Masochism	79	60% martyrdom 40% sadomasochism	1.77
Sadomasochism	38	34% masochism 66% aggression	−1.97*
Aggression	97	17% sadomasochism 83% competition	−6.15**
Competition	107	32% aggression 68% individualism	−3.73**

* $p = .05$.
** $p = .01$.

predicted by the subjects. The results are shown in Table 4. From the table it is evident that there was a tendency on the part of subjects to predict that the chooser would prefer the alternative favoring self-interest as opposed to the other's interest when confronted by two choices equivalent in terms of their actual dominance.

To summarize, the prediction phase of Experiment 2 provided a general replication of Experiment 1, yielding three main findings. First, the farther the chooser's motive was from own or relative gain maximization, the less readily it tended to be detected. Second, there was no evidence to suggest that the atypicality of motivational vectors affording negative outcomes to self in any way facilitated their recognition. Finally, there appeared to be a general expectation among subjects that the chooser would behave in such a way as to further self-interest when faced with a choice between two presumably equally attractive alternatives, although this expectation became less pronounced the farther the chooser's actual choice rule was from individualism or competition.

Analysis of the ratings of choosers collected during the impression phase of the study also produced results basically similar to those of Experiment 1. A discriminant function analysis performed on the ratings resulted in a single significant function, accounting for 48.26% of the discriminatory power of the battery of derived functions, Wilks lambda(154) = .0359, $p < .001$.

Figure 6 indicates that choosers again roughly ordered themselves from self-centered to other-centered along the dimension defined by the discriminant function, but this time clustered into three rather than two groups. Choosers with the most negative discriminant scores—the individualistic, competitive, and aggressive choosers—were those who weighed their own gain more highly than the other's gain. The most positive discriminant scores were obtained by choosers who valued the other's gain above their own gain—the altruist, the masochist, and the martyr. Inter-

mediate between these two extremes were the sadomasochistic and cooperative choosers, who had in common the characteristic of treating self and other alike (and in the long run, equitably). (See Figure 7.)

This time the larger discriminant coefficients indicated that "self-centered" classification tended to be negatively associated with perceived politeness, passivity, generosity, and kindness, and positively associated with perceived agitation, other punitiveness, stability, selfishness, and justness. Differences between this function and the function obtained in Experiment 1 could be due to differences both in the sample of adjective scales used for rating the chooser and in the range of stimulus motives investigated in the two experiments.

Figure 7. Location of group centroids on the significant discriminant function, Experiment 2.

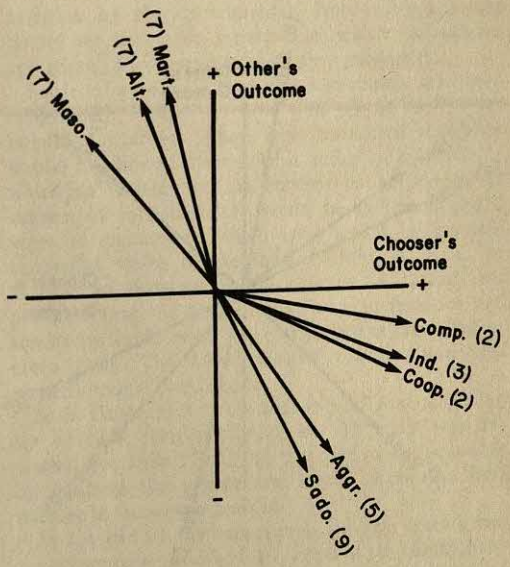


Figure 8. Mean observer-assigned vectors of inaccurate observers, Experiment 2. (*ns* are in parentheses. Maso. = masochism; Alt. = altruism; Mart. = martyrdom; Comp. = competition; Ind. = individualism; Coop. = cooperation; Aggr. = aggression; Sado. = sadomasochism.)

Finally, subjects' direct estimates of the weights that the choosers placed on their own and the other's gain were examined. Since it was anticipated that the ability to accurately estimate these weights might be related to a subject's success at predicting the chooser, weight estimates of accurate and inaccurate observers were considered separately. An accurate observer was defined as a subject who made at least 9 out of 12 correct predictions on the last prediction trial blocks. Subjects who made 8 or less correct predictions were classified as inaccurate. Mean observer-assigned weights were transformed by subtracting 5, to give the transformed scale a zero midpoint. Vectors derived from these mean assigned weights are shown in Figures 8 and 9 for inaccurate and accurate observers, respectively.

From Figures 8 and 9, it is evident that observers who accurately predicted the chooser on the last trial block were also more accurate in estimating the relative weights that the chooser placed on own and other's gain. For the accurate observers, the ordinal placement of the eight vectors corresponding to

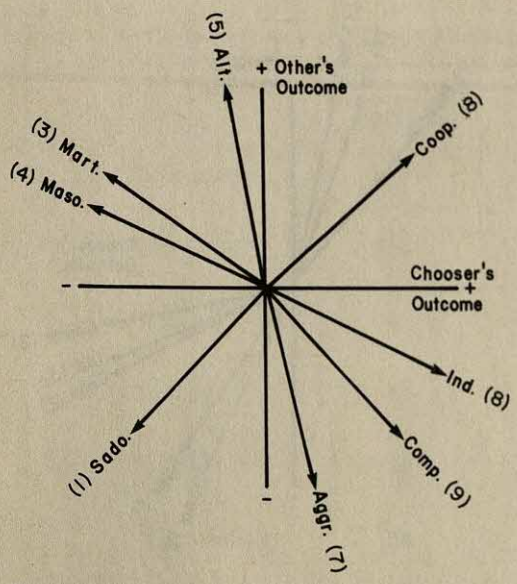


Figure 9. Mean observer-assigned vectors of accurate observers, Experiment 2. (*n*s are in parentheses. Sado. = sadomasochism; Maso. = masochism; Mart. = martyrdom; Alt. = altruism; Coop. = cooperation; Ind. = individualism; Comp. = competition; Aggr. = aggression.)

the eight chooser motivational orientations paralleled the correct order. However, several errors in the ordinal placement of these vectors occurred for the inaccurate observers. Both accurate and inaccurate observers exhibited a tendency to assimilate the relative weights assigned to the masochistic and the altruistic choosers toward martyrdom. Similarly, there was a tendency for the relative weights assigned to the aggressive and individualistic choosers to be assimilated towards competition. In the case of inaccurate observers, this latter assimilative tendency also held for the weight estimates of observers who viewed the cooperative and sadomasochistic choosers.

Prior to discussing the results of these first two studies, we will describe a third study that asked whether the rate of detecting a particular motivation vector was related to its mathematical properties.

Experiment 3

A few studies have investigated concept learning tasks that bear a resemblance to the present chooser prediction paradigm. In these experiments, the subject is typically required to make interval-level responses to a series of stimuli that vary on one or more attribute dimensions. In order to attain a reasonable level of accuracy, the subject must come to identify which of the stimulus attributes are relevant to the response, to correctly assign a relative weight to each of these attributes, and to identify the direction of the correlation between each relevant attribute and the response dimension.

The chooser prediction task of Experiments 1 and 2 is similar to an interval-concept learning problem, because subjects must decide whether one or both of two attributes, chooser's payoff and other's payoff, are relevant to determining the chooser's value for outcomes. Which of these attributes are actually important depends upon the motivational orientation of the chooser. To determine the value of outcomes for an individualistic chooser, for example, it is necessary to pay attention only to the chooser's payoff. However, both the chooser's payoff and the other's payoff must be taken into account when calculating the value of outcomes for a cooperative chooser. The relative weights placed on one's own outcome and the other's outcome and their signs (positive or negative) are shown in Table 5 for each of the chooser motivational orientation conditions of Experiment 2.

Experiment 3 was concerned with investigating the chooser prediction paradigm in the context of a nonsocial, interval-concept learn-

Table 5
Relative Weights Placed by Chooser on Own and Other's Outcomes, Experiment 2

Condition	Weight on own outcome	Weight on other's outcome
Individualism	+1	0
Cooperation	+1	+1
Altruism	0	+1
Martyrdom	−1	0
Masochism	−1	−1
Sadomasochism	−1	−1
Aggression	0	−1
Competition	+1	

ing task. In order to study just the mathematical component of the prediction task, references to a chooser making choices for self and other were deleted from the instructions and the decomposed game matrices.

The findings of previous studies indicate that subjects are better at interval-concept learning when relevant attributes are few as opposed to many (Azuma & Cronbach, 1966). In addition, variation in the relevant distribution of weights over attributes affects response accuracy. Uhl (1963) varied the predictive validities of three attributes and found a nonmonotonic relationship between subjects' response accuracy and relative weight disparity. Overall, subjects performed poorly when there was slight disparity among attribute weights, improved their accuracy when attributes were equally valid, and showed the best performance when there was a large discrepancy among weights. On the basis of these findings, it was hypothesized that subjects would find the nonsocial version of the chooser prediction task easier in conditions where one of the outcomes received a weight of zero than in conditions where both outcomes were equally important in determining an alternative's value.

Although it was recognized that the task involves making an ordinal comparison of alternatives to determine which is most highly valued, in addition to the interval judgment of the value of each alternative, this ordinal component was not expected to differentially affect accuracy across conditions in the nonsocial version of the task employed in this experiment.

Method

Subjects. The subjects were 64 introductory psychology students (13 males, 51 females) at the University of California, Santa Barbara. Eight subjects were randomly assigned to serve in each condition. All subjects received course credit for their participation.

Stimuli. A series of decomposed games identical to that used in Experiment 2 was employed. The situations were referred to as "matrices" in all instructions to subjects, and the words *Blue* and *Green* were substituted for the words *Chooser gets* and *Other gets* in the matrix display.

Procedure. Subjects arrived at the laboratory in pairs and were seated on the opposite sides of a partition by the experimenter. Both subjects could clearly see the video monitor on which instructions and stimuli for the experiment were presented.

Subjects were told that the experiment was concerned with finding out how well people could learn numerical concepts. They were instructed that they would be shown alternately a series of matrices in which an instance of the concept to be learned (a column of each matrix) would be indicated and a series of matrices in which they would be required to predict which of the four columns, A, B, C, or D, was an instance of the concept. A "hint" was provided that the column that was an instance had something to do with the entries in the blue and/or green rows. The terms *matrix, column,* and *row* were thoroughly explained.

As in Experiment 2, the sequence of first observing for 8 trials, then predicting for 12 trials was repeated five times. Subjects recorded their responses by marking the appropriate letter beside the trial number in an answer booklet.

At the end of the experiment, subjects were given a questionnaire designed to assess their comprehension of the concept. The questionnaire consisted of three parts. In the first part, subjects were asked to indicate whether for instances of the concept, the blue (green) numbers had tended to be larger, smaller, or neither larger nor smaller than the other blue (green) numbers in the matrix. Second, subjects were asked to write down the rule for deciding which column in a matrix was an instance. Finally, they were asked to fill in a blue number and a green number in a sample matrix where columns A, B, and C were given in such a way as to make column D an instance of the concept.

Results and Discussion

Table 6 gives the mean percentage of correct predictions across all trials for each condition in the present concept learning experiment. The figures for sister conditions in Experiment 2 are also shown.

A comparison between the Experiment 2 and the Experiment 3 results indicates that with the exception of the condition blue: −1–green: −1, the average accuracy for subjects in the social prediction task of Experiment 2 was higher (in some cases, quite substantially) than the average accuracy in the concept learning task.

Considering the accuracy figures for Experiment 3 alone, it is evident that the experimental hypothesis was not supported: There was no difference in performance on the task between conditions where either blue or green numbers received zero weight and conditions where both blue and green numbers

Table 6
Mean Percentage of Correct Predictions by Chooser Motive/Concept Condition, Experiments 2 and 3

Motive/concept	% correct predictions	
	Experiment 2 (motives)	Experiment 3 (concepts)
Individualism/blue: +1—green: 0	78.8	40.0
Cooperation/blue: +1—green: +1	64.1	37.5
Altruism/blue: 0—green: +1	58.8	30.7
Martyrdom/blue: −1—green: +1	41.1	31.3
Masochism/blue: −1—green: 0	51.1	34.1
Sadomasochism/blue: −1—green: −1	28.8	32.7
Aggression/blue: 0—green: −1	65.0	32.1
Competition/blue: +1—green: −1	74.7	38.3
Chance	29.2	29.2

had nonzero weight. In support of this conclusion, an analysis of variance performed on these data after transforming them by the arc sine procedure yielded no significant main effect for condition, $F(7, 56) < 1$. A statistically demonstrable improvement in accuracy did occur across trial blocks, however, $F(4, 224) = 7.689$, $p < .001$.

The results obtained from the questionnaire administered at the end of the experiment indicated that few of the subjects in the experiment ever grasped the concept they were trying to learn. (Since the pattern of results appeared to be uninfluenced by condition, it will be described for all subjects.) Of the total of 64 subjects, 42, or approximately two thirds, failed to answer any part of the questionnaire correctly. Of the remaining 22 subjects, 19 were able to correctly generate an instance of the concept. However, because a large range of answers was potentially acceptable, it is not possible to determine how many of these 19 correct instances could be attributed to guessing. Subjects' performances were much poorer on the other two parts of the questionnaire than on the instance-generation question. Only 8 subjects managed to verbalize the rule, and only 6 subjects correctly identified the relative weights of both the blue and green numbers. In total, only 3 subjects answered the entire questionnaire correctly.

Thus, the results of the present experiment do not suggest that a relationship existed between the numerical complexity of the chooser's-choice rule and the subject's ability to predict the chooser in a social context. In fact, it appears safer to conclude the opposite: that the social context of the prediction task in the earlier studies facilitated learning of the mathematical aspects of the chooser's-choice rule.

General Discussion

The results of Experiments 1 and 2 have confirmed that variations in an actor's preference for outcomes affect the overall accuracy with which observers are able to predict the actor's choice behavior in a series of decomposed games. Although the preprogrammed choosers employed in these experiments exhibited perfect consistency in making selections from among sets of own/other outcome distributions, subjects nevertheless experienced difficulty in predicting the chooser's choices when he or she behaved in a negatively self-interested or prosocial fashion.

These data seem best interpreted from within an interaction-expectancy framework: The more popular a particular social motive is perceived to be as an interaction goal, the more likely it will be anticipated (and detected) in the behavior of others. Support for this assertion was obtained both from the accuracy data and from the existence of a self-interest bias in subjects' predictions of what the chooser would do when confronted with

paired rather than single alternatives dominant with respect to the chooser's motivational vector.

Expectations concerning the chooser's likely pursuit of self-interested goals appear to have been highly pervasive in influencing subjects' responses throughout the experiments. An examination of the performance of individual subjects indicated that most of them improved their accuracy in an incremental fashion rather than by the abrupt transition that might have occurred if the subjects had been systematically testing a range of alternative hypotheses about the chooser's motive. Subjects may have been reluctant to abandon a self-interest hypothesis even when it was less than perfectly valid. Confounding of motives within the choice-dominance structures of the decomposed games would allow subjects who held certain wrong hypotheses about the chooser (i.e., subjects who believed the chooser's motivational vector to be within 90° of his or her true vector) to still make some correct responses. Exacerbated by a tendency to pay more attention to instances of the chooser's behavior that confirmed a preference for self-interest than instances that provided disconfirmation (see Bourne, Ekstrand, & Dominowski, 1971), this explanation might account for some of subjects' conservatism in utilizing the information contained in the decomposed game choices made by prosocial and negatively self-interested choosers.

As we noted previously, it is curious that differences emerged in subjects' abilities to predict self-interested and prosocial choosers, considering that fairly distinctive impressions of the personality attributes of these two groups of choosers were formed. Researchers in the area of person perception have often stressed the importance of personality impressions, particularly first impressions, in determining our reactions toward others (Hastorf, Schneider, & Polefka, 1970). However, in Experiments 1 and 2 the impressions that were formed of prosocial and negatively self-interested choosers were apparently of little value to subjects in predicting these choosers' choice behaviors, even though choices to be predicted were of the same type as those upon which impressions were based.

One possible explanation for this apparent incongruity is that subjects' ratings of the choosers' personality attributes may have been retrospective reflections upon the choosers' choices, as opposed to impressions formed early in the course of observing the choosers. Whether the personality attribute ratings represent early or retrospective impressions, however, it is interesting to note the existence of groupings of motives in terms of both the discriminant function analysis of these ratings and the vector representations derived from the subject-estimated weights of Experiment 2. These groupings suggest that the psychological representation of motivational vectors may differ from their formal representation in terms of perceived angular distance, if not in their ordering.

Finally, Experiment 3 investigated the difficulty of learning various choice rules consistent with certain social motives outside of a social context. No evidence was obtained to suggest that the mathematical complexity of a given choice rule had any influence on increasing or decreasing an observer's ability to predict. Conversely, identifying the rule-detection task as a social one positively, though differentially, affected the likelihood of correct detection.

A major limitation of the present investigations is that they have provided little information about the factors that mediate the observed relationship between the chooser's social motive and the subjects' predictive accuracy. While observer expectations about the probability that an actor will adopt particular social motives provide an overall explantion for the results, other influences may have made important contributions to the present findings. For instance, the social motivational orientations of observers may have been a significant factor in determining their relative success at the prediction task. Subjects in the experiments described here exhibited a considerable degree of individual difference in their abilities to predict the various motives. If interaction expectancies are affected by one's own social motive, as the work of Kuhlman and Wimberley (1976)

would suggest, then it follows that this motive would influence the ability to detect others' social motives, especially others with motives identical to the observer's own orientation. Future research will consider this relationship.

Reference Note

1. Kelley, H. H. *Action and perception: An attribution analysis of social interaction.* The Katz-Newcomb Lecture, University of Michigan, 18 April 1975.

References

Azuma, H., & Cronbach, L. J. Cue response correlations and the attainment of a scalar concept. *American Journal of Psychology,* 1966, *79,* 38–49.

Bourne, L. E., Ekstrand, B. R., & Dominowski, R. L. *The psychology of thinking.* Englewood Cliffs, N.J.: Prentice-Hall, 1971.

Davis, G. R. *The attribution of social motives in decomposed games.* Unpublished doctoral dissertation, University of California, Santa Barbara, 1972.

Griesinger, D. W., & Livingston, J. W. Toward a model of interpersonal motivation in experimental games. *Behavioral Science,* 1973, *18,* 173–188.

Hastorf, A. H., Schneider, D., & Polefka, J. *Person perception.* Reading, Mass.: Addison-Wesley, 1970.

Heider, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

Jones, E. S., & Davis, K. E. From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2). New York: Academic Press, 1965.

Kelley, H. H., & Stahelski, A. J. The social interaction basis of cooperators' and competitors' beliefs about others. *Journal of Personality and Social Psychology,* 1970, *16,* 66–91.

Kuhlman, D. M., & Marshello, A. Individual differences in game motivation as moderators of preprogrammed strategy effects in Prisoner's Dilemma. *Journal of Personality and Social Psychology,* 1975, *32,* 922–931.

Kuhlman, D. M., & Wimberley, D. C. Expectations of choice behavior held by cooperators, competitors, and individualists across four classes of experimental game. *Journal of Personality and Social Psychology,* 1976, *34,* 69–81.

Kuusinen, J. Affective and denotative structures of personality ratings. *Journal of Personality and Social Psychology,* 1969, *12,* 181–188.

MacCrimmon, K. R., & Messick, D. M. A framework for social motives. *Behavioral Science,* 1976, *21,* 86–100.

McClintock, C. G. Game behavior and social motivation in interpersonal settings. In C. G. McClintock (Ed.), *Experimental social psychology.* New York: Holt, Rinehart & Winston, 1972. (a)

McClintock, C. G. Social motivation—A set of propositions. *Behavioral Science,* 1972, *17,* 438–454. (b)

McClintock, C. G., & McNeel, S. P. Reward and score feedback as determinants of cooperative and competitive game behavior. *Journal of Personality and Social Psychology,* 1966, *4,* 606–613.

McClintock, C. G., Messick, D. M., Kuhlman, D. M., & Campos, F. T. Motivational bases of choice in three-choice decomposed games. *Journal of Experimental Social Psychology,* 1973, *9,* 572–590.

McClintock, C. G., & Moskowitz, J. M. Children's preference for individualistic, cooperative, and competitive outcomes. *Journal of Personality and Social Psychology,* 1976, *34,* 543–555.

Messick, D. M., & McClintock, C. G. Motivational bases of choice in experimental games. *Journal of Experimental Social Psychology,* 1968, *4,* 1–25.

Osgood, C., Suci, G., & Tannenbaum, P. *The measurement of meaning.* Chicago: University of Illinois Press, 1957.

Rapoport, A. *Two-person game theory: The essential ideas.* Ann Arbor: University of Michigan Press, 1966.

Ross, L. The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10). New York: Academic Press, 1977.

Tatsuoka, M. M. Discriminant analysis: The study of group differences. *Selected topics in advanced statistics: An elementary approach* (No. 6). Champaign, Ill.: Institute for Personality and Ability Testing, 1970.

Thibaut, J. W., & Kelley, H. H. *The social psychology of groups.* New York: Wiley, 1959.

Uhl, C. N. Learning of interval concepts: 1. Effects of differences in stimulus weights. *Journal of Experimental Psychology,* 1963, *66,* 264–273.

Received February 8, 1978 ■

urnal of Personality and Social Psychology
79, Vol. 37, No. 2, 221-233

Effects of External Evaluation on Artistic Creativity

Teresa M. Amabile
Brandeis University

This study examined the conditions under which the imposition of an extrinsic constraint upon performance of an activity can lead to decrements in creativity. Female college students worked on an art activity either with or without the expectation of external evaluation. In addition, subjects were asked to focus on either the creative or the technical aspects of the activity, or they were given no specific focus. Finally, some subjects expecting evaluation were given explicit instructions on how to make their artworks. As predicted, subjects in the evaluation groups produced artworks significantly lower on judged creativity than did subjects in the nonevaluation control groups. The only evaluation group for which this pattern was reversed had received explicit instructions on how to make artworks that would be judged creative. A possible reconciliation of these two disparate results is proposed, and practical implications are discussed.

In his autobiography, Albert Einstein describes a serious motivational problem he encountered during his student days: His physics examinations, forcing him to "cram all this stuff into one's mind" (Schilpp, 1949, p. 17), were so unpleasant that afterward he could not bring himself to consider scientific problems for an entire year. When he went on to advanced study in Zurich, he found ways to blunt the effects of educational constraint, "which smothers every truly scientific impulse" (p. 17). For example, he had a friend who agreed to work over the lecture materials so that Einstein would be freed from attending classes. In commenting upon this arrangement and its boost to his creativity motivation, Einstein later said,

> This gave one freedom in the choice of pursuits until a few months before the examination, a freedom which I enjoyed to a great extent and have gladly taken into the bargain the bad conscience connected with it as by far the lesser evil. It is, in fact, nothing short of a miracle that the modern methods of instruction have not yet entirely strangled the holy curiosity of inquiry; for this delicate little plant, aside from stimulation, stands mainly in need of freedom; without this it goes to wreck and ruin without fail. (p. 17)

Einstein's introspections and speculations about scientific inquiry are an elegant expression of the thesis to be advanced here: An intrinsically motivated state is conducive to creativity, whereas an extrinsically motivated state is detrimental. That is, if individuals engage in some activity primarily for its own sake, they will be most likely to produce creative work. If, however, they are led to engage in that activity as a means to achieve some salient extrinsic goal, their creative performance will be undermined.

This article is based on a doctoral dissertation submitted to the Department of Psychology, Stanford University. The research reported here was supported by Predoctoral Fellowship 5F31 MH05232-02, awarded to the author by the National Institute of Mental Health, and by two research grants to Mark R. Lepper: National Institute of Mental Health Grant MH-24134 and National Institute of Child Health and Human Development Grant HD-MH-09814.

The author gratefully acknowledges the advice and support of her dissertation committee members, Mark Lepper, Lee Ross, Daryl Bem, and Philip Zimbardo, and she wishes to thank several colleagues who offered comments and suggestions during various stages of the research and manuscript preparation: William DeJong, Susan Goldberg, Leslie McArthur, and Mick Watson.

Portions of this study were presented at the 49th annual meeting of the Eastern Psychological Association, Washington, D.C., March 1978.

Requests for reprints should be sent to Teresa M. Amabile, Department of Psychology, Brandeis University, Waltham, Massachusetts 02154.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3702-0221$00.75

The present conceptualization of creativity proposes that intrinsically motivated individuals will be deeply involved in the activity at hand because they will be free of extraneous and irrelevant concerns, concerns about goals extrinsic to the activity itself. They will be playful with ideas and materials because of their freedom to take risks, to explore new cognitive pathways, to engage in behaviors that might not be directly pertinent to attaining a "solution." Since they undertook the activity primarily for the enjoyment of engaging in it, they will see the activity as more like play than like work. Extrinsically motivated individuals, on the other hand, will be, at some level, concerned with the extrinsic goal to be attained and will thus not be as deeply involved in the activity. In addition, they will feel less free to engage in risk taking and will therefore rely more upon well-worn cognitive pathways. Finally, since they undertook the activity primarily for some reason extraneous to the activity itself, they will see it as more like work than like play.

Most recent intrinsic motivation research has been concerned with the "overjustification" hypothesis (Lepper, Greene, & Nisbett, 1973), derived from the attribution theories of Bem (1967, 1972), Kelley (1967, 1973), and deCharms (1968). This hypothesis states that if a person undertakes an interesting task under conditions that make salient to him the instrumentality of his behavior as a means to some extrinsic end, then that person will show less intrinsic interest in that activity later, when external constraints are absent, than a person who did not act under salient external constraints. Using the constraint-free measure of subsequent intrinsic motivation, results from a number of studies have supported this hypothesis (Amabile, DeJong, & Lepper, 1976; Condry, 1977; Deci, 1971; Kruglanski, 1975; Lepper & Greene, 1975; Lepper et al., 1973).

Recently, several theorists have begun to speculate about the effects of extrinsic constraint on *immediate* performance. In the formulation that is most relevant to the present research, McGraw (1978) has proposed a distinction between two different types of activities in terms of the differential effects that extrinsic constraint might have upon performance of those activities. McGraw describes tasks having *algorithmic* solutions as those for which the path to the solution is clear and straightforward; performance on these tasks should be enhanced by increases in extrinsic motivation. By contrast, creativity tasks require *heuristic* solutions, where it is difficult to immediately determine which operations would be relevant to a solution. Thus, creative performance should be adversely affected by increases in extrinsic motivation.

Direct Empirical Evidence

There is a handful of studies directly testing the hypothesis that decrements in creativity will accompany the imposition of extrinsic constraints. Of these, one was an overjustification study that included a measure of creativity in assessing the effects of extrinsic constraints (Kruglanski, Friedman, & Zeevi, 1971). In this experiment, one group of students was promised a reward for participation; for a second group, reward was not mentioned. Rewarded subjects produced less creative responses (as judged by two independent raters) during the experiment than did nonrewarded subjects; in addition, nonrewarded subjects later expressed greater enjoyment of the experiment.

Other studies provide further support for the proposed relationship between creativity and intrinsic motivation. In two experiments in which children made drawings under either reward expectation or no reward expectation (Greene & Lepper, 1974; Lepper et al., 1973), there was a tendency for rewarded children to produce more drawings, but of poorer quality (as judged by teachers), than nonrewarded children. A study of the effects of rewards on problem-solving performance (McGraw & McCullers, in press) found that rewarded subjects took significantly longer to break set in solving a Luchins water jar problem than did nonrewarded subjects. And in an elementary school setting (White & Owen, 1970), the creative performance of boys in a self-evaluation group was significantly higher than that of boys in peer-evaluation or teacher-evaluation groups.

While this research seems to support the intrinsic motivation view of creativity, studies within the behavior modification or token-economy traditions have obtained results that appear to contradict it. In one such study (Glover & Gary, 1976), children worked on a standard verbal creativity task in which points were awarded for fluency (number of different responses), flexibility (number of verb forms), elaboration (number of words per response), or originality (statistical infrequency of verb forms). Consistent with the experimental hypotheses, all four aspects of creativity were "demonstrated to be under experimental control" (p. 79); when fluency was rewarded, the children were fluent; when originality was rewarded, they were original; and so on. Under extinction, each aspect fell to baseline or below. Other studies of operant techniques, using both intersubject and between-group designs, have demonstrated functional control over creative performance (Halpin & Halpin, 1973; Johnson, 1974; Raina, 1968).

Thus, a cursory review of evidence on the effects of extrinsic constraint on creativity suggests a basic inconsistency in results. Overjustification studies have generally shown decrements in creativity under reward conditions, whereas behavior modification studies have shown increments under such conditions. However, as the results of the present study will suggest, a deeper analysis of the two paradigms leads to a possible resolution of the apparent contradiction.

The Present Study

A closer examination of the procedures used in the overjustification and behavior modification studies suggests that the key to their apparently contradictory results may lie in the type of instructions used. As suggested by the theoretical analysis of McGraw (1978), perhaps subjects will show decrements in creativity if extrinsic constraints are imposed, *unless* they are told specifically how to perform creatively. In order to examine this hypothesis, the present study attempted, within one experimental design, to show both decrements in creativity by an overjustification procedure *and* increments in creativity by a behavior modification procedure. In the former, a constraint was imposed upon subjects' engagement in an art activity, and they were given no specific performance instructions; in the latter, the same constraint was imposed upon subjects, and they were told specifically how to perform creatively. In accord with the intrinsic motivation model of creativity, it was expected that the simple overjustification groups—the groups working under extrinsic constraint, without explicit instructions on how to perform creatively—would show decrements, relative to control, in both creativity and intrinsic interest. And, as required by the theoretical argument developed here, it was expected that, although the specific-instructions group might show an increment in creativity compared to its control, it should not show a corresponding increment in intrinsic motivation. Thus, the major purpose of this study was to attempt a reconciliation of seemingly contradictory results by identifying those instructional sets under which extrinsic constraint might undermine creativity and those under which it might enhance creativity.

The specific extrinsic constraint employed in this study was the expectation of evaluation. Clearly, this constitutes an externally imposed constraint that is extrinsic to the activity itself; the external evaluation of an art product is not intrinsic to the activity of art itself. The use of this particular constraint is desirable for two reasons. First, expectation of evaluation has seldom, if ever, been used in previous overjustification research. As noted earlier, the extrinsic constraint usually employed is the promise of reward. Second, and most important, the external evaluation of products that may potentially be creative is so commonly employed in education and other settings as to be accepted as a fact of life. Thus, a demonstration of detrimental effects of evaluation could have significant practical implications.

Method

Design Overview

The experimental design employed here closely followed the usual overjustification paradigm: Subjects were either told or not told that their artworks

would be evaluated. Additionally, some subjects within each level of evaluation expectation were asked to focus on the creativity of their artworks; some were asked to focus on the technical aspects; others were not given any particular focus. Finally, two evaluation-expectation groups received specific instructions for performance: One was told how to make a technically good artwork, and one was told how to make a creative artwork.

Subjects

Subjects were 95 women enrolled in the introductory psychology course at Stanford University.[1] They signed up for an experiment entitled "Effects of Various Activities on Mood" in partial fulfillment of a course requirement. The experiment was run by one female experimenter.

Experimental Task

Most of the more widely used tests of creativity are amenable to showing individual differences that depend on certain specialized skills, such as verbal facility and flexibility or drawing ability. Since the present study attempted to show differences between groups of subjects on the basis of environmental manipulations, it was desirable to find a method of assessing creativity that would minimize individual differences in performance. For the purposes of the present investigation, then, a simple subjective method of assessing creativity was developed, a method that does not depend on specialized skills and yet permits reliable judgments of creativity. In addition, it was desirable, for the purposes of differentiating between creativity and other performance effects, that judgments of creativity be separable from judgments of "goodness" and liking for the products.

A study investigating the adequacy of an art activity for this task was conducted as a pretest. Using colored paper and glue, subjects made collages that were subsequently rated on several artistic dimensions by artist-judges. These judgments, over seven judges, were quite reliable; the interjudge reliability was .77 for creativity ratings and .72 for technical goodness ratings. Moreover, a factor analysis revealed two orthogonal factors, one composed of creativity and related judgments (such as novelty of idea and complexity) and one composed of technical goodness and related judgments (such as organization and neatness). In addition, as desired, judgments of liking and aesthetic appeal loaded weakly on both of these factors. Thus, this collage-making activity fits the criteria proposed earlier: It depends on skills no more specialized than manipulating paper and glue; it permits reliable subjective judgments of creativity and the other artistic dimensions; and judgments of technical competence and aesthetic appeal can be clearly separated from judgments of creativity.

Procedure

After asking the subject to complete a consent form, the experimenter explained that this study was actually a pretest for an experiment to be done the following quarter. This alleged pretest was being run in order to identify activities that would affect most individuals' moods in certain, predictable ways. The subject was told that she would randomly choose one of five different activities, and after engaging in that activity, she would complete a mood questionnaire.

At this point, the experimenter presented five numbered cards to the subject, turned them face down, shuffled them, and asked the subject to choose one. After she had done so, the experimenter consulted a small "activity card" and announced, "Number ____ is . . . the art activity." The experimenter then presented the subject with her materials: one white cardboard, 15 × 20 inches (.38 × .51 m); one small plastic squeeze-bottle of glue; and one large round aluminum tray containing 154 colored pieces of lightweight paper in various shapes and sizes. After presenting the materials, the experimenter asked the subject if she was an artist or had done much collage work. If the subject answered in the affirmative, the session was terminated. (Only three subjects were eliminated on this basis, and they were replaced.) If not, the experimenter went on to introduce the subject's task. In so doing, she stressed that the subject had complete freedom in using the materials to form a design, but that only the materials provided should be used. In addition, the subject was asked to make a design that conveyed a feeling of silliness, "as when a child is acting and feeling silly." This last instruction was given to insure that all subjects would employ the same theme in making their designs, in an effort to reduce extraneous sources of variability.

Following this general introduction, the instructions diverged to produce the eight treatment conditions. (Table 1 summarizes the experimental design.) Assignment of subjects to conditions had been predetermined by a randomized schedule, and up to this point, the experimenter was unaware of the subject's treatment condition. Now, the experimenter turned to the "manipulation" page in the subject's packet and read the crucial instructions. Subjects in the control conditions (nonevaluation – no focus,

---

[1] Males were not used, in an effort to reduce extraneous sources of variability. Pretesting had shown that on this task, females produce artworks that are judged as significantly more creative than those produced by males. The number of subjects was approximately equal in each of the eight conditions: nonevaluation – no focus = 13, nonevaluation – technical focus = 12, nonevaluation – creativity focus = 11, evaluation – technical focus = 11, evaluation – creativity focus = 11, evaluation – specific technical focus = 12, evaluation – specific creativity focus = 12.

Table 1
Experimental Design

Evaluation expectation	Instruction				
	No focus	Technical focus	Specific technical focus	Creativity focus	Specific creativity focus
Absent	C_1	C_2		C_3	
Present	E_1	E_2	E_2	E_3	E_3

Note. The letter in each box indicates whether a condition is control or experimental, and the subscripts link experimental groups to their appropriate controls.

nonevaluation – technical focus, and nonevaluation – creativity focus) were told:

> There is one more important point that I should make before you begin. We won't be using your design as a source of data. We are not interested at all in the activity itself or what you do with the activity. We are *only* interested in the mood you report on the questionnaire. So we do not care about the design itself at all—its only purpose is to provide you with this experience so we can see how it affects your mood.

In addition, subjects in the nonevaluation – technical focus condition were asked to concentrate on the technical aspects of the activity "for this particular mood induction," and subjects in the nonevaluation – creativity focus condition were asked to focus on the creative aspects.

The basic instructions for the five experimental groups (evaluation – no focus, evaluation – technical focus, evaluation – creativity focus, evaluation – specific technical focus, evaluation – specific creativity focus) were:

> There is one more important point that I should make clear before you begin. In addition to your questionnaire, we will be looking at your finished design as an important source of data. We have five graduate artists from the Stanford Art Department working with us, and when this experiment is over, we will have them come in to judge each artwork. They will make a detailed evaluation of your design, noting the good points and criticizing the weaknesses. And since we know that our subjects are interested in how they were evaluated, we will send you a copy of each judge's evaluation of your design in about two weeks.

In addition, subjects in the evaluation – technical focus condition were told that the judges would base their evaluation on how technically good the designs were; subjects in the evaluation – creativity focus condition were told that the judges would base their evaluations on how creative the designs were. Those in the evaluation – specific technical focus condition were told that the judges would make their technical evaluation on the basis of these six elements: (a) the neatness of the design, (b) the balance of the design, (c) the amount of planning evident, (d) the level of organization in the design, (e) the presence of actual recognizable figures or objects in the design, and (f) the degree to which the design expresses something to them. Subjects in the evaluation – specific creativity focus condition were told that the judges would base their creativity evaluation on these seven elements: (a) the novelty of the idea, (b) the novelty shown in use of the materials, (c) the amount of variation in the shapes used, (d) how asymmetrical the design is, (e) the amount of detail in the design, (f) the complexity of the design, and (g) the amount of effort evident. These components were the ones that had, in fact, clustered closely with pretest judges' ratings of technical goodness and creativity, respectively.

After asking the subject if she had any questions, the experimenter looked at the digital clock that had been placed in the room and said, "Okay, it's (time) right now; I'll be back in 15 minutes." She then left the room.

When the 15 minutes had elapsed, the experimenter reentered the experimental room, asked the subject to stop working, and—in keeping with the cover story—presented her with a "Mood Questionnaire." This questionnaire contained 15 mood adjectives and asked the subject to rate herself on each. This was the first of two questionnaires presented to the subject. The second, an "Art Activity Questionnaire," included a number of items designed to assess the subject's interest in and attitude toward the art activity. After completing this questionnaire, the subject was asked a series of questions designed to probe suspicions about the experimental situation, the tasks, or the instructions.[2] The experimenter then fully explained the purposes and hypotheses of the study,

[2] Fifteen subjects, distributed evenly throughout the eight conditions, expressed suspicions about being watched during a 10-minute free-time period that elapsed between administration of the two questionnaires. Since the behavioral measure obtained during that period proved not to be useful, these subjects were retained, and their data appear in all analyses reported here.

noting the subject's reactions and comments. No subject appeared distressed by the deception that had been employed, and all expressed interest in the hypotheses and outcomes of the study.

Dependent Measures

Creativity. The creativity of the designs was the dependent measure of primary interest in this study. Creativity and several other artistic dimensions (including technical goodness) were assessed according to the judging procedure developed during pretesting. Fifteen artists, nine males and six females, served as judges of the designs made by subjects in this study. Each judge had had at least 5 years of experience doing studio art (painting, drawing, or design); most were graduate students enrolled in the Stanford Art Department. These artists were paid for their participation.

The judges worked in individual sessions, rating the 95 designs, which had been affixed to the walls of a lab room. A different random ordering of the designs was used for each judge. Before beginning a judging session, the experimenter gave the judge a brief summary of the general instructions the subjects had received, *without* mentioning the crucial evaluation/nonevaluation instructions or the particular instructions focus that some subjects had received. The judge was shown a sample of the materials that each subject had worked with, as well as a sample of the continuous scale (with 5 reference points marked) on which all ratings would be made.

The judge was told that he or she would judge the 95 designs (always in the same order for a particular judge) on each of 16 different artistic dimensions: (a) expression of meaning, (b) degree of representationalism, (c) silliness, (d) detail, (e) degree of symmetry, (f) planning evident, (g) novelty of the idea, (h) balance, (i) novelty in use of materials, (j) variation of shapes, (k) effort evident, (l) complexity, (m) neatness, (d) overall organization, (o) creativity, and (p) technical goodness. The judge was left alone to do each rating task, which consisted of rating all 95 designs on one of the 16 artistic dimensions. Before rating the designs on any given dimension, the judge was given a brief written definition of that dimension.[3] A different random ordering of the 16 dimensions to be judged was used for each of the 15 judges.

In essence, the rating of the designs can be viewed as a discrimination task for the judges. It was expected that on many dimensions of judgment, the judges would, on the average, see differences between the artworks made by the nonevaluation groups and the artworks made by the evaluation groups. Could the judges, for example, see the artworks of the nonevaluation – no focus group as more creative than those of the evaluation – no focus group? Viewed in this way, the experimental design employed here was a repeated measures design, with judges as "subjects"; for any one artistic dimension, each judge provided data on each of the eight experimental conditions.

Other measures. As noted earlier, a questionnaire designed to assess interest in the art activity was administered to subjects at the end of the experimental session. This questionnaire contained items that, in previous research (Amabile et al., 1976), had served as indicators of intrinsic interest. In addition, an attempt was made to obtain a behavioral measure of interest in the activity. Subjects were left alone in the experimental room for a period of 10 minutes between administration of the Mood Questionnaire and administration of the Art Activity Questionnaire. Some collage-making materials and pieces from a structure-building game had been left on another table in the room, and the amount of time subjects spent playing with these activities during this free time was observed from behind a one-way mirror. It was expected that this behavioral measure would provide results analogous to those obtained on the intrinsic interest questionnaire. The behavioral measure proved not to be useful, however, since few subjects in any condition played with either set of materials during the free-time session.

Results

Preliminary Analyses

Reliability. As a first step in analysis of the creativity results, it was essential to determine that the subjective ratings made by artist-judges were reliable at an acceptable level. Thus, a Spearman–Brown interjudge reliability was calculated for each dimension of judgment (see Nunnally, 1967, p. 233). In general, the reliabilities calculated in this manner were quite high: 12 of the 16 dimension reliabilities were above .80, and the median reliability was .84. Of particular interest is the reliability of the major dependent measure for this study, creativity; this value was quite satisfactory at .79. Only one dimension of judgment, balance, failed to reach an acceptable reliability level; this reliability, at .48, was so much lower than the others that judgments of balance were excluded from further analyses.

These generally high reliabilities bear a special significance, given the nature of the judgment task employed in this study. The

[3] These definitions were all nondirective. For example, the creativity "definition" was, "Using your own, subjective definition of creativity, the degree to which the design is creative."

judges were asked to perform a particularly difficult feat: to work alone in a small room for nearly 3 hours, viewing 95 designs affixed to the walls, making 1,520 separate judgments, and spending barely 5 sec on each one. Given these circumstances, and given particularly that the judges were not trained in any way to agree with one another on this task, their high level of agreement is all the more encouraging.

Factor analysis. A factor analysis (varimax rotation) was performed on the dimensions of judgment to determine if the eight "creativity" dimensions and the seven "technical" dimensions did in fact cluster together.[4] The results of this analysis suggest that the judgments did cluster in this fashion. First, there are clearly two orthogonal factors. Moreover, these were the two primary factors identified in the analysis, and they do indeed seem to be a creativity factor and a technical factor. Seven of the eight original creativity dimensions (creativity, novelty of material use, novelty of idea, effort evident, variation of shapes, detail, and complexity) load high and positive on the creativity factor; all of these loadings are +.80 or higher on that factor and ±.30 or lower on the technical factor. In addition, five of the seven original technical dimensions (organization, neatness, planning evident, balance, and expression of meaning) load high and positive on the technical goodness factor; all of these loadings are +.45 or higher on this factor and ±.16 or lower on the creativity factor.

The results of this analysis differed from those of the pretest in three minor respects. Two component dimensions that had previously loaded high on their respective factors (asymmetry on the creativity factor and representationalism on the technical factor) loaded sufficiently low on those factors here that they were discarded from further analyses. In addition, judges' ratings of "technical goodness" as a single dimension did not, in this study, load any higher on the technical factor than on the creativity factor.[5] Thus, the composite technical goodness measure used in subsequent analyses did not include the single technical goodness dimension of judgment.

Creativity

It was expected that with the exception of the specific creativity instructions group, the artworks produced by subjects who expected evaluation would be judged lower on creativity than the artworks produced by subjects who did not expect evaluation. Judge ratings on creativity and the six creativity component dimensions strongly support this hypothesis.

A composite creativity measure was formed by combining the normalized rating for each of the creativity dimensions: novelty of material use, novelty of idea, effort evident, variation of shapes, detail, and complexity.[6] Means for this composite measure are presented in Table 2.

An overall analysis of variance for the seven groups excluding the specific creativity instructions group (evaluation – specific crea-

---

[4] The factor analysis was done on the average ratings (over all judges) for each of the 95 designs on each of the 16 dimensions.

[5] This result is not surprising, since, in this study, artists' judgments of technical goodness correlated .68 with their creativity judgments. However, it clearly implies that while the technical goodness cluster of judgments on the factor analysis *does* capture the technical aspect of the artworks, the single technical goodness dimension does not. It appears that, unlike the artist-judges in the pretest study, the judges in the present study did not assess simply the *technical* goodness of the designs with their ratings on this one dimension, but an overall Gestalt of "goodness," which included aspects of creativity.

[6] Since the experimental design was treated as a repeated measures design for the creativity and technical goodness measures, with judges as subjects, each subjective measure was analyzed by first obtaining one number for each judge on each of the eight experimental conditions (an average for each condition) and then carrying out repeated measures analyses of variance, dependent t tests, and any additional analyses on these. Thus, for example, the creativity composite scores were formed by first normalizing all judgments on each of the six creativity components and then averaging the normalized ratings for each judge on each component for each condition. Finally, one number was obtained for each judge on each condition by forming a composite of the six components (a simple average). By using this technique of analysis for subjective measures, it is possible to examine overall rated differences between conditions and, at the same time, to account for differential scale usage by allowing each judge to act as his or her own control.

Table 2
Mean Judge Ratings of Creativity

Evaluation expectation	Instruction				
	No focus	Technical focus	Specific technical focus	Creativity focus	Specific creativity focus
Absent	.356	.181		.160	
Present	−.499	−.056	−.466	−.472	.777

Note. These numbers are the means of composites of six normalized components of creativity that clustered on the factor analysis of artist judgments.

tivity focus) was statistically significant, $F(6, 84) = 12.13$, $p < .001$. Thus, a planned contrast was performed on these seven groups to test the hypothesis that control groups (nonevaluation) were judged higher on creativity than experimental groups (evaluation). This contrast was clearly significant, $F(1, 84) = 45.81$, $p < .001$.

This pattern is borne out by a series of paired comparisons between control groups and the relevant experimental groups. As expected, only when evaluation subjects were given specific instructions on how to make a creative design did they produce artworks judged as significantly more creative than those of nonevaluation subjects. The mean rated creativity for this specific instructions group (evaluation – specific creativity focus) is significantly higher than that of the relevant control, $t(14) = -3.88$, $p < .01$;[7] indeed, this group is higher than any other on judged creativity. In all other cases, the nonevaluation groups are significantly higher on judged creativity than the comparable evaluation groups: For the no focus groups (nonevaluation – no focus vs. evaluation – no focus), $t(14) = 9.44$, $p < .001$; for the technical focus groups (nonevaluation – technical focus vs. evaluation – technical focus), $t(14) = 2.07$, $p < .06$; (nonevaluation – technical focus vs. evaluation – specific technical focus), $t(14) = 3.62$, $p < .01$; and for the creativity focus groups (nonevaluation – creativity focus vs. evaluation – creativity focus), $t(14) = 3.79$, $p < .01$.[8]

Technical Goodness

The technical goodness of the artworks—the degree of technical competence displayed by subjects in their work—was chosen as a feature that might be compared with creativity. Means for the composite technical goodness measure (which included organization, neatness, planning, and expression of meaning) are presented in Table 3. As can be seen, for the no focus groups (nonevaluation – no focus and evaluation – no focus) and the creativity focus groups (nonevaluation – creativity focus, evaluation – creativity focus, and evaluation – specific creativity focus), the evaluation groups are rated lower on technical goodness than their nonevaluation controls. Paired comparisons bear out this observation: The nonevaluation – no focus group is rated significantly higher than the evaluation – no focus group, $t(14) = 3.16$, $p < .01$, and the nonevaluation – creativity focus group is rated higher than both the evaluation – creativity focus group, $t(14) = 6.16$, $p < .01$, and the evaluation – specific creativity focus group, $t(14) = 5.00$, $p < .01$. However, this pattern seems to be reversed for the technical focus groups; here, the nonevaluation control (nonevaluation – technical focus) is rated lowest of the three, although it is significantly lower than only the specific technical focus group, $t(14) = -2.21$, $p < .05$.

It appears that at first glance, the pattern of technical goodness results mimics the pat-

[7] All significance tests reported here are two-tailed.

[8] The results from the single creativity measure mirror the results reported for the composite measure in all respects except that the technical focus evaluation groups (evaluation – technical focus and evaluation – specific technical focus) were not different from their control (nonevaluation – technical focus) at acceptable levels of significance.

rn that had been obtained for creativity: ith the exception of the evaluation group hich was told exactly what to do (the spe- fic instructions group in each case), non- aluation groups are rated higher than aluation groups. There is, however, one in- resting deviation from this parallelism. The eativity focus group that was told to expect aluation, but was *not* told specifically what do to receive a good evaluation (evalua- on – creativity focus) was rated very low n creativity (significantly lower than its ntrol, nonevaluation – creativity focus). By ntrast, however, the technical focus group hat was told to expect evaluation, but was ot told specifically what to do to receive a ood evaluation (evaluation – technical fo- us), was *not* rated low on technical good- ess; in fact, it was rated higher (although ot significantly so) than its control group.

Self-Rated Interest

Several items on the Art Activity Question- naire administered to subjects just prior to debriefing were intended to measure their attitude toward the art activity. A composite intrinsic interest measure was formed using six of these items; all six loaded higher than .50 on the intrinsic interest factor obtained in a factor analysis of questionnaire items, and they all correlated significantly with one another. These six items were (a) "Did you view your engagement in the art activity as motivated more by intrinsic factors, like your own interest, or by extrinsic factors, like the experimenter's instructions?" (b) "Was the art activity more like work or more like lei- sure activity?" (c) "How playful did you feel during the activity session?" (d) "How satisfied were you with your performance on the art activity?" (e) "How much do you like your finished design?" and (f) "How much pressure did you feel during the activity session?" This last item loaded high and nega- tive on the intrinsic interest factor, so it was subtracted from the other five in forming the composite.

Means for this composite measure are pre- sented in Table 4. It was expected that over- all, the control groups (nonevaluation) would be higher in self-rated interest than the ex- perimental groups (evaluation). Recall that on the creativity measure, the specific crea- tivity instructions group (evaluation – spe- cific creativity focus) was expected to be an exception to the general pattern. This excep- tion was *not* predicted on the intrinsic interest measure, however. On the contrary, it was ex- pected that even though the specific creativ- ity instructions subjects might exhibit supe- rior creativity in accord with their task in- structions, their intrinsic interest would still be undermined by evaluation expectation.

This overall pattern of results was, in fact, obtained. An analysis of variance on all eight groups yielded a significant overall effect, $F(7, 87) = 2.68$, $p < .025$, and a planned con- trast testing the specific trend of nonevalua- tion groups being higher on intrinsic interest than evaluation groups was statistically sig- nificant, $F(1, 87) = 4.08$, $p < .05$.

In comparison with the creativity results, however, the intrinsic interest results are not as strong. Indeed, only two experimental– control paired comparisons are statistically significant: that for the two no focus groups (nonevaluation – no focus vs. evaluation – no

Table 3
Mean Judge Ratings of Technical Goodness

Evaluation expectation	Instruction				
	No focus	Technical focus	Specific technical focus	Creativity focus	Specific creativity focus
Absent	.254	−.058	.231	.258	−.269
Present	−.081	−.014		−.322	

Note. These numbers are the means of composites of four normalized components of technical goodness that clustered on the factor analysis of artist judgments.

Table 4
Mean Self-Ratings of Intrinsic Interest

Evaluation expectation	Instruction: No focus	Technical focus	Specific technical focus	Creativity focus	Specific creativity focus
Absent	.311	.294		−.059	
Present	−.184	.222	−.568	.207	−.158

Note. These numbers are the means of composites of six normalized measures of intrinsic interest that clustered on the factor analysis of questionnaire items.

focus), $t(22) = 2.07$, $p < .05$, and that for the specific technical focus group and its control (nonevaluation – technical focus vs. evaluation – specific technical focus), $t(22) = 2.77$, $p < .02$.

Despite the failure of specific comparisons, however, the overall planned contrast suggests that it is reasonable to assert that intrinsic interest was undermined by evaluation expectation in this study. This result is particularly important when the two specific instructions groups are considered: Although the technical instructions group (evaluation – specific technical focus) was very high on rated technical goodness, and the creativity instructions group (evaluation – specific creativity focus) was very high on creativity, both of these groups were quite low on intrinsic interest. In other words, as predicted, the specific creativity instructions group (evaluation – specific creativity focus) did *not* exhibit a high level of intrinsic interest to match its high level of creativity.

Discussion

The main hypothesis of the present study—that under nonspecific instructional sets, the expectation of external evaluation will lead to decrements in creativity—was strongly supported by the artist-judges' ratings of creativity. Whether subjects were given no particular focus, a technical focus, or a creativity focus in their instructions, subjects in the evaluation conditions were, overall, rated lower on creativity than subjects in the nonevaluation conditions. In keeping with the theoretical analysis presented earlier, only the specific creativity instructions evaluation group (evaluation – specific creativity focus) was higher than its nonevaluation control on rated creativity. Thus, the present results suggest a reconciliation between the seemingly contradictory findings on the effect of constraint on creativity provided by the overjustification and behavior modification literatures.

Reinterpretation of Previous Findings

The behavior modification studies appeared to demonstrate that creativity can be increased by the offer of rewards for creative performance. This conclusion clearly contradicts the intrinsic motivation view of creativity proposed here: that the imposition of extrinsic constraints (such as rewards or external evaluation) can lead to decrements in creativity. The key to the reconciliation lies in the nature of the task and the nature of the instructions given. In accord with the present thesis, if a creative performance depends on some degree of risk taking and set breaking—some level of production beyond the obvious and commonplace—the imposition of salient extrinsic constraints, establishing an extrinsic motivation, will result in lower levels of creativity. In order to evoke a spontaneously creative performance, a task must have some degree of ambiguity—a nonobvious solution or method of approach. In McGraw's (1978) terms, it must require a *heuristic* rather than an *algorithmic* solution. This was clearly *not* the case in the behavior modification studies cited earlier, and it was not the case in the specific creativity instructions condition in the present study.

The behavior modification studies, for the most part, used verbal tests of creativity. The instructions given to children under reward conditions in those studies effectively eliminated any ambiguity about what constituted a good ("creative") performance; if the experimenters wanted to demonstrate that "fluency" was under experimental control, they told the children that they would be rewarded for producing a large number of ideas. Not at all remarkably, children produced large numbers of ideas under these conditions. And, apparently, the instructions given to the specific creativity instructions group in the present study also succeeded in reducing the ambiguity in the task; subjects were told to come up with a novel idea, to use the materials in a novel way, to make a detailed and complex design, and so on. Since, as was evident from pretest results, judges considered complex, detailed designs with novel ideas and novel use of materials to be creative, this group achieved very high creativity scores. What is crucial here is that, within the same experimental design, when subjects were given evaluation instructions but *not* told specifically what to do, their creativity was dramatically lower. And, when subjects did *not* expect evaluation, their creativity remained high—no matter what they were asked to focus on.

The present thesis proposes motivational state as a mediator of the observed performance effects. The results obtained on subjects' intrinsic motivation, while generally supportive of this proposition, are not as strong as might have been desired. An attempt at obtaining a behavioral measure of intrinsic interest failed because of the generally low level of any behavior emitted by subjects during the free-time period. The self-rating questionnaire on intrinsic interest did show significant interest differences consistent with the hypothesis. However, because some specific paired comparisons between experimental and control groups did not yield the predicted significant differences, further research is required before the observed creativity decrement can reliably be attributed to decrements in intrinsic motivation.

Practical Implications

In determining the practical implications of the present results, it is perhaps most instructive to consider the two no focus groups (nonevaluation – no focus and evaluation – no focus). These groups permit a test of the basic, uncluttered version of the hypothesis: Evaluation group subjects will be less creative than nonevaluation group subjects and will show less intrinsic interest. These two groups provide an analogue to conditions that exist in many social settings; in classrooms, for example, children are often given art projects and writing projects that will be graded, but the criteria for evaluation are left unspecified. Appropriately, these two groups provide the clearest, most consistent pattern of results obtained in this study: The nonevaluation group was significantly higher on rated creativity and technical goodness and significantly higher on self-rated intrinsic interest. In fact, with the exception of the specific creativity instructions group, the no focus nonevaluation group was higher than any other on rated creativity; moreover, it was the highest of all eight groups on self-rated intrinsic interest and second highest on rated technical goodness. In all respects, then, the nonevaluation group here performed better than the evaluation group and felt, apparently, more intrinsically motivated.

The specific creativity instructions group (evaluation – specific creativity focus) is of particular interest when considering practical applications of this work, since this group was higher in rated creativity than any other. Indeed, this result might seem to suggest that there is no reason *not* to use an extrinsic constraint, if it is accompanied by specific performance instructions. It can be argued, however, that there are several flaws in this approach. First, for most activities, it is difficult or impossible to specify the behaviors that would lead to a creative performance. Second, the high creativity of subjects in the specific creativity instructions group was *not* accompanied by a high level of technical competence; in fact, the rated technical competence of this group was significantly lower than control. Finally, and perhaps most

importantly, the high creativity of this group was not accompanied by a high level of intrinsic interest.

It is important to note that the present analysis and the present study deal with relatively low, "nonprofessional" levels of creativity. While we might expect that some professional scientists or artists could succumb to the overjustification effect, it seems eminently clear that many highly creative people go on being creative in the face of numerous extrinsic constraints. (See Simonton, 1977, for a historical look at this question.) There are two possible explanations for the seeming impotency of the overjustification effect at this level. First, it is probably the case that these exceptional individuals have largely internalized the norms and standards by which their work would be judged. Thus, external evaluation becomes less salient and less necessary. Second, highly creative and successful individuals almost certainly have a very high level of intrinsic interest in their work, and they are probably well aware of this interest. The overjustification effect is proposed to occur only when internal states are ambiguous or nonsalient (see Bem, 1972).

The practical implications of these results are perhaps most relevant for educational settings. They suggest that if creativity is to be fostered and interest maintained, care must be taken in the use of control in the classroom. Surely, some children would have no interest in even attempting a task—say, an art activity—unless some control were exerted over their behavior. But the present results suggest that if control is *over*sufficient, undesirable consequences could result. Consider what the state of modern physics might be if Einstein had not been clever enough early in his career to circumvent the detrimental effects of external evaluation.

References

Amabile, T., DeJong, W., & Lepper, M. Effects of externally imposed deadlines on subsequent intrinsic motivation. *Journal of Personality and Social Psychology*, 1976, *34*, 92–98.

Bem, D. Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review*, 1967, *74*, 183–200.

Bem, D. Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6). New York: Academic Press, 1972.

Condry, J. The enemies of exploration. *Journal of Personality and Social Psychology*, 1977, *35*, 459–466.

deCharms, R. *Personal causation*. New York: Academic Press, 1968.

Deci, E. Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, 1971, *18*, 105–115.

Glover, J., & Gary, A. Procedures to increase some aspects of creativity. *Journal of Applied Behavior Analysis*, 1976, *9*, 79–84.

Greene, D., & Lepper, M. Effects of extrinsic rewards on children's subsequent intrinsic interest. *Child Development*, 1974, *45*, 1141–1145.

Halpin, G., & Halpin, G. The effect of motivation on creative thinking abilities. *Journal of Creative Behavior*, 1973, *7*, 51–53.

Johnson, R. Differential effects of reward versus no-reward instructions on the creative thinking of two economic levels of elementary school children. *Journal of Educational Psychology*, 1974, *66*, 530–533.

Kelley, H. Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 15). Lincoln: University of Nebraska Press, 1967.

Kelley, H. The processes of causal attribution. *American Psychologist*, 1973, *28*, 107–128.

Kruglanski, A. The endogenous—exogenous partition in attribution theory. *Psychological Review*, 1975, *82*, 387–406.

Kruglanski, A., Friedman, E., & Zeevi, G. The effects of extrinsic incentives on some qualitative aspects of task performance. *Journal of Personality*, 1971, *39*, 606–617.

Lepper, M., & Greene, D. Turning play into work: Effects of adult surveillance and extrinsic rewards on children's intrinsic motivation. *Journal of Personality and Social Psychology*, 1975, *31*, 479–486.

Lepper, M., Greene, D., & Nisbett, R. Undermining children's intrinsic interest with extrinsic rewards: A test of the overjustification hypothesis. *Journal of Personality and Social Psychology*, 1973, *28*, 129–137.

McGraw, K. The detrimental effects of reward on performance: A literature review and a prediction model. In M. Lepper & D. Greene (Eds.), *The hidden costs of reward: New perspectives on the psychology of human motivation*. Hillsdale, N.J.: Erlbaum, 1978.

McGraw, K., & McCullers, J. Evidence of a detrimental effect of extrinsic incentives on breaking a mental set. *Journal of Experimental Social Psychology*, in press.

Nunnally, J. *Psychometric theory*. New York: McGraw-Hill, 1967.

Raina, M. A study into the effect of competition on creativity. *Gifted Child Quarterly*, 1968, *12*, 217–220.

Schilpp, P. *Albert Einstein: Philosopher-scientist*. Evanston, Ill.: Library of Living Philosophers, 1949.

Simonton, D. Creative productivity, age, and stress: A biographical time-series analysis of 10 classical composers. *Journal of Personality and Social Psychology*, 1977, *35*, 791–804.

White, K., & Owen, D. Locus of evaluation for classroom work and the development of creative potential. *Psychology in the Schools*, 1970, *7*, 292–295.

Received February 20, 1978 ■

Corrections to Lynch and Cohen

In the article, "The Use of Subjective Expected Utility Theory as an Aid to Understanding Variables That Influence Helping Behavior," by John G. Lynch, Jr., and Jerry L. Cohen (*Journal of Personality and Social Psychology*, Vol. 36, No. 10, pp. 1138–1151), the following corrections should be made:

1. The last two lines of Table 1 (p. 1140) should read

$SEU_{not\ help} = \sum_{i=3}^{4} P_i U_i$	−5.4	−2.5	−.4
$SEU_{help} - SEU_{not\ help}$	6.4	2.5	−.6

2. The symbols used in the curves in the left and right panels of Figure 1 (p. 1143) are not related to the manipulated levels of probability in the manner specified in the legends of these panels. Therefore, the symbols in the legends should be ignored. In both legends, probability levels are listed in ascending order of subjective magnitude, and the higher the level of probability, the steeper the slope of the curve to which it properly corresponds.

3. The third sentence in the first full paragraph on page 1146 should read, "The Probability 1 × Utility 1 × Probability 2, Probability 1 × Utility 1 × Utility 2, Probability 1 × Probability 2 × Utility 2, and Utility 1 × Probability 2 × Utility 2 interactions were all significant, with $F(8, 26) = 5.22$, 9.52, 3.89, and 4.61, respectively ($p < .001$)."

Journal of Personality and Social Psychology
1979, Vol. 37, No. 2, 234-246

Menstrual Cycle Affects Kinesthetic Aftereffect, An Index of Personality and Perceptual Style

A. Harvey Baker
Queens College of the
City University of New York

Brian L. Mishara
University of Massachusetts—Boston

Irene W. Kostin
Educational Testing Service
Princeton, New Jersey

Laurence Parker
Cornell University

Research suggests that kinesthetic aftereffect (KAE) scores reflect status on a postulated stimulus intensity modulation (SIM) mechanism that damps down subjective stimulus intensity for some (reducing) and increases it for others (augmenting). Such a mechanism would help account for empirically observed individual differences in such behaviors as pain tolerance, sensory deprivation reactivity, and stimulation seeking. It was hypothesized and confirmed in three adult female samples that KAE varies curvilinearly over the menstrual cycle: Greater KAE reduction occurs at the cycle's beginning and end. Neither tiredness, oral contraception, medication, attention, nor social expectations can explain this finding. Of the behaviors studied in the KAE literature, only five are also encompassed by the menstrual cycle literature. Four of these (antisocial behavior, acute schizophrenic episodes, accidents, and activity level) show similar curvilinearity over the cycle. We hypothesize that cyclical variation in the SIM mechanism mediates the curvilinear pattern observed for both these four behaviors and KAE.

This study explores the relationship between menstrual cycle and kinesthetic aftereffect (KAE). Prior findings support the hypothesis that KAE scores reflect individual differences in a postulated mechanism that modulates the subjective intensity of incoming stimulation. If KAE scores vary over the menstrual cycle, then the stimulus intensity modulation (SIM) mechanism may similarly vary over the cycle, providing a useful framework for understanding certain reported behavioral variations across the cycle.

A brief report of this research was presented at the meeting of the American Psychological Association, New Orleans, September 1974. We thank Kathleen McGraw, Patricia Renner, and Valerie Finkelstein for serving as experimenters in these studies and Roy Freedle, Donald Goodenough, Leland Van den Daele, Laraine Schwartz, and Ruth Sharabany for their critical comments of an earlier draft of this aritcle.

Requests for reprints should be sent to A. Harvey Baker, Department of Psychology, Queens College, City University of New York, Flushing, New York 11367.

Kinesthetic Aftereffect as an Index of Stimulus Intensity Modulation

Several researchers (e.g., Baker, Mishara, Kostin, & Parker, 1976; Mishara & Baker, 1974; Petrie, 1967; Silverman, 1967) have hypothesized that the central nervous system functions as if it had a modulator mechanism, which differs from person to person. This mechanism acts like a volume control, modulating the subjective intensity of incoming stimulation. For people with the volume control "set low" ("reducers"), the subjective intensity of incoming stimulation is presumably attenuated. Under normal circumstances, reducers are stimulus deprived and seek more stimulation to compensate. Although they can handle high-intensity stimulation well, they are very uncomfortable when environmental stimulation is low, as in sensory deprivation

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3702-0234$00.75

situations. In contrast, people with the volume control "set high" ("augmenters") are subjectively overloaded with stimulation, show stimulation-avoidant behaviors, and cope well with situations involving minimal intensity.

Procedures used to measure this postulated mechanism include average evoked response to stimuli of varying intensities (e.g., Lukas & Siegel, 1977; Silverman, Buchsbaum, & Henkin, 1969); indexes of Russian "strength of the nervous system" (Sales, Guydosh, & Iacona, 1974; Sales & Throop, 1972); electrodermal response (Depue & Fowles, 1976); the R–A Scale, a self-report questionnaire regarding preferred degree of stimulus intensity (Vando, 1974); a self-report index assessing a wide range of behaviors and attitudes (Baker et al., 1976); and KAE (e.g., Buchsbaum & Silverman, 1968; Deaux, 1976; Petrie, 1967; Schooler & Silverman, 1969).

KAE has been the most widely used and has been related to three other SIM indexes (see text below). KAE involves changes from pretest to test trials in width judgments of a standard block that occur with intervening rubbing of an aftereffect-inducing block different in width from the standard block. Two major KAE variants exist. In the one-hand variant (Petrie, 1967), the subject rubs the inducing block with one hand while resting the other. There is a two-hand variant (e.g., Buchsbaum & Silverman, 1968); however, it has been less frequently used in KAE/SIM validity studies than the one-hand version, and it may not tap the same psychological process (Schooler & Silverman, 1969). We therefore treat only the one-hand variant here.

When, as in this study, the aftereffect-inducing block for one-hand KAE is larger than the standard block, judged width of the test block decreases on the average following aftereffect induction (e.g., Bakan & Thompson, 1965). Reducers' width judgments decrease the most following aftereffect induction, and augmenters' decrease the least (or even increase) (Mishara, Baker, Parker, & Kostin, 1973).

Consistent with the SIM hypothesis, reducing scores on KAE are associated with (a) greater pain tolerance (Petrie, Collins, & Solomon, 1958; Poser, 1960; Ryan & Foster, 1967; Sweeney, 1966); (b) lesser sensory-deprivation endurance (Petrie et al., 1958, Study 2); (c) higher "need for stimulation" (Mishara & Baker, 1974; Ryan & Foster, 1967; Sales, 1971, 1972); (d) delinquency (Compton, 1967; Petrie, McCulloch, & Kazdin, 1962); (e) more accidents (Petrie, 1967, p. 25); (f) schizophrenia (Petrie, Holland, & Wolk, 1963); (g) greater activity level (Petrie, 1967, p. 100; Sales, 1971); and (h) use of amphetamines rather than barbiturates (Deaux, 1976). Regarding other indexes of SIM status, KAE reducing scores are associated with (i) reducing, as assessed by average evoked response (Schooler, Buchsbaum, & Carpenter, 1976); (j) higher sensory thresholds (as well as with two other indexes of a "strong" nervous system) (Sales & Throop, 1972); and (k) higher scores on a self-report index (Baker et al., 1976).

Despite these validity findings, critics have recently contended that KAE lacked retest reliability (e.g., McDonald, 1974; Morgan & Hilgard, 1972; Platt, Holtzman, & Larson, 1971), showed intermittent validity (e.g., Sales & Throop, 1972; Weintraub, Green, & Herzog, 1973), and thus should not be used as a personality index. However, we (Baker et al., 1976; Baker, Mishara, Parker, & Kostin, 1978; Mishara & Baker, 1974) and others (Bakan & Thompson, 1962) have demonstrated that systematic individual differences in practice effects bias second-session KAE scores. Low retest reliability, which is caused by such practice effects, can be disregarded. When KAE–personality findings are restricted to one-session studies, there is clear evidence of validity (Baker et al., 1976). Internal consistency, the appropriate reliability statistic for a one-session task, is high in 10 samples (Mishara & Baker, 1978). We conclude that the KAE–SIM formulation shows good construct validity when a one-session design is used.

Kinesthetic Aftereffect and the Menstrual Cycle

SIM is presumed to reflect not only stable individual differences (trait) but also tran-

sient organismic fluctuations (state) (Petrie, 1960; Petrie et al., 1963). If KAE is a sensitive index of current modulator status, it should reflect both between-people differences in trait, as the validity findings reviewed above indicate, and transient changes in state. The latter possibility is supported by Gupta's (1974) finding that state-altering variables such as Dexedrine and phenobarbital significantly affect KAE performance.

The present KAE–menstrual cycle hypothesis is based on two interrelated assumptions: (a) Strong stimulation (whether external or internal) is presumed to lead to a shift toward reducing on the part of the SIM mechanism.[1] This shift results in a damping down of the subjective intensity of incoming stimulation (cf. Petrie, 1967, pp. 51–60). The KAE scores should reflect this shift toward reducing. (b) One important form of internal stimulation, pain (e.g., lower abdominal cramps, tender breasts) is presumed to vary curvilinearly over the menstrual cycle, with maximal pain occurring at the cycle's beginning and end (Dalton, 1964; Moos et al., 1969).[2] If these two assumptions are correct, then (a) there should be a shift toward SIM reduction at the menstrual cycle's beginning and end, when pain is maximal, and (b) KAE scores should index this shift in SIM. Where, as here, the aftereffect-inducing block is larger than the test block, we hypothesize greater KAE reduction (larger aftereffect) at the beginning and end of the menstrual cycle—in other words, we hypothesize a curvilinear relationship between KAE and locus in the menstrual cycle.[3]

Relationship of This Study to the Menstrual Cycle Literature

Very few of the reported menstrual cycle findings have withstood recent critical reviews (O'Connor, Shelly, & Stern, 1974; Parlee, 1973, 1974; Redgrove, 1971; Sommer, 1973; Zimmerman & Parlee, 1973). Improper generalizations from sample to population, problems in drawing causal inferences from correlational data, inadequate statistical treatment, and failures in replicating earlier findings are some central problems raised by the critics.

The reviews generally agree that a few consistent, replicable findings of periodicity have been reported regarding (a) self-reported symptoms such as mood changes (Coppen & Kessel, 1963; Moos et al., 1969), (b) incidence of criminal behavior (Dalton, 1961; Morton, Additon, Addison, Hunt, & Sullivan, 1953), and (c) incidence of psychiatric crises (Dalton, 1959; Glass, Heninger, Lansky, & Talan, 1971; Jacobs & Charles, 1970). As Parlee (1974) and Sommer (1973) note, few replicated findings exist involving objectively measured nonverbal behaviors such as performance on cognitive tasks, reaction time, and athletic performance. Parlee (1974) and Sommer (1973) conclude that (a) most of the replicable findings are in behavioral areas that can be influenced by social–cultural expectations concerning how women "should" behave or feel at different phases of the cycle and that (b) few, if any, consistent and significant findings exist in areas where no clear cultural expectations exist. They hold, in effect, that most psychological and behavioral changes reported do not reflect phenomena intrinsically associated with the cycle (such as hormonal or other physiological changes) but rather are "artifactual," arising from social–cultural expectations.

Recently, a few carefully done studies have focused on objective, nonverbal behaviors—

---

[1] This assumption is supported by findings that (a) high-intensity auditory stimulation by itself increases pain tolerance (Gardner & Licklider, 1959), and high-intensity auditory stimulation combined with suggestion also increases pain tolerance (Melzack, Weisz, & Sprague, 1963) and decreases self-reported pain (Lavine, Buchsbaum, & Poncy, 1976); that (b) pain tolerance also increases with nonpainful somatosensory stimulation (Higgins, Tursky, & Schwartz, 1971; Satran & Goldstein, 1973); and that (c) the strong internal stimulation of organic pain raises the pain threshold (decreases pain sensitivity) in undamaged areas of the body (Hazouri & Mueller, 1950; Mersky & Evans, 1975).

[2] Our assumption here is that these changes in self-report of pain over the menstrual cycle are primarily due to hormonal and other physiological changes over the cycle, a view consistent with Dalton (1964) and Lennane and Lennane (1973), among others. We are aware that some would argue (e.g., Parlee, 1974) that such findings reflect social expectancy effects, an alternative hypothesis that will be considered briefly in the discussion section.

[3] Petrie (1967, p. 103) mentions, without presenting any supportive data, pilot findings consistent with our prediction regarding premenses.

not likely to be mediated by social expectations. For example, both conditionability to an aversive stimulus (Asso & Beech, 1975) and rod-and-frame test performance (Klaiber, Broverman, Vogel, & Kobayashi, 1974) vary over the menstrual cycle. More such studies must be done before the significance of these findings can be determined. Here, we attempt to relate another objectively measured, nonverbal behavior to the menstrual cycle.

Method

Apparatus

Following Petrie (1967), apparatus included a 30-inch (76.20 cm) long tapered comparison wedge, a 2.5-inch (6.35 cm) wide aftereffect induction block, and a test block (whose width is specified below for each sample). A mounted ruler ran the length of the comparison wedge. Around the ruler were a pair of finger guides, allowing the exact location of the subject's hand on the wedge to be recorded at the end of each trial. The subject was blindfolded and seated in front of a table that had the apparatus on it.

Subjects

Three independent groups of right-handed women ($ns = 47$, 14, 55), predominantly of college age ($Ms = 20.5$, 20.6, and 20.9 years, respectively), were tested. The samples are reported separately because of minor procedural differences: Samples 1 and 2 were tested concurrently by two female experimenters using different-sized KAE test blocks—2 inches (5.08 cm) and 1.5 inches (3.81 cm), respectively. Sample 3 was run 1 year later by a different female experimenter with the 1.5-inch (3.81 cm) wide test block. Three comparison samples of right-handed males ($ns = 31$, 10, and 59; mean ages = 19.6, 21.1, and 21.2 years, respectively) were run concurrently with the three female samples.

Procedure

Because of the differential bias in repeated KAE administrations discussed above, the more ideal procedure of using multiple KAE administrations over the menstrual cycle's course was contraindicated. An 18-trial KAE procedure was thus given *once* to each subject, following Petrie's (1967) detailed procedures exactly. The most important features of Petrie's standardized procedure include the following: (a) On each of the 18 judgments, subjects held the test block with the thumb and forefinger of the right hand and indicated its width on the tapered block with the thumb and forefinger of the left hand. (b) There were two practice trials, four pretest trials, and then three aftereffect-induction periods (90, 90, and 120 sec long, respectively), each followed by four test judgments (12 test trials in all). (c) Before KAE, subjects rested their hands for 45 min so that nothing touched their thumbs and forefingers (cf. Petrie, 1967, pp. 107–127 for further details).

Scores

The KAE score was the mean of the 12 test judgments minus the mean of the four pretest judgments. Negative scores represent the usual aftereffect (width judgments on test trials smaller than on pretest trials). The most negative scores (largest aftereffect) indicate reducing; minimally negative—or even positive—scores (smallest aftereffect) represent augmenting.

Interview

As part of a large questionnaire battery orally administered by the experimenter during the 45-min rest period, the subject estimated onset date of next menses. Based on this estimate, *raw* menstrual cycle score (number of days before onset of menses) was calculated.

In Sample 3, a more refined procedure—occasionally used in recent menstrual cycle studies (e.g., Golub, 1976)—was adopted: Cycle length (date of most recent onset and date of onset of next menses) was estimated at the time of the KAE administration, and then the actual date of onset of the next menses was verified by follow-up phone calls.[4] This permitted, for Sample 3 only, conversion of the raw menstrual cycle score into an adjusted menstrual cycle score[5] based on a standard 29-day cycle.[6]

Information on possible confounding variables was also collected: (a) rating of tiredness now as com-

---

[4] The validity of subjects' estimates can be assessed (a) by correlating number of days until next onset, based on estimated versus confirmed onset dates, or (b) by computing the mean absolute difference between estimated and confirmed dates. In Sample 3 and in a larger follow-up study regarding menstrual cycle estimates, the validity correlations were high ($rs \geq .92$) and the absolute differences adequately small (1.0 and 2.1 days for Sample 3 and the follow-up sample, respectively).

[5] For example, suppose a particular subject had a cycle length of 24 days and that she was administered KAE on the 6th day before onset of next menses. The day on which she was tested was first converted into a percent of her cycle, in this case 25%, and then the percent was reconverted to days based on the standard 29-day cycle, in this case to 7.25 days. Employment of a standard cycle is a more refined procedure that has occasionally been used in the menstrual cycle literature (e.g., Moos et al., 1969).

[6] The decision to specify the standard cycle as 29 days is based on the mean cycle length found in a large study ($n = 2{,}542$) by Sheldrake and Cormack (1976) and on the median mean cycle length determined from a review of 21 other studies (Presser, 1974).

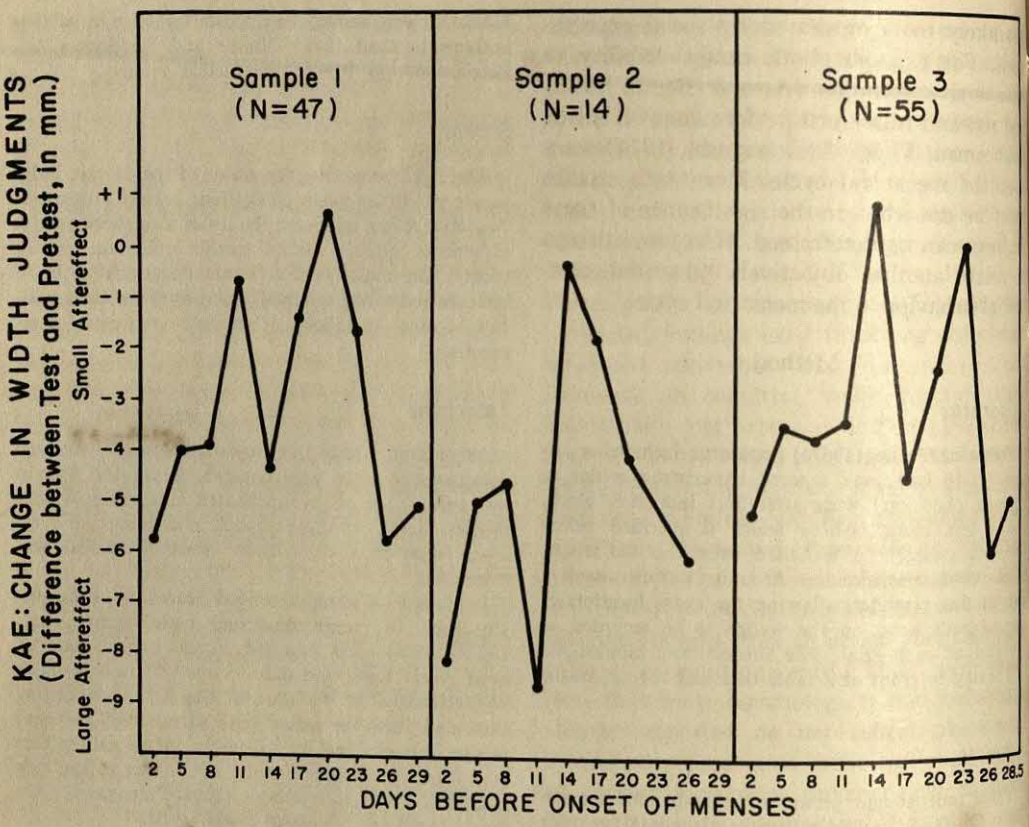


Figure 1. Menstrual cycle affects kinesthetic aftereffect (KAE).

pared to usually (on a 7-point scale); (b) oral contraceptive usage; and (c) any kind of medication or drug usage during the past 12 hours.

Since the same experimenters conducted the interviews and administered the KAE task, the experimenters were not told the purposes and hypothesis of the study.

Results

Figure 1 presents the KAE scores for each sample as a function of days before onset of next menses. To smooth the curves, reported day of the cycle was collapsed into 10 equal intervals of 3 days each. For Sample 2, no subjects fell in two of these intervals, so only 8 data points are depicted. For Sample 3, the last interval consists of 2 days only—Days 28 and 29 of a standard 29-day cycle (see Method section).

These curves each show irregularity—most notably, a dip near midcycle (an issue treated briefly in Footnote 11). Despite this irregularity, the curves generally show the predicted curvilinear pattern.

Test of the Main Hypothesis

Correlational analysis. One way to test for curvilinearity is to ascertain the degree to which the observed data can be fitted by a curve that looks like an inverted U, with (a) the most negative scores (largest aftereffect) at the beginning and end of the cycle, (b) the least negative scores (smallest aftereffect) at the middle, and (c) days equidistant from the midpoint having the same value.

The choice of method to test for curvilinearity was shaped by the size of the present samples, the goal of treating menstrual cycle as a continuous variable (many studies have broken up the cycle into "phases," the definition of which differs from study to study), and the goal of using all information available. The following procedure met these con-

straints: First, the raw menstrual cycle score (i.e., subject's estimate) was transformed into the absolute value of each day's deviation from midcycle (Day 15).[7] This procedure folds one arm of the predicted inverted U-shaped curve onto the other arm (days with identical predicted KAE, for example, 1 and 29, have the same value). Since we expected a curve, we then squared these deviations from Day 15 to produce our transformed menstrual cycle score. According to the curvilinear hypothesis, (a) the least negative KAE scores (augmentation) occur at midcycle, where the smallest values of the transformed scores occur (e.g., raw menstrual cycle score of Day 15 corresponds to a value of 0^2 on the transformed scale); and (b) the most negative KAE scores (reduction) occur at both ends of the cycle, where the largest values of the transformed scale occur (for example, raw scores of 1 and 29 both correspond to a transformed value of 14^2). Thus, a negative linear relationship should obtain between KAE and the transformed scores.

For Samples 1, 2, and 3, the Pearson product-moment correlations between KAE and the transformed menstrual cycle scores were $-.237$ ($p \approx .05$), $-.602$ ($p < .05$), and $-.245$ ($p < .05$), respectively.[8] The average correlation (McNemar, 1969) was $-.285$ ($p < .002$). Since the main hypothesis involved a directional prediction, p values are for one-tailed tests.

Two double checks on the correlational analysis. These correlations approximate one-way analysis of variance with menstrual cycle as the independent variable (with 29 levels) and with curvilinearity tested by quadratic trend analysis. Because sample size was far too small, one-way analyses of variance using trend analyses could only be employed with far fewer than 29 menstrual cycle levels. The raw menstrual cycle scores were here collapsed into 10 levels for Samples 1 and 3 (see Figure 1) and into 4 levels for Sample 2. Resulting Fs from the quadratic trend analyses (dfs = 1, 44; 1, 10; 1, 52) were 2.90, 4.59, and 3.68, respectively. Combining probabilities for independent tests on the same hypothesis (Winer, 1962), $\chi^2 = 19.40$, $p < .01$.

For the second double check, the three samples were pooled, and subjects were classified into two categories occasionally used (e.g., Dalton, 1964) in menstrual cycle studies: (a) paramenstruum (week before and days of menses)[9, 10] versus (b) midcycle (the rest of the cycle). Paramenstruum females ($M = -4.92$ mm, $SD = 4.08$ mm) showed greater KAE reducing, $t(114) = 3.18$, $p < .001$, than midcycle females ($M = -2.34$ mm, $SD = 4.35$ mm). Since this single simple analysis has the advantage of providing a large n for each cell, certain of the later analyses of secondary issues are patterned on it.

Some Secondary Issues

Male–female comparisons. Although few menstrual cycle researchers (e.g., Dalton, 1964) have used comparison groups of males, Parlee (1973) recently enunciated a rationale for employing this admittedly not perfect control procedure: She asserts that the use of nonmenstruating individuals (for example, males) can help in both description and understanding by providing a baseline with which to compare behaviors associated with different parts of the menstrual cycle.

Paramenstruum females showed greater reduction (larger aftereffect) than males in

---

[7] Day 15 is the day exactly in the middle of the standard 29-day cycle used here.

[8] When the day of menstrual cycle raw score is converted into an adjusted score for a standard 29-day cycle, which was possible only for Sample 3, the value of the correlation becomes slightly higher, $r = -.258$.

[9] For menstruating subjects, days of menses at time of KAE testing ranged from Day 1 to Day 5 of menstrual flow.

[10] Because our hypothesis was based, in part, on cyclical variation of painful symptoms, we used the same cutoff points as Moos (1968), whose extensive ($n = 839$) study found that more women complain of painful symptoms during menstruation and during *the week* before menses relative to intermenses. When we review other menstrual cycle studies, we use the term "paramenstruum" to refer to findings clearly associated with the cycle's beginning and end either as defined by Dalton (1964; that is, the 4 days before and the 4 days after menses) or as defined by us or including slight variations on this definition (e.g., 1 week before menses and 1 week after menses, as used by Glass et al., 1971).

each sample. For paramenstruum females, Ms = −4.97, −5.26, and −4.86 mm (SDs = 4.97, 1.97, and 3.60 mm; ns = 16, 5, and 19) for Samples 1–3, respectively. For males, Ms = −1.98, −1.95, and −3.45 mm (SDs = 5.75, 2.38, and 2.99 mm; ns = 31, 10, and 59). Comparisons of paramenstruum female and male KAE scores within each sample gave t values (dfs = 45, 13, and 76) of 1.76, 2.69, 1.57; $p < .10$, $p < .02$, and $p < .14$, respectively. Combining probabilities across samples (Winer, 1962), $\chi^2 = 16.36$, $p < .02$. (All p values are for two-tailed tests, since there was no a priori directional hypothesis here.)

Midcycle females in Samples 1 and 2 (Ms = −2.37 and −2.85 mm; SDs = 5.07 and 3.86 mm; ns = 31 and 9, respectively) also showed greater reduction (*larger* aftereffects) than the corresponding males (see above), but this difference was clearly not significant, ts(60) and (17) = .28 and .63, *ns*. In Sample 3, midcycle females ($M = -2.19$ mm; $SD = 3.60$ mm; $n = 36$) showed less reduction (smaller aftereffect) than the corresponding males. This difference, opposite in direction to the nonsignificant male–female differences for Samples 1 and 2, approached the conventional level of significance, $t(93) = 1.67$, $p < .10$. Taken together, these data suggest that there is no difference between males and midcycle females on KAE.[11]

Given that paramenstruum females do, but midcycle females do not, differ from males, one may ask whether females in general differ from males. For all females pooled, KAE scores are −3.24, −3.71, and −3.06 mm (SDs = 5.13, 3.43, and 3.75; ns = 47, 14, and 55) for Samples 1–3, respectively. (See preceding text for corresponding male values.) Male–female comparisons within each sample gave t values (dfs = 76, 22, and 112) of .97, 1.42, and .61, respectively, *ns*. In Samples 1 and 2, females show greater reduction (larger aftereffect); in Sample 3, they show lesser reduction (smaller aftereffect). Overall there is no evidence of a significant sex difference when all females are involved in the comparison. This finding is consonant with reports from three earlier, single-session studies employing the same one-hand KAE procedure used here (Barthol, 1958; Charles & Duncan, 1959; Spitz & Lipman, 1961).

Possible confounding variables. We analyzed three variables that might have confounded the observed menstrual cycle–KAE relationship: (a) tiredness, which reportedly fluctuates over the cycle (Dalton, 1964); (b) oral contraceptive use, which affects mood and behavior (Kutner, Phillips, & Haag, 1971); and (c) use of drugs or medication, which might vary systematically over the cycle, and which can affect KAE (Gupta, 1974). Each variable was assessed by self-report. Two separate 2 × 2 unweighted-means, independent group analyses of variance were undertaken on the KAE scores with (a) menstrual status (paramenstruum vs. midcycle) as one variable and (b) degree of tiredness (more tired than usual vs. the same or less tired than usual) and oral contraceptive status (use vs. not use), respectively, as the other variable. No significant interaction was observed between degrees of tiredness and menstrual status, $F(1, 112) = .06$, *ns*, or between oral contraceptive status

[11] Perhaps a true male/midcycle female difference exists that is obscured by the presence of the dip that each sample's curve displays somewhere in the middle of the cycle (see Figure 1). Since our hypothesis is based in part on findings of variation in painful symptoms over the cycle, it is relevant to note here that in addition to the more common report of painful symptoms at the beginning and end of the cycle, some women also report experiencing pain in the middle of the cycle, about the time of ovulation. Indeed, Vollman (1974) regards such midcycle pain as one test of time of ovulation, noting that for some women this pain is so severe as to be incapacitating. Moos et al. (1969) plotted reports of painful symptoms and found maximal values at either end of the cycle and also an increase in self-reported pain in the middle of the cycle. In this context, the increase in reduction on KAE in the middle of the curves for our three samples is consistent with the assumption that an increase in internal pain leads to damping down (reduction) on the part of the postulated modulator mechanism. Exploratory t tests were undertaken in Sample 3, where the most precise menstrual cycle measurement was made. These indicate that the mean of Days 16–18 (depicted as Day 17 on the graph) differed from both the mean of Days 22–24, $t(11) = 2.98$, $p < .02$, two-tailed test, and from the mean of Days 13–15, $t(91) = 2.41$, $p < .05$. This exploratory "finding" of a midcycle dip warrants further investigation.

and menstrual status, $F(1, 98) = .07$, *ns*. (The latter *n* is smaller here because the experimenter inadvertently forgot to inquire about oral contraceptive status for some subjects.)

Finally, we assessed whether the KAE–menstrual cycle finding might artifactually reflect differential use of drugs and medication over the cycle. If, for example, drugs and medications that lead to greater reducing (Gupta, 1974) were used more at paramenstruum and drugs that lead to greater augmentation were used more during midcycle, such differential drug use might have produced our finding. To see if the findings would hold, independent of drug usage, we simply dropped all subjects who reported using drugs and/or medication in the 12 hours before testing and computed a *t* test between the remaining midcycle and paramenstruum females. There was still a menstrual cycle–KAE effect, $t(89) = 2.66$, $p < .005$.

Discussion

Observed Kinesthetic Aftereffect/Menstrual Cycle Relationship

Effect magnitude. In three samples, KAE scores show the same curvilinear pattern over the menstrual cycle. How large is this effect? Effect magnitude can be assessed from two perspectives—means (effect size) or variability (explained variance): (a) Using Cohen's (1969) *d* statistic, paramenstruum and midcycle females differ in KAE means by .52, .79, and .74 *SD* for Samples 1–3, respectively. Cohen described effects of this size as "medium" (Sample 1) and "large" (Samples 2 and 3). However, (b) the average correlation of .285 shows that menstrual cycle explains 8% of the variance in KAE scores, a significant but modest amount. We suspect that improved methodology and specification of the midcycle dip would strengthen the correlational results (see Footnote 11).

Possible confounding variables. Elimination of possible confounding variables strengthens the findings. Empirical analyses show that tiredness, oral contraceptive use, and the effects of drugs and/or medication do not account for the results. On logical grounds, one can eliminate a possible attentional hypothesis: If cyclical variation in pain and discomfort (Moos et al., 1969) directly affects attention, then attention to the KAE task would presumably be lower at menses and premenses. Since KAE is minimal with lessened attention (Bakan & Thompson, 1965), one might predict *greater* KAE at midcycle, whereas smaller aftereffects were found here.

Social expectancy effects. It appears that the KAE findings cannot be explained in either of the two primary ways in which social expectancy effects have been invoked to account for certain prior menstrual cycle findings (Parlee, 1974; Sommer, 1973): (a) When interpreting a correlational relationship between the menstrual cycle and some other variable, the *researcher* may, because of his or her social expectations, ascribe causal status to the menstrual cycle when it is an effect. For example, Dalton (1964) interpreted the greater frequency of accidents at paramenstruum as being caused by psychological processes associated with the menstrual cycle and ignored the alternative hypothesis (Parlee, 1973) that stress associated with the accident accelerated menses' onset, thus leading to this "finding." We can see no way in which taking KAE might have any such effect on menstrual cycle locus. (b) The *subject*, because of her social expectations, may behave differently at or near menses—for example, doing more poorly on a task because she believes women do poorly then. KAE, however, is not a "right or wrong" task that focuses on simple performance errors. Nor do we know of any social expectancies about when augmenting rather than reducing errors occur on KAE. So, neither type of social expectancy effect is likely to occur here.

The method of this study. Precise physiological assessment (assay of blood samples, assessment of basal temperature, and indexes of major metabolites in the urine) of the menstrual cycle variable is important for accurate measurement, since pronounced individual differences exist in occurrence of ovulation and in patterning of hormonal secretions. But these procedures are quite expensive and

difficult. An admittedly simpler and more economical approach seems appropriate when first exploring a given domain for possible menstrual cycle relationships. In a situation in which social expectancies have an effect, use of self-report measures might well increase the role of such expectations, and thus Type I errors would be more likely. With such biasing effects ruled out, the decrease in accuracy in menstrual cycle assessment due to self-reports should simply make verification of a directional hypothesis such as ours more difficult. Studies using more precise menstrual cycle measurement procedures should simply find a stronger effect.

Implications for KAE methodology. Is it feasible to continue to use KAE as a measure of trait since we and Gupta (1974) have shown that state variables affect KAE? Yes. Only 2 (Deaux, 1976; Schooler et al., 1976) of the 18 KAE validity studies cited in the introduction controlled for state. Thus, state need not be controlled to obtain significant personality findings with KAE. Of course, a useful task for improving KAE's predictive power would be the development of some means of separating state and trait variance.

Possible Behavioral Implications

In contrast to most menstrual cycle studies, we observed a relationship that (a) holds across independent samples, (b) involves objective measurement on a laboratory-type task, and (c) cannot be explained by social expectancy or (d) by other obvious sources of confounding. What are the possible behavioral implications of this finding?

According to KAE validity studies, between-people differences in KAE scores (trait) correlate with between-people differences in diverse psychological domains. Based on our present findings for females, is what appears to be a within-person cyclical variation in KAE scores (state) similarly associated with a within-person cyclical variation in some or all of these same psychological domains?

Our literature review shows that five of these domains have been studied in relation to the menstrual cycle. For these five, the KAE validity literature indicates that KAE reducing scores are associated with delinquent behavior, accidents, schizophrenia, greater activity level, and higher sensory threshold. The present findings associate KAE reducing scores with the paramenstruum. Is it possible that relative to the rest of the cycle, women at paramenstruum show more delinquent behavior, more accidents, higher incidence of schizophrenic symptomatology, higher activity levels, and higher sensory thresholds?

The findings for the first three of these five domains show just this pattern. Relative to the rest of the cycle, there is significantly higher incidence during the paramenstruum of (a) crime (Dalton, 1961), unpremeditated acts of violence (Morton et al., 1953), and disorderly behavior among prisoners (Dalton, 1961); (b) accidents (Dalton, 1964); and (c) admission of newly hospitalized schizophrenics (Dalton, 1959) and hospital emergency ward treatment of schizophrenics (Glass et al., 1971).

Regarding the fourth domain, activity level, Morris and Udry (1969, 1970) report in two samples evidence that greater activity level (assessed by pedometer) occurs at the beginning and end of the cycle. Three peaks of activity were found—the largest around the time of ovluation, and two others, one occurring early in menstruation and the other in the premenstrual week. Although one might *describe* these curves as U-shaped (with an additional peak in the middle, much like our curves), no statistical test for curvilinearity was reported. At best, one can say that the observed data are consistent with the KAE curve over the menstrual cycle.

The results from the fifth domain—laboratory studies of sensory thresholds—are mixed. Only one study (of olfactory sensitivity) supports our expectation that higher sensory thresholds should occur at the beginning and end of the menstrual cycle (Vierling & Rock, 1967). In contrast, a study of visual threshold (Diamond, Diamond, & Mast, 1972), although reporting high threshold during menstruation, observed low threshold during the premenstrual week. A different pattern was reported for taste threshold (Glanville & Kaplan, 1965), where the lowest thresholds were

observed during menstruation as compared to two other periods during the cycle.

What conclusions can be drawn from these findings? With respect to sensory thresholds, the results are confusing. This may well mean that the parallel that we were seeking to draw between KAE and the menstrual cycle and between thresholds and menstrual cycle does *not* obtain. Another possibility is that these findings may reflect a problem in the threshold methodology employed (i.e., a confounding of sensitivity with response bias). Wong and Tong (1974) recently applied signal detection theory to the menstrual cycle/two-flash threshold relationship. They showed that a failure to distinguish between changes in response criterion and changes in sensitivity (a failure true of each of the threshold studies reviewed above) can considerably distort the relationship between threshold sensitivity and menstrual cycle. It remains an issue for future research to determine which of these two possibilities is most likely.

The first four domains reviewed each involved naturalistic observation, and only the last domain involved laboratory investigation. Ordinarily, one would give greater weight to studies based on laboratory research, because of the greater precision and greater control of extraneous variables normally associated with this methodology. Here, only sensory threshold studies used laboratory methods, but certain methodological limitations may be involved. Under these circumstances, there is little basis for weighting the laboratory evidence heavily and discounting the evidence from the domains involving naturalistic observation.

For these four domains, behaviors related to reducing in the KAE–trait literature show greater association with the cycle's beginning and end. For each domain, other possible explanations for this curvilinear pattern exist (see next section). Yet, within the limitations of the existing evidence, there is a convergence both with our curvilinear hypothesis and with the present curvilinear KAE findings. It is also noteworthy that recent menstrual cycle literature reviews (see the introductory sections of this article) concur that consistent and replicable menstrual cycle findings have been reported for criminal behavior and for psychiatric crises (which does subsume schizophrenic crises as one important component).[12] Thus, for two of these four domains, there is the most agreement that behavioral variation does occur over the cycle. It would be worthwhile to explore whether other behaviors known empirically to correlate with KAE show a similar curvilinear pattern over the menstrual cycle.

Possible Theoretical Implication

Our hypothesis regarding KAE was based, in part, on the assumption that the postulated SIM mechanism varies curvilinearly over the cycle. The finding that KAE shows the predicted curvilinear pattern thus constitutes one basis for inferring that the underlying process of SIM (presumed to underlie individual differences in KAE and thus to mediate the relationship between KAE and various external behaviors) also varies curvilinearly over the menstrual cycle. Obviously, this type of inference requires additional independent confirmation, preferably from other procedures that assess the postulated process (for example, the average evoked response measure of reducing–augmenting).

Suppose that future research clearly establishes that amount of activity level and incidence of criminal behavior, schizophrenic symptomatology, and accidents vary curvilinearly over the cycle. We would then tentatively offer the single parsimonious explanation that the modulator mechanism varies curvilinearly over the menstrual cycle and that this variation mediates the cyclic pattern observed for these four behavioral domains. There are various other alternative hypotheses that might also explain the cyclical pattern observed for any one of these domains. Dalton (1964), for example, explained sev-

---

[12] Although there is general agreement that incidence of "psychiatric crises" shows periodicity over the menstrual cycle, it is important to add that three studies have additionally indicated that the particular phenomenon of concern here, schizophrenic symptomatology, also shows such periodicity (i.e., Dalton, 1959; Glass et al., 1971; Jacobs & Charles, 1970).

eral of these findings in terms of cyclical variations in fatigue over the cycle. Parlee (1974) and Sommer (1973) have enunciated a social expectancy hypothesis that might account for the crime, accident, and psychiatric crises findings. Based on the known relationship between KAE as an index of trait and these same domains, we are here simply advancing a new, promising hypothesis. No claim is made that all possible alternative hypotheses have been precluded. We suggest that SIM may provide a useful theoretical foothold for understanding certain types of behavioral variation across the menstrual cycle. Further research assessing the implications of SIM formulations for the menstrual cycle seems warranted.

A General Implication for Social Expectancy Research

Finally, this study has a general implication for research areas in which social expectancy effects are a possible confounding variable. Undeniably, there is a cultural stereotype that certain kinds of pain and discomfort occur at the paramenstruum and another stereotype that behavioral decrements of various sorts (for example, poorer work performance) also occur at this time. But either or both of these stereotypes may also reflect true underlying phenomena. One approach to this problem is (a) to generate a hypothesis based on the assumption that a cultural stereotype—in this case, that the internal stimulation of pain and discomfort is greatest at the paramenstruum—is, at least in part, true and (b) then to test this hypothesis with respect to some measure—such as KAE—that is sufficiently unfamiliar to the subject so that no systematic social expectations exist. In general, a positive finding would offer support for the idea that the particular cultural stereotype involved reflects, at least to some degree, a true, underlying phenomenon. In this study, our positive findings regarding KAE offer some support for the assumption that aversive internal stimulation is greatest at the paramenstruum.

References

Asso, D., & Beech, H. R. Susceptibility to the acquisition of a conditioned response in relation to the menstrual cycle. *Journal of Psychosomatic Research*, 1975, *19*, 337–344.

Bakan, P., & Thompson, R. On the relation between induced and residual kinesthetic aftereffects. *Perceptual and Motor Skills*, 1962, *15*, 391–396.

Bakan, P., & Thompson, R. Attention and kinesthetic aftereffects. *American Journal of Psychology*, 1965, *76*, 39–45.

Baker, A. H., Mishara, B., Kostin, I., & Parker, L. Kinesthetic aftereffect and personality: A case study of issues involved in construct validation. *Journal of Personality and Social Psychology*, 1976, *34*, 1–13.

Baker, A. H., Mishara, B., Parker, L., & Kostin, I. When "reliability" fails, must a measure be discarded?—The case of kinesthetic aftereffect. *Journal of Research in Personality*, 1978, *12*, 262–273.

Barthol, R. Individual and sex differences in cortical conductivity. *Journal of Personality*, 1958, *26*, 366–378.

Buchsbaum, M., & Silverman, J. Stimulus intensity control and the cortical evoked response. *Psychosomatic Medicine*, 1968, *30*, 12–22.

Charles, J., & Duncan, C. The distance gradient in kinesthetic figural aftereffect. *Journal of Experimental Psychology*, 1959, *57*, 164–170.

Cohen, J. *Statistical power analysis for the behavioral sciences*. New York: Academic Press, 1969.

Compton, N. H. Perceptual characteristics of delinquent girls. *Perceptual and Motor Skills*, 1967, *24*, 596–598.

Coppen, W., & Kessel, N. Menstruation and personality. *British Journal of Psychiatry*, 1963, *109*, 711–721.

Dalton, K. Menstruation and acute psychiatric illness. *British Medical Journal*, 1959, *1*, 148–149.

Dalton, K. Menstruation and crime. *British Medical Journal*, 1961, *2*, 1752–1753.

Dalton, K. *The premenstrual syndrome*. Springfield, Ill.: Chares C Thomas, 1964.

Deaux, E. A psychophysiological influence on the choice of illicit drugs. *Physiological Psychology*, 1976, *4*, 400–404.

Depue, R., & Fowles, D. Electrodermal activity and schizophrenia: The problem of stimulus intensity modulation. *Psychological Bulletin*, 1976, *83*, 192–193.

Diamond, M., Diamond, L., & Mast, M. Visual sensitivity and sexual arousal levels during the menstrual cycle. *Journal of Nervous and Mental Disease*, 1972, *155*, 170–176.

Gardner, W., & Licklider, J. Auditory analgesia in dental operations. *Journal of the American Dental Association*, 1959, *59*, 1144–1149.

Glanville, E., & Kaplan, A. Taste perception and the menstrual cycle. *Nature*, 1965, *205*, 930–931.

Glass, G., Heninger, G., Lansky, M., & Talan, K. Psychiatric emergency related to the menstrual

cycle. *American Journal of Psychiatry*, 1971, *128*, 705–711.

Golub, S. The effect of premenstrual anxiety and depression on cognitive function. *Journal of Personality and Social Psychology*, 1976, *34*, 99–104.

Gupta, B. S. Stimulant and depressant drugs on kinesthetic figural aftereffect. *Psychopharmacologia*, 1974, *36*, 275–280.

Hazouri, L., & Mueller, A. Pain threshold studies on paraplegic patients. *Archives of Neurological Psychiatry*, 1950, *64*, 607–613.

Higgins, J., Tursky, B., & Schwartz, G. Shock-elicited pain and its reduction by concurrent tactile stimulation. *Science*, 1971, *172*, 866–867.

Jacobs, T., & Charles, E. Correlation of psychiatric symptomatology in an outpatient population. *American Journal of Psychiatry*, 1970, *126*, 1504–1508.

Klaiber, E., Broverman, D., Vogel, W., & Kobayashi, Y. Rhythms in plasma MAO activity, EEG, and behavior during the menstrual cycle. In M. Ferin et al. (Eds.), *Biorhythms and human reproduction*. New York: Wiley, 1974.

Kutner, S., Phillips, N., & Haag, E. Oral contraceptives, personality and changes in depression. *Contraception*, 1971, *4*, 327–336.

Lavine, R., Buchsbaum, M., & Poncy, M. Auditory analgesia: Somatosensory evoked response and subjective pain rating. *Psychophysiology*, 1976, *13*, 140–148.

Lennane, K., & Lennane, R. Alleged psychogenic disorders in women—A possible manifestation of sexual prejudice. *New England Journal of Medicine*, 1973, *228*, 288–292.

Lukas, J., & Siegel, J. Cortical mechanisms that augment or reduce evoked potentials in cats. *Science*, 1977, *198*, 73–74.

McDonald, A. The lack of effect of D-amphetamine on perceptual reactance and personality. *Journal of Abnormal Psychology*, 1974, *83*, 87–90.

McNemar, Q. *Psychological statistics* (4th ed.). New York: Wiley, 1969.

Melzack, R., Weisz, A., & Sprague, L. Stratagems for controlling pain: Contributions of auditory stimulation and suggestion. *Experimental Neurology*, 1963, *8*, 239–247.

Mersky, H., & Evans, P. Variations in pain complaint threshold in psychiatric and neurological patients with pain. *Pain*, 1975, *1*, 73–79.

Mishara, B., & Baker, A. H. *Stimulus intensity modulation: A perceptual cognitive approach to life style in the elderly* (Research Bulletin 74-35). Princeton, N.J.: Educational Testing Service, 1974.

Mishara, B., & Baker, A. H. Kinesthetic aftereffect scores are reliable. *Applied Psychological Measurement*, 1978, *2*, 239–247.

Mishara, B., Baker, A. H., Parker, L., & Kostin, I. W. Kinesthetic figural aftereffects: Norms from four samples, and a comparison of methods for classifying augmenters, moderates and reducers. *Perceptual and Motor Skills*, 1973, *37*, 315–325.

Moos, R. The development of a menstrual distress questionnaire. *Psychosomatic Medicine*, 1968, *30*, 853–867.

Moos, R., Kopell, B., Melges, F., Yalom, I., Lunde, D., Clayton, R., & Hamburg, D. Fluctuations in symptoms and moods during the menstrual cycle. *Journal of Psychosomatic Research*, 1969, *13*, 37–44.

Morgan, A., & Hilgard, E. The lack of retest reliability for individual differences in the kinesthetic aftereffect. *Educational and Psychological Measurement*, 1972, *32*, 871–878.

Morris, N., & Udry, J. Depression of physical activity by contraceptive pills. *American Journal of Obstetrics and Gynecology*, 1969, *104*, 1012–1014.

Morris, N., & Udry, J. Variations in pedometer activity during the menstrual cycle. *Obstetrics and Gynecology*, 1970, *35*, 199–201.

Morton, J., Additon, H., Addison, R., Hunt, T., & Sullivan, J. A clinical study of premenstrual tension. *American Journal of Obstetrics and Gynecology*, 1953, *65*, 1182–1191.

O'Connor, J., Shelly, E., & Stern, T. Behavioral rhythms related to the menstrual cycle. In M. Ferin et al. (Eds.), *Biorhythms and human reproduction*. New York: Wiley, 1974.

Parlee, M. The premenstrual syndrome. *Psychological Bulletin*, 1973, *80*, 454–465.

Parlee, M. Stereotypic beliefs about menstruation: A methodological note on the Moos menstrual distress questionnaire and some new data. *Psychosomatic Medicine*, 1974, *36*, 229–240.

Petrie, A. Some psychological aspects of pain and the relief of suffering. *Annals of the New York Academy of Science*, 1960, *86*, 13–27.

Petrie, A. *Individuality in pain and suffering*. Chicago: University of Chicago Press, 1967.

Petrie, A., Collins, W., & Solomon, P. Pain sensitivity, sensory deprivation, and susceptibility to satiation. *Science*, 1958, *128*, 1431–1433.

Petrie, A., Holland, T., & Wolk, I. Sensory stimulation causing subdued experience-audioanalgesia and perceptual augmentation and reduction. *Journal of Nervous and Mental Disease*, 1963, *127*, 312–321.

Petrie, A., McCulloch, R., & Kazdin, P. The perceptual characteristics of juvenile delinquents. *Journal of Nervous and Mental Disease*, 1962, *134*, 415–421.

Platt, D., Holtzman, P., & Larson, D. Individual consistencies in kinesthetic figural aftereffects. *Perceptual and Motor Skills*, 1971, *32*, 787–795.

Poser, E. Figural aftereffect as a personality correlate. *Proceedings of the XVIth International Congress of Psychology*. Amsterdam: North Holland, 1960.

Presser, H. Temporal data relating to the human menstrual cycle. In M. Ferin et al. (Eds.), *Biorhythms and human reproduction*. New York: Wiley, 1974.

Redgrove, J. Menstrual cycle. In W. Colquhoun (Ed.), *Biological rhythms and human performance*. London: Academic Press, 1971.

Ryan, E., & Foster, R. Athletic participation and perceptual reduction and augmentation. *Journal of Personality and Social Psychology*, 1967, *6*, 472–476.

Sales, S. Need for stimulation as a factor in social behavior. *Journal of Personality and Social Psychology*, 1971, *19*, 124–134.

Sales, S. Need for stimulation as a factor in preferences for different stimuli. *Journal of Personality Assessment*, 1972, *36*, 55–61.

Sales, S., Guydosh, R., & Iacona, W. Relationship between strength of the nervous system and the need for stimulation. *Journal of Personality and Social Psychology*, 1974, *29*, 16–22.

Sales, S., & Throop, W. Relationship between kinesthetic aftereffect and strength of the nervous system. *Psychophysiology*, 1972, *9*, 492–497.

Satran, R., & Goldstein, M. Pain perception: Modification of threshold of intolerance and cortical potentials by cutaneous stimulation. *Science*, 1973, *180*, 1201–1202.

Schooler, C., Buchsbaum, M., & Carpenter, W. Evoked response and kinesthetic measures of augmenting/reducing in schizophrenics: Replications and extensions. *Journal of Nervous and Mental Disease*, 1976, *163*, 221–232.

Schooler, C., & Silverman, J. Perceptual styles and their correlates among schizophrenic patients. *Journal of Abnormal Psychology*, 1969, *74*, 459–470.

Sheldrake, P., & Cormack, M. Variations in menstrual cycle symptom reporting. *Journal of Psychosomatic Research*, 1976, *20*, 169–177.

Silverman, J. Variations in cognitive control and psychophysiological defense in schizophrenics. *Psychosomatic Medicine*, 1967, *29*, 225–251.

Silverman, J., Buchsbaum, M., & Henkin, R. Stimulus sensitivity and stimulus intensity control. *Perceptual and Motor Skills*, 1969, *28*, 71–78.

Sommer, B. The effect of menstruation on cognitive and perceptual-motor behavior: A review. *Psychosomatic Medicine*, 1973, *35*, 515–534.

Spitz, H., & Lipman, R. A. Comparison of mental retardates and normals on kinesthetic aftereffects. *Journal of Abnormal & Social Psychology*, 1961, *62*, 686–687.

Sweeney, D. Pain reactivity and kinesthetic aftereffect. *Perceptual and Motor Skills*, 1966, *22*, 763–769.

Vando, A. The development of the R–A scale: A paper-and-pencil measure of pain tolerance. *Personality and Social Psychology Bulletin*, 1974, *1*, 28–29.

Vierling, J., & Rock, J. Variations in olfactory sensitivity to exaltolide during the menstrual cycle. *Journal of Applied Physiology*, 1967, *22*, 311–315.

Vollman, R. Some conceptual and methodological studies in human reproduction. In M. Ferin et al. (Eds.), *Biorhythms and human reproduction*. New York: Wiley, 1974.

Weintraub, D., Green, G., & Herzog, T. Kinesthetic aftereffects day by day: Trends, task features, reliable individual differences. *American Journal of Psychology*, 1973, *86*, 827–844.

Winer, B. *Statistical principles and experimental design*. New York: McGraw-Hill, 1962.

Wong, S., & Tong, J. Menstrual cycle and contraceptive hormonal effects on temporal discrimination. *Perceptual and Motor Skills*, 1974, *39*, 103–108.

Zimmerman, E., & Parlee, M. Behavioral changes associated with the menstrual cycle: An experimental investigation. *Journal of Applied Social Psychology*, 1973, *3*, 335–344.

Received February 24, 1978

Journal of Personality and Social Psychology
1979, Vol. 37, No. 2, 247–260

Evaluative Judgments of Aspects of Life as a Function of Vicarious Exposure to Hedonic Extremes

Marshall Dermer, Sidney J. Cohen, Elaine Jacobsen, and Erling A. Anderson
University of Wisconsin—Milwaukee

In two experiments, the hypothesis was corroborated that vicarious exposure to hedonic extremes—especially the hedonically negative—results in contrast regarding evaluative judgments of aspects of life that have evolved or been acquired in the course of life *beyond* the laboratory. In Experiment 1, participants who wrote about hedonically negative events occurring at the turn of the century expressed greater satisfaction on a composite index of present life quality than participants who wrote about hedonically positive events. In Experiment 2, participants who wrote about hedonically negative events, personal tragedies, scored higher on a composite index of satisfaction with life, health, and physical appearance than participants who wrote about hedonically positive events. The findings for the composites corroborate a comparison level model of evaluative judgment. The findings for individual items, however, suggest that aspects of life are not evaluated in terms of a single utility scale and standard—the comparison level. Other findings are discussed that appear to contradict a simple affective model of evaluation in which the positivity of evaluations is postulated to increase with the positivity of affective states.

A number of psychologists have generalized perceptual (Helson, 1964) and judgmental (Volkmann, 1951) principles, corroborated in the traditional experimental psychology laboratory, to social judgment. The most well-known generalization is that of Thibaut and Kelley (1959) who proposed a theory of the evaluation of outcomes resulting from social interaction. Although outcomes might differ in their specifics, it was assumed that all outcomes could be characterized in terms of their utility or hedonic value. The comparison level was conceptualized as a "psychologically meaningful midpoint for the scale of outcomes—a neutral point of satisfaction–dissatisfaction," (p. 81) and was defined as the "average value of all the outcomes known to the person (by virtue of personal or vicarious experience), each outcome weighted by its salience (or the degree to which it is instigated for the person at the moment)" (p. 81). An outcome, therefore, is judged to be positive or negative to the extent that its hedonic value is, respectively, above or below the comparison level.

Comparison level theory has been corroborated in a number of studies in which satisfaction with outcomes attained *within* experiments has increased as a function of manipulations hypothesized to lower the comparison level (e.g., Brickman, 1975; Friedland, Arnold, & Thibaut, 1974). Judgmental theories, however, apply to the evaluation of outcomes that are of greater importance to individuals than those usually bestowed upon participants in laboratory experiments. Thibaut and Kelley, for example, illustrated

This research was supported by grants from the Graduate School of the University of Wisconsin—Milwaukee to Marshall Dermer. Sidney Cohen and Erling Anderson contributed to the design and execution of Experiment 1, and Elaine Jacobsen contributed to the design and execution of Experiment 2. Special thanks are due Ellen Berscheid, Philip Brickman, and Harry Upshaw for their encouragement and Joni Gonnering for skillfully transcribing all written materials.

Requests for reprints should be sent to Marshall Dermer, Department of Psychology, University of Wisconsin—Milwaukee, Milwaukee, Wisconsin 53201.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3702-0247$00.75

their theory in terms of marital satisfaction. Volkman (1951, p. 286) also presumed the generality of judgmental principles. More recently, Brickman and Campbell (1971) generalized adaptation level theory (a conceptual cousin of comparison level theory) to satisfaction with, for example, personal wealth and competence. These generalizations are intriguing, but are without direct experimental support.

In this report, two experiments are presented in which the generality of comparison level theory was tested for evaluations of *important and often vital life outcomes that have evolved or been acquired in the course of a participant's life beyond the laboratory.* Furthermore, a condition in which participants were vicariously exposed to the hedonically negative was common to both experiments. The importance of this condition in testing comparison level theory, for this judgmental domain, is emphasized.

Experiment 1 [1]

In the past 10 years, a social indicator movement has evolved in which it has been proposed that the well-being of the nation can best be assessed by measuring a broad range of variables. Andrews and Withey's (1976) respondents, in particular, indicated their feelings regarding various aspects of life along rating scales similar to those used in laboratory studies of contrast. Furthermore, the "objects" of judgment (for example, the respondent's life as a whole and standard of living) differed from those of previous laboratory studies of contrast. The study of contrast in relation to judgments of the quality of life offers, consequently, an opportunity to replicate the response assessment methodologies of previous experiments and to generalize comparison level theory to a new class of stimuli. Furthermore, although the theoretical orientation of the present report tends toward linguistic rather than perceptual or subjective interpretations of contrast (see Manis, 1971) and comparison level theory (see Upshaw, 1969), as Andrews and Withey (1976) have noted, "anything that can be done to improve the human lot that is reflected as felt improvement is a condition to be coveted" (p. 10).

Hedonic Manipulation

In the present experiment, comparison level theory was tested by presenting historical information regarding life at the turn of the century. In the "good old days" condition, life was presented as the embodiment of traditional American ideals. For example, educational instruction was described as personalized, and students were depicted as having excellent foundations in reading, writing, and arithmetic; the air was clean and the water pure; there was time for people to genuinely relate to one another; food was wholesome and nutritious; people enjoyed their work and took pride in the products and services they provided; and government was efficient. In the "bad old days" condition, a description was presented of what historically appears to have been the plight of the vast majority of Americans. This information and the contrast principle are described in Otto Bettmann's (1974) prefatory remarks to his *The Good Old Days—They Were Terrible!*:

> I have always felt that our times have overrated and unduly overplayed the fun aspects of the past. What we have forgotten are the hunger of the unemployed, crime, corruption, the despair of the aged, the insane and the crippled . . . In most of our nostalgia books . . . the period's dirty business is swept under the carpet of oblivion. What emerges is a glowing picture of the past, of blue-skied meadows where children play and millionaires sip tea.
>
> If we compare this purported Arcadia with our own days we cannot but feel a jarring discontent, a sense of despair that fate has dropped us into the worst of all possible worlds. (pp. xii–xiii)

Female students evaluated the present quality of life at the beginning of the semester and some weeks later, immediately after having been exposed to one of the experimental conditions. It was generally hypothesized, on the basis of comparison level theory, that pretreatment-to-posttreatment changes in evaluations would be more positive in the

[1] Special thanks are due Linda S. Mezydlo and David Stamm for preparing stimulus materials and helping conduct this experiment.

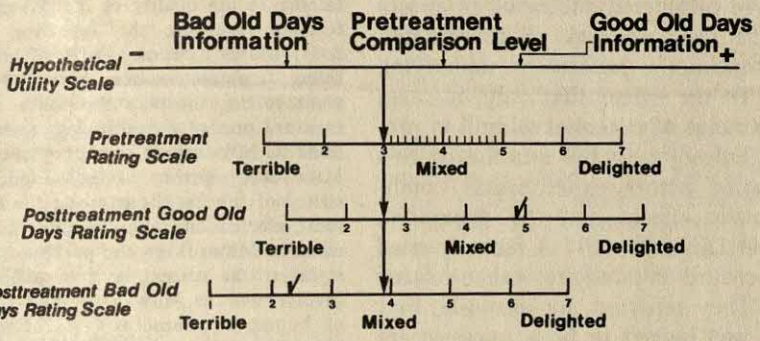


Figure 1. Location of pretreatment comparison level and information presented in experimental conditions, in relation to a hypothetical utility scale and rating scales.

bad than in the good old days condition. In particular, it was hypothesized that evaluations of present life quality would increase from pretreatment to posttreatment in the bad old days condition (positive contrast) but tend to decrease (negative contrast) somewhat or remain unchanged in the good old days condition.

The prediction of differential absolute magnitudes of contrast across experimental conditions and therefore a *stronger* test of comparison level theory in the hedonically negative than in the positive condition is illustrated in Figure 1. As indicated on the uppermost horizontal (utility) scale, the good old days information was assumed, by virtue of conveying traditional American values, to be closer in hedonic value to participants' pretreatment comparison levels than the bad old days information. The second horizontal line represents the pretreatment rating scale, centered at the hedonic value of the pretreatment comparison level. As Upshaw (1969, p. 347) has noted, comparison level theory implicitly assumes that the "origin" or midpoint of each judge's reference scale is anchored at the comparison level. The third horizontal line represents the posttreatment rating scale in the good old days condition, which has been shifted about three eighths of a unit upward in relation to the pretreatment rating scale. This shift follows from the definition of the comparison level as the average of all known outcomes, each outcome weighted by its salience, although the absolute magnitude of the shift is arbitrary. The fourth horizontal line represents the posttreatment rating scale in the bad old days condition, which has been shifted about seven eighths of a unit downward in relation to the pretreatment rating scale. Again, the absolute magnitude of the shift is arbitrary but is consistent with the definition of the comparison level as an average—the assumed hedonic value of the bad old days information and the magnitude of the shift postulated in the good old days condition.

The vertical line cutting across the three rating scales in Figure 1 represents an aspect of life that was evaluated 3 on the premeasure. As a consequence of the hypothesized scale shifts, its posttreatment evaluation is 2.7 in the good old days condition (a shift of .3 unit downward) and 3.9 in the bad old days condition (a larger shift of .9 unit upward). Finally it is important to note that although the posttreatment scales are postulated to shift, the assumptions just mentioned imply that posttreatment evaluations of the bad old days information in the bad old days condition should more greatly depart from the posttreatment scale midpoint than corresponding evaluations in the good old days condition. This implication is illustrated by the location of check marks on the posttreatment scales.

Scale Anchoring

Mere prior exposure to a stimulus complex that contrasts with a stimulus under evaluation has not generally been considered sufficient for generating contrast effects (see Eiser & Stroebe, 1972, p. 48). Judgmental

theorists have considered anchoring processes in which the response scale is coordinated with the continuum putatively underlying judgments. To the extent that only one segment of the range of potential stimuli is presented *and* judged, with the location of this segment varied across experimental conditions, linguistic effects may be generated. Pepitone and DiNubile (1976) recently used such a procedure in studying crime-related judgments. They reported, for example, that a homicide was judged to be a more severe criminal violation when participants first had read *and* publicly judged the seriousness of an assault case than in a condition in which the first case was another homicide. The overt recording of the first judgment appeared to be a necessary condition for producing a contrast effect.

In the present experiment, an anchoring manipulation was also included. After participants had been exposed to the stimulus materials and had written about them, all participants indicated their feelings regarding the present quality of life in Milwaukee. In the unanchored condition, participants simply expressed their judgments in terms of aspects of present life quality. In the anchored condition, participants were required to first evaluate an aspect of their present life in terms of the quality of life in Milwaukee in 1900, *immediately before* expressing their judgment of the same aspect of present life quality. Theoretically, this manipulation should increase the salience of the hedonic information and consequently result in greater contrast because of greater shifts in comparison levels in the anchored than in the unanchored conditions.

Method

Participants

Seventy-three women enrolled in introductory psychology at the University of Wisconsin—Milwaukee participated in the experiment in exchange for extra credit in their courses. Women were selected as a matter of convenience.

Pretreatment Measures

At the beginning of the semester, a survey was distributed that included a series of questions pertaining to the quality of life. Women indicated their feelings regarding the following aspects of their lives: life as a whole; personal health; standard of living; amount of time for doing the things they want to do; the way the police and courts in this area are presently operating; present working conditions in Milwaukee; present quality of education in Milwaukee; quality of their food; the present efficiency of the fire department; the honesty of present local, elected officials; the present physical environment in Milwaukee—the purity of the air, lakes, and streams; the present level of public spirit and civic involvement in Milwaukee; and the present quality of human relationships (e.g., between family members, doctors and patients, and merchants and customers). Items were included that appeared relevant to the hedonic manipulation. Respondents answered each question by placing a check mark *anywhere* along a 67-point rating scale similar to that of Andrews and Withey (1976, p. 18):

1.....2.....3.....4.....
Terrible Unhappy Mostly dissatisfied Mixed

.....5.....6.....7
Mostly satisfied Pleased Delighted

Those women whose ratings were between Unhappy (12) and Pleased (56) were invited to participate several weeks later and randomly assigned to conditions.

Procedure

Participants were run in small groups ranging in size from 2 to 10 women and responded independently to materials. Participants were told that their conceptions of life in Milwaukee at the turn of the century were being studied. Specifically, they would be asked to write a description of a day in the life of a typical Milwaukeean living in 1900. Participants were told that since they were probably not intimately familiar with life in 1900, materials describing this era would be presented. Participants were instructed to imagine vividly what life was like in Milwaukee in 1900 and to take notes if they wished.

Audiovisual Materials and Presentation Procedure

In all conditions, participants viewed an identical series of 16 slides; materials susceptible to multiple interpretations were used. Accompanying each slide was a tape-recorded narrative by which hedonic information was manipulated. In the good old days condition, life was described positively, whereas in the bad old days condition negative aspects were emphasized. For example, portions of the narrative associated with a slide of an elaborate fountain

park were as follows:

(*Good old days*). Beautifully carved fountains such as this one on Kilbourn Avenue provided a cooling respite for Milwaukeeans and contributed to the relaxed, slow-paced, leisurely atmosphere of the turn of the century. Note the graceful, flowing lines of the artfully sculpted statue that was characteristic of the quality craftsmanship of the European immigrants. Fountains like this one supplied a thirsty public with cool drafts of crystal-clear spring or pure well water on sunny afternoons.

(*Bad old days*). This fountain, located in the middle of a dirt highway later known as Kilbourn Avenue, is an example of the poor public hygiene practices of the turn of the century. Horses, birds, dogs, as well as people, drank and washed themselves in this public germ spreader. People all drank from the same filthy communal drinking cup. It was not known for years that fountains such a this one and the public drinking cups used with them were transmission vehicles for tuberculosis, smallpox, and other dreaded diseases which tore at the heart of the community by exiling infected people to isolation, or to convalescent hospitals for extended terms. Most people struck down and killed by tuberculosis were between the ages of 18 and 30.

Some narratives were longer than others, but for a given slide the narratives were of approximately equal length across experimental conditions. After the presentation of each slide, participants were given 15 sec to vividly imagine the scene without the accompanying narrative. Overall, participants spent a total of 32 min viewing the slides, listening to the narratives, and taking notes.

Descriptive Task

A booklet was next distributed, in which participants first described their conceptions of a day in the life of a typical Milwaukeean by indicating what this person would be doing at 1-hour intervals beginning at 6:00 A.M. when the person woke up, and finishing at midnight when the person retired to bed. Only upon completing the descriptive task were participants instructed to proceed.

Posttreatment Measures and Anchoring Manipulation

Two sets of measures were assessed after the descriptive task. The first (Present) set was identical to the pretreatment measures. The second (Past) set was similar to the Present set, but participants judged their life in terms of the quality of life prevalent in 1900. For example, the Past item corresponding to the first pretreatment measure was: "How would you feel about your life as a whole if it were similar in quality to that of a typical Milwaukeean living in 1900?" The questions were introduced as a way that the researchers might better understand participants' conceptions and feelings regarding life in Milwaukee.

In the unanchored sequence, participants first completed the 13 items in the Present set before responding to those in the Past set. In the anchored sequence, participants completed the first item in the Past set before completing the first item in the Present set. The remaining items were presented in this alternating fashion.

Debriefing

After completing their ratings, participants were informed that for experimental purposes, a balanced historical description had not been presented and that four volumes depicting life in Milwaukee in 1900 had been placed on reserve at the library. Participants were encouraged to approach the experimenter if they had any questions they wanted immediately answered and to contact the first author if they had questions later. Finally, participants were asked not to discuss the study with others. After the data had been analyzed, a complete description of the study and the findings was mailed to all participants.

Results

Manipulation Check

Responses were averaged across the 13 Past items to form an index of the judged quality of life in 1900 (coefficient alpha, based on the pooled within-cell covariance matrix, was .91). A univariate Hedonic × Anchoring Condition analysis of variance revealed only a reliable hedonic main effect, $F(1, 69) = 181.11$, $p < .0001$. Participants in the good old days condition indicated that if their present lives were similar in quality to the life portrayed, they would feel on the average "mostly satisfied" ($M = 45.4$), or about 11 points above the scale midpoint (34), whereas those in the bad old days condition would feel "unhappy" ($M = 15.1$), or about 19 points below the scale midpoint. These findings are consistent with the assumed hedonic value of the stimulus materials, as illustrated by the location of the materials on the utility scale in Figure 1 and the check marks on the accompanying rating scales.

It is also important to note that univariate analyses of variance for each item in the Past set revealed reliable hedonic effects ($p <$

Table 1
Average Judgments of Present Life Quality—Composite

Condition	*n*	Pre-treatment	Post-treatment	Difference	$F(1, 69)$	*p*
Unanchored						
Good old days	19	39.7	39.7	.0	.00	1.00
Bad old days	18	38.5	41.0	2.5	3.49	.07
Anchored						
Good old days	18	39.7	37.9	−1.8	1.78	.19
Bad old days	18	39.1	44.9	5.8	18.56	.00005

Note. Scores range from 1 (Terrible), through 34 (Mixed), to 67 (Delighted).

.0001) generally consistent with the overall index. The major exception was for evaluations of the fire department, in which the evaluations in the good old days condition ($M = 32.4$), although more positive than those in the bad old days condition ($M = 10.4$), were just below the scale midpoint. Finally, it is important to note that the seven largest F ratios were detected for the personal health, police and courts, food, elected officials, physical environment, public spirit, and human relationship items.

Major Analyses

Life quality composite. Responses were averaged across the 13 pretreatment items (coefficient alpha = .70) and the 13 posttreatment items (coefficient alpha = .81) to form indexes of present life quality. These composites are presented in Table 1. Analysis of pretreatment-to-posttreatment differences revealed a hedonic main effect, $F(1, 69) = 14.12$, $p = .0004$, as well as a Hedonic × Anchoring interaction trend, $F(1, 69) = 3.62$, $p = .06$. The test for the anchoring main effect was not reliable ($p = .56$).

The reliable hedonic main effect is due to pretreatment-to-posttreatment change in evaluations being more positive in the bad ($M = 4.2$) than in the good ($M = -.9$) old days conditions. This finding corroborates the most general comparison level prediction. Furthermore, the more specific hypothesis of greater positive contrast in the bad old days condition than negative contrast in the good old days condition was also corroborated. Inspection of the tests in Table 1 for pretreatment-to-posttreatment simple effects explicitly reveals that the positive contrast hypothesis was weakly corroborated in the unanchored bad old days condition ($p = .07$) and strongly corroborated in the anchored bad old days condition ($p = .00005$). Negative contrast, that judgments would decline in positivity after participants were exposed to the good old days information, was, however, detected in neither the unanchored ($p = 1.00$) nor the anchored condition ($p = .19$). The Hedonic × Anchoring interaction trend is, as predicted, due to the difference between the good and bad old days conditions being more enhanced when judgments were anchored (the difference is 7.6) than when they were unanchored (2.5).

Individual items. To investigate whether all items were uniformly influenced by the treatments, pretreatment-to-posttreatment difference scores were calculated for each item and a multivariate analysis of variance was conducted.

A reliable multivariate anchoring main effect was detected, $F(13, 57) = 2.04$, $p = .03$. This effect appears to be due to evaluations of the fire department increasing in the anchored ($M = 8.8$) as compared to the unanchored ($M = 2.0$) conditions ($p < .003$).

Of greater interest, a reliable multivariate hedonic main effect was also detected, $F(13, 57) = 2.16$, $p = .02$. Pretreatment-to-posttreatment difference scores were reliably (at $p \leq .05$) more positive in the bad than in the good old days condition for all items listed in Table 2 except personal health ($p = .06$). Inspection of the Difference column in Table 2 for pretreatment-to-posttreatment

Table 2
Average Judgments of Present Life Quality—Individual Items

Item and condition	Pre-treatment	Post-treatment	Difference	$F(1, 69)$	p
Personal health					
Good old days	48.0	47.2	−0.8	0.26	.61
Bad old days	44.9	48.5	3.6	4.70	.03
Police and courts					
Good old days	36.4	31.7	−4.7	5.25	.02
Bad old days	33.1	35.9	2.8	1.81	.18
Quality of food					
Good old days	43.7	41.7	−2.0	1.15	.29
Bad old days	43.9	47.2	3.3	3.06	.08
Elected officials					
Good old days	33.8	28.3	−5.5	7.50	.008
Bad old days	33.6	33.9	0.3	0.02	.89
Physical environment					
Good old days	28.0	26.7	−1.3	0.52	.47
Bad old days	23.4	31.0	7.6	17.33	.00009
Public spirit					
Good old days	33.1	31.8	−1.3	0.48	.49
Bad old days	30.0	36.2	6.2	10.86	.002
Human relationships					
Good old days	37.0	32.2	−4.8	3.97	.05
Bad old days	34.9	43.4	8.5	12.14	.0009

Note. Scores range from 1 (Terrible), through 34 (Mixed), to 67 (Delighted).

changes in judgments reveals that, generally, evaluations tended to decrease in the good old days condition and increase in the bad old days condition. Furthermore, inspection of the F ratios indexing these changes suggests that, generally, increments in judgments (positive contrast) in the bad old days condition tended to be larger than decrements (negative contrast) in the good old days condition.

The hedonic main effect and pretreatment-to-posttreatment simple effect findings described above are consistent with the analyses reported for the composite. The multivariate test for the Hedonic × Anchoring interaction was, however, not reliable, $F(13, 57) = 1.49$, $p = .15$.

Discussion

Life Quality Composite

Comparison level theory most generally implies that pretreatment-to-posttreatment changes in judgments should be more positive in the bad than in the good old days condition. This is precisely what was detected, as indicated by the reliable test for the hedonic main effect, when pretreatment-to-posttreatment differences for the composite were analyzed. Furthermore, it was hypothesized that contrast effects would be greater in the bad than in the good old days condition. This is precisely what was detected, as indicated by the tests for pretreatment-to-posttreatment simple effects reported in the last two columns of Table 1. It must be noted, however, that whereas these tests are consistent with a more precise contrast hypothesis, the comparisons must be interpreted cautiously because of the absence of a no-treatment control group.

The results for the composite also support the hypothesis of greater contrast in the anchored than in the unanchored condition, although the test was just short of being reliable at conventional levels of significance. This finding is similar to that reported by Pepitone and DiNubile (1976), who detected

contrast only when judgments of an initial stimulus in a two-stimulus "sequence were 'anchored,' that is overtly recorded and thus publicly committed" (p. 448). It is not clear, however, on the basis of either study, which aspects of such anchoring manipulations are necessary for producing contrast. Whereas Pepitone and DiNubile have stressed public commitment, according to comparison level theory, merely requiring participants to judge *but not publicly record initial judgments* should be sufficient to induce or enhance contrast, by virtue of increasing the salience of the hedonic value of the prior stimulus.

Individual Items

The contrast hypotheses were corroborated for 7 of the 13 life quality items. The pattern of reliable hedonic effects for the set of items appears clearly due to differential manipulation of aspects of life. Those aspects of life along which the general contrast hypothesis was corroborated were those most strongly manipulated.

Finally, it will be recalled that a reliable anchoring effect was detected for the fire department item. This finding, that evaluations of the fire department increased more in the anchored than in the unanchored conditions *across* levels of the hedonic factor, can be reconciled with adaptation level theory if not with comparison level theory.[2] Recall that the manipulation check pertaining to the efficiency of the fire department indicated that evaluations were often below the scale midpoint in the good old days condition and even more so in the bad old days condition. The net result in either hedonic condition was to increase the salience of how bad the fire department was in 1900. In the anchored conditions, the manipulation checks were collected before the evaluation of the fire department's present efficiency, and thus, greater salience and positive contrast would be expected in this condition.

Theoretical Implications

The present findings, although corroborating comparison level theory, suggest that the theory may not be sufficiently refined to adequately represent the complexities of evaluative judgment. Upshaw (1969) has emphasized an important theoretical distinction between comparison and adaptation level theories. Although outcomes might differ in their specifics, according to comparison level theory, all outcomes are judged in terms of a single utility scale such that "a nagging wife, a 10% salary raise, and a slice of apple pie are all evaluated in terms of a single CL [comparison level]" (pp. 347–349). This conception of the comparison level as a generalized hedonic standard of judgment would imply that contrast effects should have been detected, within the limits of experimental error, across all 13 measures of life quality. As indicated, however, the general contrast hypothesis was corroborated for only 7 of the measures. (The differential impact of the hedonic manipulation upon the measures is not due to differential experimental error.)[3] According to an adaptation level model of judgment, there is a different reference scale for each aspect of life, with each scale having a particular adaptation level (Upshaw, 1969, p. 347). This model of judgment is consistent with the finding that the strongest contrast effects were observed along those dimensions that were most strongly manipulated, and the anchoring main effect detected for evaluations of the efficiency of the fire department.

Experiment 2[4]

In Experiment 2, we attempted to constructively replicate Experiment 1 by essentially following Abraham Maslow's (Maslow, 1972) suggestion regarding exercises in deprivation:

> All you have to do is to go to a hospital and hear all the simple blessings that people never before realized *were* blessings—being able to urinate, to

---

[2] Upshaw (1969) has distinguished between adaptation and comparison level models of judgment. This distinction is presented in the next section.

[3] These error terms are for the analyses of pretreatment-to-posttreatment difference scores.

[4] Special thanks are due Margaret Grade and Betty McMackin for help in conducting this experiment and Carol Schultz for her extensive assistance in collecting and developing stimulus materials.

eep on your side, to be able to swallow, to scratch n itch, etc. Could *exercises* in deprivation educate s faster about all our blessings? (p. 108)

Vomen who had completed pretreatment easures earlier in the semester were asked o imagine vividly a series of events. In the edonically negative condition, the events ere personal tragedies, whereas in the comarison condition the events were all positive. fter imagining *each* event, participants decribed what they would do, think of, and eel. Later, they expressed their level of satisaction with various aspects of life, including heir life in general, health, and physical apearance, on scales identical to the pretreatnent measures.[5] Given the results of an earlier xperiment (see Footnote 5), and given that he latter aspects of life appeared a priori to e most strongly manipulated across condiions, it was predicted that on a composite of hese measures, pretreatment-to-posttreatment changes in evaluation would be more positive in the hedonically negative than in the hedonically positive condition. This prediction, of course, is analogous to the general contrast prediction, which was strongly corroborated in Experiment 1 and was the major hypothesis tested in the study.

Participants also described their moods. According to a simple affective model of evaluation, the positivity of evaluative responses should covary directly with the positivity of affective states (Byrne, 1971, chapter 13). Both comparison and adaptation level theories, however, suggest that evaluations on the composite index would be most positive in the hedonically negative condition. The mood descriptions also served as a manipulation check.

Method

Participants

Eighty-three women enrolled in introductory psychology at the University of Wisconsin—Milwaukee participated in the experiment in exchange for extra credit in their courses. Women were again selected as a matter of convenience.

Pretreatment Measures

At the beginning of the semester, a survey was distributed that included a series of questions regarding the respondent's satisfaction with her life, health, physical appearance, relations with other people, sex life, and financial situation. Respondents answered each question by placing a check mark *anywhere* along the following rating scale:

1..... 2..... 3..... 4.....

Exceptionally dissatisfied | Very dissatisfied | Dissatisfied | Slightly dissatisfied

.... 5..... 6..... 7..... 8

Slightly satisfied | Satisfied | Very satisfied | Exceptionally satisfied

Responses to this 71-point scale were assigned integers from 10 (Exceptionally dissatisfied) to 80 (Exceptionally satisfied). Those women whose ratings were between 30 and 65 were invited to participate several weeks later and were randomly assigned to conditions. (Several women participated although they had not completed the "sex life" question.)

Procedure

Participants were run in small groups ranging in size from two to seven persons and responded independently to the materials. Each session lasted about 2 hours. All instructions and questions were contained in a single booklet.

Instructions. At the beginning of each session, participants were informed that they would be asked to read and respond to a series of articles and that their responses would be confidential. It was also explained that at any time they could decline to answer a question or participate further, but would receive full credit toward their psychology course.

Participants were told that the ability of people to vividly imagine "life events" was being studied. They anticipated reading a series of life events,

---

[5] Before the present series of experiments had been conducted, a study conceptually similar to Experiment 2 had been run (Fisher, 1976). Women role played either being slow walkers or having been in an automotive accident and confined to wheelchairs for the rest of their lives. After traveling about campus for an hour and discontinuing role playing, participants in the hedonically negative condition tended to express more positive judgments on a composite index of satisfaction with life, health, and physical appearance than participants in the control condition. For various methodological reasons (e.g., the absence of pretreatment measures and some participants failing to comply with the role-playing instructions) the difference between conditions was not reliable, $F(1, 78) = 1.65$, $p = .10$, directional alternative. We believed the findings to be suggestive, however, and consequently designed a more precise study—Experiment 2.

imagining the events happening to themselves and describing their reactions to the events.

Life events. In the hedonically negative condition, participants successively imagined (a) that they were severely burned and permanently disfigured—especially their face and hands—as a consequence of a gas explosion that destroyed their home or apartment and killed someone they dearly loved; (b) that they were blind; (c) that they were in an automobile accident that resulted in their confinement to a wheelchair for the rest of their lives; and (d) that they were severely suffering from Hodgkin's disease, a cancer of the lymphatic system that most often attacks young adults. In the hedonically positive condition, participants successively imagined that they were (a) winners of an all-expenses-paid tour for themselves and a friend to southwestern Europe and Morocco; (b) multimillionaires who with their loved one enjoyed a spectacular world cruise on the Queen Elizabeth 2; (c) well-paid private secretaries who travelled through northeastern Europe with their wealthy employer; (d) and winners of an all-expenses-paid vacation of their own design, for themselves and a friend, in Missouri.

To stimulate participants' imaginations, in the negative condition women read appropriate articles from the medical and rehabilitation literature, whereas in the positive condition they read travel brochures. Participants could not proceed to a new life event until they had read and described their reactions to earlier life events.

Posttreatment Measures

Upon completing the last life event, subjects were told that the role-playing task was over and that additional information was being collected to better understand responses. Participants first described how they felt while role playing, by completing the Nowlis (1970) Mood Adjective Check List. Later, they described their satisfaction with the six aspects of life assessed at the beginning of the semester, on identical rating scales. Participants then completed a second set of mood scales to describe their current mood. The last question in the booklet asked participants to describe anything that might help the researchers understand participants' responses. This question was included generally to enhance an understanding of the effects of the treatment and to check for experimental demand. Although many women wrote extensively, especially in the hedonically negative condition, no one indicated knowledge of any of the hypotheses being investigated.

Debriefing

After the data were analyzed, a complete description of the study and the findings was mailed to all participants. Those wishing additional information were invited to discuss the study with the researchers.

Table 3
Average Mood Scores as a Function of Experimental Condition and Phase

Factor	Condition: Hedonically negative (n = 42)	Condition: Hedonically positive (n = 41)	Univariate hedonic tests, $F(1, 81)$
	During role-playing task		
Elation	5.2	9.8	55.82
Anxiety	7.8	4.7	39.05
Sadness	7.8	4.4	45.86
	After satisfaction ratings		
Elation	6.3	6.3	0
Anxiety	5.9	4.1	13.15
Sadness	5.3	3.9	10.85

Note. High scores indicate greater elation, anxiety, and sadness. The range of possible scores for elation is 4 to 16, and for the remaining factors is 3 to 9. All nonzero F statistics are reliable at $p < .002$.

Results

Manipulation Check

Multivariate analyses of variance for mood during the role-playing task, $F(3, 79) = 34.40$, $p < .0001$, and after the satisfaction ratings were made, $F(3, 79) = 5.09$, $p < .003$, indicated reliable hedonic effects. As indicated in Table 3, participants' reports of mood were generally more negative in the hedonically negative than in the hedonically positive condition.

Major Analyses

Satisfaction composite. The three satisfaction measures hypothesized to be most influenced by the treatments were averaged to form a composite (the alpha coefficients for the pretreatment and posttreatment indexes were, respectively, .52 and .59). An analysis of variance of pretreatment-to-posttreatment differences revealed, as predicted, greater positivity in judgments of satisfaction in the hedonically negative than in the hedonically positive condition, $F(1, 81) = 3.32$, $p < .04$, directional alternative. As indicated in Table 4, participants' judgments of satisfaction increased an average of 4.9 points in the hedoni-

Table 4
Average Judgments of Satisfaction

Variable	Pre-treatment	Post-treatment	Difference	$F(1, 81)$	p
	Hedonically negative condition ($n = 42$)				
Composite	53.6	58.5	4.9	20.37	.00002
Life	55.0	61.1	6.1	18.78	.00004
Health	54.3	61.2	6.9	15.96	.0001
Physical appearance	51.4	53.2	1.8	1.77	.19
	Hedonically positive condition ($n = 41$)				
Composite	53.8	55.9	2.1	3.65	.06
Life	55.0	56.3	1.3	0.83	.36
Health	55.3	57.8	2.5	2.04	.16
Physical appearance	51.0	53.6	2.6	3.61	.06

Note. Scores range from 10 (Exceptionally dissatisfied), through 45, the midpoint, to 80 (Exceptionally satisfied).

cally negative condition but only 2.1 points in the positive condition. The tests for pretreatment-to-posttreatment simple effects, in the last two columns of Table 4, reveal the hedonically negative condition to be associated with the largest increment in evaluations.

It will be recalled that according to a simple affective model of evaluation, the positivity of evaluations increases with the positivity of affective states. The present findings contradict this model, since judgments of satisfaction were most positive in the hedonically negative condition. Nor does it appear that the positivity of evaluative responses increases with *changes* toward more positive affective states. According to such a revised model, the participants in the hedonically negative condition judged their lives most positively because of the affective "relief" resulting from the termination of the role-playing task. Contrary to the revised model, as participants' moods in the hedonically negative condition decreased (from first to second assessment) in terms of sadness and anxiety, their pretreatment-to-posttreatment changes in judged satisfaction, on the composite, *also decreased*, $r(40) = .34$, $p < .05$, for sadness; $r(40) = .30$, $p < .06$, for anxiety. Changes in elation did not covary ($r = -.01$) with changes in the composite.

Individual items. For each item, pretreatment-to-posttreatment difference scores were calculated. The omnibus multivariate hedonic test was reliable, $F(3, 79) = 2.65$, $p = .05$, as were the univariate tests for satisfaction with life, $F(1, 81) = 5.63$, $p < .01$, directional alternative, and health, $F(1, 81) = 3.23$, $p < .04$, directional alternative. As indicated in Table 4, pretreatment-to-posttreatment increments in judged satisfaction with life and health were greater in the hedonically negative than in the hedonically positive condition. The weak trend for judgments of satisfaction with physical appearance is contrary to that hypothesized and, of course, is not reliable, using a directional statistical decision rule, $F(1, 81) = .15$.

Similar analyses of pretreatment-to-posttreatment change were conducted for judgments of satisfaction with relations with other people and financial situation. Whereas the means for the relations item were in the direction specified by comparison level theory, $F(1, 81) = 2.00$, $p < .08$, directional alternative, the means for the financial item were in the opposite direction, $F(1, 81) = .02$. Furthermore, although the absence of pretreatment measure scores for some participants precluded a pretreatment-to-posttreatment analysis of satisfaction with sex life, the posttreatment means were also opposite to what would be expected if the comparison level, $F(1, 81) = .71$, were a generalized standard.

Discussion

Satisfaction Composite

The finding that pretreatment-to-posttreatment changes in judgments of satisfaction were reliably more positive in the hedonically negative than in the hedonically positive condition corroborates the major prediction derived from comparison level theory. Furthermore, it is interesting to note that as in Experiment 1 (bad old days condition), the hedonically negative condition was associated with the largest increments in evaluations. Although the stimulus material was not judged by participants in terms of hedonic value (as in Experiment 1), the greater potency of the hedonically negative material may have resulted from its departing more from participants' pretreatment comparison levels than did the hedonically positive information.

The findings for the satisfaction index do not appear to corroborate Byrne's (1971) affective model of evaluation, since judgments of satisfaction were greatest in the most affectively negative condition. Obviously, more research is needed to delimit the generally accepted positive relationship between affective states and evaluations.

Individual Items

The contrast hypothesis was corroborated for two—life and health—of the three items comprising the satisfaction composite. Evaluative judgments of physical appearance were virtually uninfluenced, although the within-cell experimental error for this item (see Footnote 3) was slightly smaller than the error associated with the items yielding reliable contrast effects. Furthermore, it will be recalled that satisfaction regarding three aspects of life besides the composite items were measured and analyzed, but contrast effects were not generally detected. Although comparison level theory may be of heuristic value, these findings do not appear consistent with the assumption that all outcomes are judged in terms of a single utility scale and standard—the comparison level.

General Discussion

Summary of Major Findings

These experiments corroborate the hypothesis that vicarious exposure to hedonic extremes—especially the hedonically negative—results in contrast effects regarding evaluative judgments of aspects of life that have evolved or been acquired in the course of life *beyond* the laboratory. Although the present results are generally consistent with comparison level theory, the findings appear better described by an adaptation level model of judgment (Upshaw, 1969, p. 347). Furthermore, the finding that the hedonically negative condition of Experiment 2 was associated with the most negative moods, but the most positive judgments, appears to contradict Byrne's (1971) affective model of evaluation.

Potential Limitations

Participant sex. Although only women participated in the present studies, the findings should hold for males, since contrast effects have been detected with males in other studies (e.g., Brickman, 1975).

Treatment complexity. The treatments used in the present experiments were complex. It is, therefore, not precisely known which aspects of the treatments—especially the hedonically negative treatments—mediated effects. Clearly, participants were not merely exposed to stimuli. For example, across both experiments, participants described reactions to the stimulus materials before evaluating aspects of life.

Magnitude of effects. In relation to the maximum contrast effects that could be detected in these experiments, the effects of the hedonically negative materials, though statistically reliable, were small. A number of considerations are relevant to this result:

1. Several participants indicated that the hedonically negative materials had created a conflict. On the one hand, they were inclined to increase the positivity of their judgments, yet they felt that this was inappropriate, since their judgments were typically not so positive. These participants reported expressing their more typical judgments on the post-

treatment measures, thus hindering the detection of large contrast effects.

2. In designing these experiments, it was assumed that most participants had not been extensively or recently exposed to hedonically negative information and that the hedonic value of such information consequently was substantially below participants' pretreatment comparison levels. These considerations would suggest the production of large contrast effects. According to Thibaut and Kelley (1959), however, the weights associated with the comparison level reflect the "salience" of outcomes, where "salience" refers to the extent that an individual might think about outcomes before judgment. But, as illustrated by the mood ratings in Experiment 2, hedonically negative outcomes may have immediately aversive properties. Thinking, for example, of injury, disease, or death generally elicits aversive emotional responses, amounts to self-punishment, and may therefore ordinarily be avoided.[6] Consequently, the salience and weighted contribution of the hedonically negative to the comparison level may be small.

3. Strictly speaking comparison and adaptation levels are weighted averages. The determinants of values and weights, however, have not been precisely specified. Although exposure to hedonically negative information for more than an hour, as in Experiment 2, may appear substantial, the contribution to comparison or adaptation levels may be small, since these constructs reflect all of a participant's previous experience.

Beyond the Hedonically Positive

Psychologists appear, in general, to have assumed that individuals are primarily oriented towards the hedonically positive when making assessments, although the psychological basis for this assumption has not always been specified. Festinger (1954) made this assumption for the assessment of abilities when he postulated that individuals compare themselves with those performing better. Numerous research instruments involve a comparison between a respondent's description of his or her present status and "ideal" status, and the concept of relative deprivation, of course, depends upon an individual not having that which he or she desires (see Cook, Crosby, & Hennigan, 1971). But if our theories are general, they should also explain when and how the hedonically negative may influence judgment. The young students we most often study may be bombarded with advertisements rendering the hedonically positive salient, and they may believe that they will always be healthy and their outcomes constantly improving; but as theorists we should not be blinded by these aspects of our culture.

Thibaut and Kelley (1959) attempted to specify a psychological theory of outcome salience in proposing that salient outcomes are those that an individual believes he or she can to some degree control. Since vicarious exposure to the hedonically positive may provide information about response-positive reinforcement contingencies (see Berger, 1977) and consequently mediate positive reinforcement, it is not surprising that individuals may attend to those achieving more positive outcomes and think about these activities and subsequent outcomes. Such behavior may momentarily decrease judgments of satisfaction with current outcomes but may mediate more positive future outcomes. Eventually, however, there may be little or no possibility of achieving the hedonically positive! This may be true, for example, because of injury, disease, old age, or social constraints. Under these circumstances, the hedonically negative may become salient even though it may

---

[6] Several months after Experiment 2 was completed, but before debriefing, we contacted participants by telephone and solicited their reactions to the experiment. The participants in the hedonically positive condition appear to have had some trouble recalling the experiment and generally did not express much interest in it. The participants in the hedonically negative condition, however, appeared to have little trouble recalling the study, thought participation was a valuable experience, and were eager to learn more about the study. Although thinking about the hedonically negative was aversive, the participants almost without exception deemed the experience valuable. Several participants reported being happier with their lives a few days after participation. One participant, however, reported discerning the major hypothesis the day following the experiment.

momentarily elicit aversive emotional behavior. The hedonically negative may provide information about response-negative reinforcement contingencies and thus facilitate an individual's avoiding outcomes far more aversive than those immediately resulting from vicarious exposure to, or thinking about, the hedonically negative. And, of course, as illustrated in the present experiments, such information would be expected to enhance judgments of satisfaction with present life outcomes.

Finally, it is interesting to note the applied implications of the present experiments. The students in Experiment 1 do not appear to have had much knowledge of the plight of the masses and the details of day-to-day life in earlier times. American historians only as recently as the 1960s have come to write history from the "bottom up" (see Thernstrom, 1964), although they were long aware that history has ordinarily been written and taught from the perspective of the bourgeoisie and aristocracy. The current findings suggest that it might be beneficial to incorporate descriptions of the lives of ordinary people into primary and secondary history curricula. Certainly 19th-century novelists have provided graphic descriptions. To paraphrase Sir Walter Raleigh, one may learn to be appreciative from comparing people's forepast miseries with one's own like errors and ill deservings. Similarly, the results of Experiment 2 suggest that we might more likely count our blessings if we were not so isolated from the hedonically negative.

References

Andrews, F. M., & Withey, S. B. *Social indicators of well-being*. New York: Plenum Press, 1976.

Berger, S. M. Social comparison, modeling, and perseverance. In J. M. Suls & R. L. Miller (Eds.), *Social comparison processes*. Washington, D.C.: Hemisphere, 1977.

Bettmann, O. L. *The good old days—They were terrible!* New York: Random House, 1974.

Brickman, P. Adaptation level determinants of satisfaction with equal and unequal outcome distributions in skill and chance situations. *Journal of Personality and Social Psychology*, 1975, *32*, 191–198.

Brickman, P., & Campbell, D. T. Hedonic relativism and planning the good society. In M. H. Appley (Ed.), *Adaptation-level theory*. New York: Academic Press, 1971.

Byrne, D. *The attraction paradigm*. New York: Academic Press, 1971.

Cook, T. D., Crosby, F., & Hennigan, K. M. The construct validity of relative deprivation. In J. M. Suls & R. L. Miller (Eds.), *Social comparison processes*. Washington, D.C.: Hemisphere, 1977.

Eiser, J., & Stroebe, W. *Categorization and social judgment*. New York: Academic Press, 1972.

Festinger, L. A theory of social comparison processes. *Human Relations*, 1954, *7*, 117–140.

Fisher, M. L. *When tragedy may lead to happiness: Extending comparison level theory into the realm of life satisfaction*. Unpublished master's thesis, University of Wisconsin—Milwaukee, 1976.

Friedland, N., Arnold, S. E., & Thibaut, J. Motivational bases in mixed-motive interactions: The effects of comparison levels. *Journal of Experimental Social Psychology*, 1974, *10*, 188–199.

Helson, H. *Adaptation-level theory*. New York: Harper & Row, 1964.

Manis, M. Context effects in communication. In M. H. Appley (Ed.), *Adaptation-level theory*. New York: Academic Press, 1971.

Maslow, B. G. (Ed.). *Abraham H. Maslow: A memorial volume*. Monterey, Calif.: Brooks/Cole, 1972.

Nowlis, V. Mood: Behavior and experience. In M. B. Arnold (Ed.), *Feelings and emotions*. New York: Academic Press, 1970.

Pepitone, A., & DiNubile, M. Contrast effects in judgments of crime severity and the punishment of criminal violators. *Journal of Personality and Social Psychology*, 1976, *33*, 448–459.

Thernstrom, S. *Poverty and progress*. Cambridge, Mass.: Harvard University Press, 1964.

Thibaut, J. W., & Kelley, H. H. *The social psychology of groups*. New York: Wiley, 1959.

Upshaw, H. S. The personal reference scale: An approach to social judgment. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 4). New York: Academic Press, 1969.

Volkmann, J. Scales of judgment and their implications for social psychology. In J. H. Rohrer & M. Sherif (Eds.), *Social psychology at the crossroads*. New York: Harper, 1951.

Received February 24, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 2, 261–272

Actor–Observer Differences in Information Inference and Causal Attribution

Susan V. Eisen
McLean Hospital
Belmont, Massachusetts

Two experiments were conducted to assess (a) differences in the information available to persons trying to understand the causes of their own behavior (actors) versus those trying to understand the causes of another's behavior (observers) and (b) the effects of information differences on causal explanations. In Experiment 1, actors reported positive behaviors to be less distinctive and more consistent with past behavior than did observers, whereas the reverse was true for negative behaviors. Consistent with this difference, actors attributed desirable behaviors more to their own internal dispositions than did observers, whereas the opposite occurred for undesirable behaviors. In Experiment 2, when all subjects were given the consensus, distinctiveness, and consistency information generated by actors in Study 1, both actors and observers attributed positive acts more to internal factors than negative acts. When given the information generated by the observers, neither actors nor observers exhibited this bias. Thus, when given the same information, actors and observers no longer showed differences in causal explanations.

Recent theory and research have suggested that under a variety of circumstances, causal explanations for one's own behavior differ from explanations of the same behavior performed by someone else (Jones & Nisbett, 1972; Monson & Snyder, 1977). Jones and Nisbett (1972) discussed three factors that may contribute to these differential perceptions of causality. They proposed that actors and observers differ in their visual perspectives, their motivations, and the information available to them.

The effects of visual perspective have received much attention, and results have generally supported the contention that actors tend to see the causes of their behavior as environmental, whereas observers view the causes of the same behavior as originating within the actor (Arkin & Duval, 1975; Nisbett, Caputo, Legant, & Maracek, 1973; Regan & Totten, 1975; Storms, 1973). Similarly, much research has focused on actor–observer attribution differences that appear to be motivationally based (e.g., Arkin, Gleason, & Johnston, 1976; Bradley, 1978; Miller & Ross, 1975; Sicoly & Ross, 1977; Snyder, Stephan, & Rosenfield, 1976; Stevens & Jones, 1976).

In contrast, the question of the effects of actors' and observers' informational differences on causal attribution has been relatively neglected. It is this question that the present research addresses. Jones & Nisbett (1972) suggested that because they know their own history, actors are in a better position than observers to assess how distinctive and consistent a current behavior is compared with past behavior. Although observers may know little about someone else's past behavior, they probably have an idea of whether they them-

This article is based on a dissertation submitted to the Department of Psychology, Brandeis University, in partial fulfillment of the requirements for the PhD degree. The author would like to thank Leslie McArthur for her advice and support as chairperson of the thesis committee and for extensive comments on earlier versions of this manuscript. Thanks also go to Teresa Amabile for her careful reading and helpful comments.

Requests for reprints should be sent to Susan V. Eisen, McLean Hospital, 115 Mill Street, Belmont, Massachusetts 02178.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3702-0261$00.75

selves would (or did) perform the behavior (consensus). Thus, observers may evaluate an actor's behavior by comparing it with the behavior of others, while actors may evaluate their own behavior by comparing it with past behavior. These three types of information (distinctiveness, consistency, and consensus) were first proposed as useful for causal attribution by Kelley (1967) and were found to have a significant impact on perceptions of causality (Hansen & Lowe, 1976; McArthur, 1972; Orvis, Cunningham, & Kelley, 1975).

The difference in information available to actors and observers may contribute to divergent perceptions of causality (Jones & Nisbett, 1972). For example, if Mr. Smith fails to show up for a meeting, his supervisor may infer that Smith's unreliable behavior is widespread (low in distinctiveness and high in consistency) and that hardly anyone else fails to attend (low consensus). In line with Kelley's (1967) model, the supervisor would attribute Smith's transgression to his unreliable nature. On the other hand, Smith may know that he rarely misses meetings and attribute his absence to external factors. Consistent with the difference in information available to actors and observers, Hansen and Lowe (1976) found that in attributing causality for a behavior, actors relied more heavily on distinctiveness and consistency information, whereas observers favored consensus. However, to date there have been no attempts to ascertain whether actors think their behavior is more distinctive and less consistent with past behavior than observers think it is.

The purposes of the present research were (a) to investigate actors' and observers' reports of how they think others would behave in given situations (consensus) and their knowledge or assumptions about the distinctiveness and consistency of given behaviors; (b) to assess actors' and observers causal attributions for a variety of behaviors; and (c) to examine the effect on causal attributions of providing actors with the information reported by observers, and vice versa.

Assessment of actor–observer differences in information assumptions was accomplished in Experiment 1 by having actors and observers estimate consensus, distinctiveness, and consistency information for a series of behaviors. Causal attributions were studied in Experiment 2 by giving one group of subjects no information, another group the information generated (assumed) by actors in Study 1, and a third group, the information generated by observers in Study 1. This method is analogous to that used by Storms (1973). But, rather than reversing visual perspectives, actors' and observers' "informational perspectives" were reversed.

Experiment 1

Experiment 1 was designed to assess differences in actors' and observers' estimates of consensus, distinctiveness, and consistency surrounding each of 12 behaviors. For this purpose, 12 "personality–achievement inventory" items were devised, each indicative of a different behavior. Three valences (positive, neutral, and negative) and four categories (emotions, beliefs, actions, and accomplishments) of behavior were employed, yielding one item for each combination of valence and category. All subjects completed the items and were provided with false feedback informing them of the behavioral dimension each item measured. Actors' feedback referred to their own responses, while observers' feedback referred to the responses of another subject whose completed items they were shown. Based on the feedback, actors generated consensus, distinctiveness, and consistency for each of their own responses, while observers generated the same types of information for their matched actor's responses.

Choice of this methodology was guided by two major considerations: (a) In order for me to be able to generalize the findings, investigation of a variety of behaviors was desired, and (b) relatively novel behaviors, for which actors are not likely to have a vast store of distinctiveness and consistency information, had to be chosen in order for the information manipulations used in Study 2 to be believable. Since taking a "personality–achievement inventory" is a fairly novel experience in the lives of most people, it appeared to be a good mechanism for accommodating these considerations.

Table 1
Personality and Achievement Dimensions Represented by Each Item

Valence	Behavioral category			
	Emotions	Beliefs	Actions	Accomplishments
Positive	*Happy*/ Unhappy	*Progressive*/ Conservative	*Helpful*/ Unhelpful	Verbal reasoning *Correct*/incorrect
Neutral	*Excitable*/ Unexcitable	Skeptical/ *Naive*	Reserved/ *Outspoken*	Language skills *50%*
Negative	Affable/ *Hostile*	Tolerant/ *Intolerant*	Reliable/ *Unreliable*	Spatial relations Correct/ *incorrect*

Note. The italicized item represents the feedback given to subjects regarding their own (actors) or their matched actor's (observers) response.

Based on Jones and Nisbett's (1972) suggestion, it was predicted that (a) actors would estimate higher distinctiveness and lower consistency for their responses than would observers and (b) actors would be more certain of their distinctiveness and consistency estimates than would observers.

The predictions for consensus information were more complicated. Since observers as well as actors completed the inventory items, they knew whether they gave the same response as the person whose completed items they were shown. Since only 12 items were employed, observers could easily recall their responses. Thus, they already had "self-based" consensus information (Hansen & Donoghue, 1977). Consequently, it was predicted that (c) observers who gave the same response on an item as their matched actor (high self-based consensus) would estimate higher consensus than would actors, whereas observers who gave a different response (low self-based consensus) would estimate lower consensus than would actors; and (d) observers having high self-based consensus would be more certain of their consensus estimates than actors.

Method

Experimental Design

The experiment was a 2 × 2 × 3 × 4 design (Sex of Subject [male, female] × Role [actor, observer] × Valence of Behavior [positive, neutral, negative] × Category of Behavior [emotions, beliefs, actions, accomplishments]).

Subjects

Eighteen male and 18 female undergraduates volunteered to serve as subjects for research on "personality assessment," in partial fulfillment of an introductory psychology course requirement. Half the subjects of each sex were randomly assigned to serve as actors; the remaining subjects served as observers and were each "yoked" to a same-sex actor. That is, each observer was shown the responses of a same-sex actor and generated information for that actor's responses. Subjects were run individually by a female experimenter.

Personality–Achievement Items

Twelve forced-choice self-report items from a supposed "personality–achievement inventory" were constructed. Each item indicated a positive, neutral, or negative emotion, belief, action, or accomplishment. The nine personality behavior dimensions (emotions, beliefs, and actions) were selected from Anderson's (1968) likableness ratings of 555 adjectives. Positive behaviors were rated in the top 100, neutral behaviors between 240 and 300, and negative behaviors in the bottom 100. The behavioral dimensions selected appear in Table 1. Although the dimensions could be conceived of as traits, the items themselves dealt with specific, concrete situations. For example, the following item represented the happy–unhappy (positive emotion) personality dimension:

A. When I feel depressed, my friends are usually able to cheer me up.
B. When I'm dating someone special, I feel more cheerful than usual.

The three achievement items were actual tasks: an analogy, an anagram, and an embedded-figure problem.

Three randomly ordered blocks of four items each were constructed, with the restriction that an achieve-

ment task was always the last item in each block. This was done to facilitate timing of the problems. (The personality items were not timed.) Each of the three blocks was randomly ordered to yield three orders of items.

Procedure

Subjects were told that we were studying personality assessment techniques and that they would be asked to complete 12 items from a larger, widely used personality–achievement inventory. The items would be scored, the subjects would be told the results, and they would be asked some questions about their responses. Participants in the experiment were then asked to read the instructions, complete the first page containing three personality items, look at the example of the problem that was to follow on the next page, and then call the experimenter to time the "test" problem. The experimenter informed subjects how much time they would have to do the problem, instructed them to begin, and informed them when time was up. This procedure was followed until all 12 items were completed.

Actors. After completing the items, actors were left alone for approximately 3 minutes while the experimenter scored their inventories. Upon returning, the experimenter gave the subjects the scored items, along with written feedback and the dependent-measure questionnaire.

Observers. Upon completion of the inventory items, observers were told that their tests would be scored and feedback would be given later, but in the meantime, the experimenter would like them to look at someone else's completed and scored inventory and answer some questions about the other person's responses. The experimenter left the room, returning several minutes later with a same-sex actor's scored inventory and the dependent-measure questionnaire.

Feedback

The feedback (scoring) was presented in the dependent-measure booklet. At the top of each page appeared a personality or achievement dimension represented by one item, followed by a sentence specifying which end of the dimension the subject's response indicated. The dimensions were numbered to correspond to the numbers of the inventory items. Below is an example of one dimension and the feedback given:

> Subscale: HAPPY–UNHAPPY
> On this item you (actors)/this student (observers) gave the *happy*/unhappy response.

For each item, the end of the dimension that the response supposedly indicated (the italicized word) was circled in red ink to appear as if the feedback were tailored to the individual's response. In fact, everyone received the same feedback regardless of how the item was answered. Thus, even though subjects did not answer all the items in the same way, all were told that their responses indicated the same behavior. Debriefing revealed that subjects believed that the items were from an actual personality-achievement inventory and that the feedback was valid. The experimenter went through a sample item explaining how the items were scored, ensuring that subjects understood the instructions.

Dependent Measures

Based on the feedback, subjects were asked to estimate for each item (a) what percentage of the other subjects in the experiment also gave the *X* response on that item (consensus information); (b) if they took the whole inventory (rather than just the 12 items selected), on what percentage of the items in that subscale they (or their matched actor) would also give the *X* response (distinctiveness information); and (c) if they (or their matched actor) took the inventory at several different times, what percentage of the time they (or their matched actor) would give the *X* response on that item (consistency information).

Inferences were made on a 9-point scale from 10% to 90% of the people, items, and time for consensus, distinctiveness, and consistency information, respectively. Subjects also rated on a 9-point scale how certain they were about each of their information estimates.

Results

Data Analysis

Information estimates and certainty data were analyzed using a 2 × 2 × 3 × 4 (Sex of Subject × Role × Valence × Category) analysis of variance. Sex of subject was a between-subjects factor; the remaining three independent variables were within-subjects factors. Since actors and observers were yoked, role was considered to be a within-subjects factor, and actor–observer pairs were the units of analysis.

Since more than one dependent measure was obtained, use of multivariate analysis of variance was considered. However, because of the difficulty in interpreting multivariate analysis in which there are repeated measures on more than one factor and multiple levels of the independent variables, univariate analyses were done. An attempt was made to guard against a Type I error by adopting the more stringent Greenhouse–Geisser criterion for rejection of the null hypothesis (Winer, 1971).

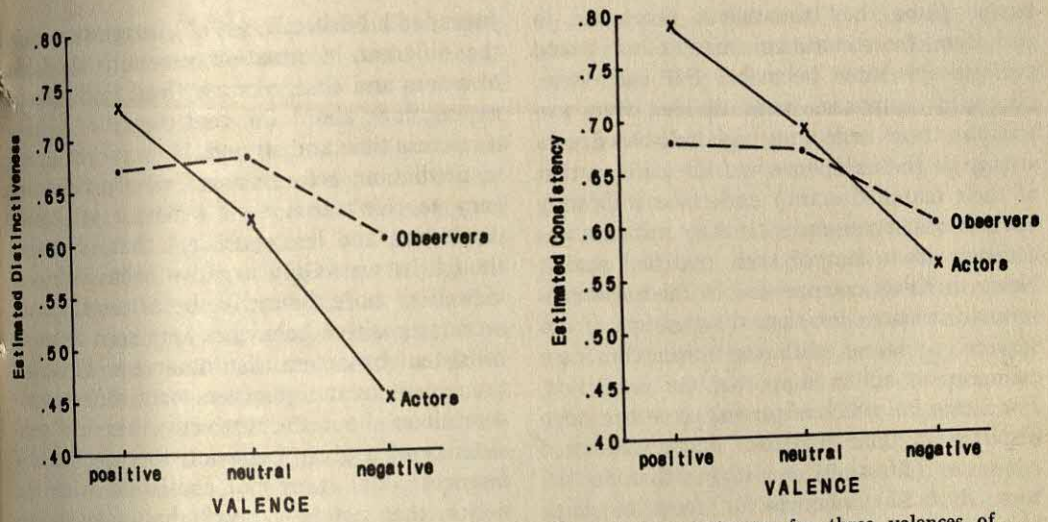


Figure 1. Actors' and observers' distinctiveness and consistency estimates for three valences of behavior. (The higher the number, the lower the estimated distinctiveness.)

All effects reported are either still significant according to this stringent criterion, or if not, an F_{max} test for homogeneity of variance was performed and found to be nonsignificant. Significant three-way or higher order interactions that are not psychologically meaningful are not reported.

Information Estimates

Distinctiveness information. As predicted, actors estimated their responses to be significantly higher in distinctiveness than observers did, $F(1, 16) = 5.48$, $p < .05$. However, a significant Role × Valence interaction, $F(2, 32) = 14.45$, $p < .001$, modified this main effect, indicating that only for negative behaviors did actors infer higher distinctiveness than did observers ($p < .05$; see Figure 1). Actors generated significantly higher distinctiveness for negative than neutral behaviors and significantly lower distinctiveness for positive than neutral behaviors ($ps < .05$). Observers estimated higher distinctiveness for negative than neutral behaviors ($p < .05$), but there was no significant difference in their estimates for positive and neutral behaviors.

A significant main effect of valence was also obtained, $F(2, 32) = 37.34$, $p < .001$. The more positive the response, the lower the estimated distinctiveness.

Consistency information. Contrary to prediction, the main effect of role on consistency estimates did not approach significance. Both actors and observers estimated that consistency of behavior (how often they would give the same response on the same item) would be fairly high (see Figure 1). However, a significant Role × Valence interaction was obtained, $F(2, 32) = 6.71$, $p < .001$. Consistent with the findings for distinctiveness information, actors inferred lower consistency than did observers only for undesirable behaviors, although the difference was not significant. Contrary to prediction, actors estimated significantly higher consistency than did observers for desirable behaviors ($p < .05$). Actors' consistency estimates for positive, neutral, and negative behaviors all differed significantly from each other ($ps < .05$). Observers generated lower consistency for negative than neutral responses ($p < .05$), but there was no difference in their estimates for positive versus neutral responses.

As for distinctiveness inference, a highly significant valence main effect was also found, $F(2, 32) = 32.30$, $p < .001$. Both actors and observers generated highest consistency for positive behaviors and lowest consistency for negative behaviors.

Consensus information. As predicted, observers used their knowledge of their own be-

havior (how they themselves responded to each item) to estimate how many others would perform the same behavior. For each item, observers' responses were divided into two groups—those reflecting high self-based consensus (if their response was the same as that of their matched actor) and those indicating low self-based consensus (if their response was different from that of their matched actor). Newman-Keuls comparisons of the mean consensus estimates for these two groups of observers' responses with the mean consensus estimates of actors supported the prediction. For items on which observers gave the same response as their matched actors, generated consensus ($M = 69$) was higher than for actors ($M = 57$), whereas for items on which observers differed, generated consensus ($M = 52$) was lower than for actors. Each of the means differed significantly from the others ($ps < .01$).

A significant main effect of valence was also found, $F(2, 32) = 12.18$, $p < .01$. The more positive the behavior, the higher the estimated consensus ($M = 65$ for positive, $M = 59$ for neutral, $M = 54$ for negative behaviors). Again, each mean differed significantly from each of the others ($ps < .05$).[1]

Certainty

As predicted, observers having high self-based consensus were significantly more certain than actors of their consensus estimates, $t(18) = 2.27$, $p < .05$. Observers having low self-based consensus were no less certain than actors, $t(18) < 1$. Also as expected, actors were more certain of their distinctiveness and consistency estimates than were observers, $F(1, 16) = 6.20$, $p < .025$, for distinctiveness, and $F(1, 16) = 7.12$, $p < .025$, for consistency.

Significant main effects of valence were also obtained on certainty for each type of information ($ps < .01$). The more positive the behavior, the more certain subjects were of their information estimates.

Discussion

The actor–observer differences in distinctiveness and consistency estimates support Jones and Nisbett's (1972) suggestion that the different information generally available to actors and observers may lead to divergent assumptions about the consistency of behavior across time and stimuli. However, contrary to prediction, actors showed no *general* tendency to think that their behavior was more distinctive and less consistent than observers thought it was. Only negative behaviors were viewed as more distinctive by actors than observers; positive behaviors were seen as more consistent by actors than observers. Thus, although actors and observers made different assumptions about the distinctiveness and consistency of a given behavior, the specific differences found were not consistent with the notion that actors "know" that a given behavior is more distinctive and/or less consistent than an observer thinks it is. Rather, the differences in information inference were primarily determined by desirability of the behavior.

Focusing on causal attributions for positive and negative behaviors, Miller and Ross (1975) and Monson and Snyder (1977) have suggested that actor–observer differences in attribution may derive from informational differences. Actors may attribute "good" behaviors more to themselves because they selectively remember past instances of desirable behavior (Monson & Snyder, 1977). Such selective recall would allow for an actor's inference of high consistency and low distinctiveness for positive acts, and low consistency and high distinctiveness for negative acts. Results of the present study support this notion. Actors generated higher distinctiveness than observers for negative responses and higher consistency than observers for positive responses.

However, the question that comes to mind is, What could be the basis for actors' selective recall? There is no compelling cognitive

---

[1] Significant Valence × Category interactions were also obtained for consensus, distinctiveness, and consistency estimates. They did not appreciably change the implication of the valence main effects, but did suggest that the strength of the valence effect varies somewhat as a function of type of behavior. Similar Valence × Category interactions were also obtained on measures of causal attribution in Experiment 2.

reason for actors to recall desirable behaviors more than undesirable ones. It could be suggested that positive behaviors are perceived as more common by actors than by observers. If this were the case, actors' distinctiveness and consistency estimates might be logical inferences based on perceived frequency of occurrence of the behaviors. (If people are generally helpful, and I give a helpful response, than it would be logical to think that I would usually give a helpful response.) Low distinctiveness and high consistency would be expected for frequently occurring behaviors. The valence main effect for consensus estimates suggests that positive behaviors were indeed seen as most common (highest generated consensus), followed by neutral and negative behaviors, respectively. However, actors did not see positive behaviors as more common than observers did; nor did they see negative acts as more unusual. Thus, the relative frequency of positive behaviors and uniqueness of negative behaviors does not provide an adequate cognitive explanation for the actor–observer differences in distinctiveness and consistency inference.

The failure to find a compelling cognitive explanation for the results suggests that actors' information estimates may be motivationally biased. Sicoly and Ross (1977) recently found evidence for self-serving biases in attribution that could not be explained in nonmotivational terms. The present results suggest that such attribution biases may be traceable to biased estimates of the distinctiveness and consistency of behavior, which are influenced by actors' motivations.

Experiment 2

The purpose of the second experiment was to assess actor–observer differences in causal attribution and to determine whether the discrepancies in information inference found in Experiment 1 influence attributions. The design was identical to that of Experiment 1, with the additional between-subjects factor of information. One group of actors and observers was given the consensus, distinctiveness, and consistency information estimated by actors in Experiment 1. Another group was given the information estimated by observers in Experiment 1. A third group was not given any information. All subjects made causal attributions for their own or their matched actor's behavior. The following predictions were made:

1. Actors given no information would attribute positive behaviors more to internal causes than would observers, and negative behaviors more to external causes than observers would. This prediction was based on the Role × Valence interactions on distinctiveness and consistency estimates found in Experiment 1.

2. Since actors' and observers' informational inferences for neutral behaviors did not differ in Experiment 1, it was predicted that actors' and observers' causal attributions for neutral behaviors would not differ, given no information; nor were they expected to differ in the two information conditions.

3. When provided with identical information, actors' and observers' causal attributions were not expected to differ. Thus, no Role × Valence interactions were expected to occur in the two information conditions. Specifically, (a) actors and observers presented with the information estimated by actors in Experiment 1 would exhibit the strong positivity bias of actors given no information (more internal attribution for positive than negative responses and more external attribution for negative than positive responses); (b) actors and observers given the information generated by observers in Experiment 1 would exhibit the attributional tendencies of observers given no information.

Method

Subjects

Fifty-four male and 54 female undergraduates served as subjects in partial fulfillment of an introductory psychology course requirement. Twenty-seven subjects of each sex were randomly assigned to serve as actors and 27 as observers.

Procedure

The materials and procedure were the same as those used in Experiment 1.

Information Manipulations

Subjects in the information conditions were told that in addition to the feedback regarding their responses, they would be given information dealing with the norms (consensus), validity (distinctiveness), and reliability (consistency) of the inventory items. The information was presented for each item along with the feedback in the following form:

> *Norms*. *X*% of the subjects in this experiment also gave the *Y* response.
> *Validity*. If you/this student took the whole inventory (rather than just these 12 items), you/he/she would give the *Y* response on *X*% of the items in this subscale.
> *Reliability*. If you/this student took the inventory at different times, you/he/she would give the same response on this item *X*% of the time.

Subjects were told that the information was based on past research with the inventory and that they might find it useful in helping them to decide what caused them (or their matched actor) to give a particular response. The actual percentages provided were those generated by subjects in Experiment 1. One actor and one observer were each given the exact consensus, distinctiveness, and consistency estimates generated by one subject in the first experiment. Thus, 18 actors and 18 observers got the information generated by the 18 actors in Experiment 1; 18 actors and 18 observers received the information generated by the 18 observers in Experiment 1. An additional 18 actors and 18 observers received no information.

Dependent Measures

Based on the feedback from the personality–achievement items, subjects were asked to decide what was the most probable cause of each response. The four alternatives provided by McArthur (1972) were presented, as follows:

> 1. Something about me (actors)/this student (observers) probably caused me (actors)/him or her (observers) to give the *Y* response.
> 2. Something about the specific item probably caused me/him/her to give the *Y* response.
> 3. Something about the particular circumstance probably caused me/him/her to give the *Y* response.
> 4. Some combination of 1, 2, and 3 above probably caused me/him/her to give the *Y* response.

Subjects were asked, if they chose Number 4, to specify the particular combination of factors they thought could best explain why they (or their matched actor) gave the *Y* response.

Results

Data Analysis

The design of this experiment was identical to that of Experiment 1, with the additional between-subjects factor of information. Three levels of information were varied: (a) information generated by actors in Experiment 1; (b) no information; and (c) information generated by observers in Experiment 1.

A 3 × 5 (Item Order × Locus of Attribution) chi-square performed on the 1,296 forced-choice attributions was not significant. Therefore, all further analyses were done without regard to item order. Analyses of variance were performed on internal (person) and external (stimulus and circumstance attribution.[2] (The measure of external attribution was derived from summing stimulus, circumstance, and stimulus–circumstance attributions.)

Causal Attribution Effects

Internal attribution. It was predicted that actors given no information would exhibit a "positivity" bias in causal attribution for positive and negative behaviors that would be reduced or eliminated when they were provided with the information generated by observers. Observers' attributions were not expected to reflect as strong (if any) a bias, given no information, but were expected to be biased when given the information generated by actors. To test these predictions, the simple interactions were computed for each level of information. The results supported the predictions. When subjects were given no information, a significant Role × Valence interaction, $F(2, 96) = 10.09$, $p < .001$, revealed that actors made more internal attributions for positive responses and fewer internal attributions for negative responses than observers did, although only the latter difference was significant ($p < .05$). Newman-Keuls tests revealed that actors made significantly fewer internal attributions for negative than for neutral or positive responses ($ps < .05$), whereas ob-

[2] Analyses done on person–stimulus and other internal–external combination attributions are not reported because they are not directly relevant to the hypotheses.

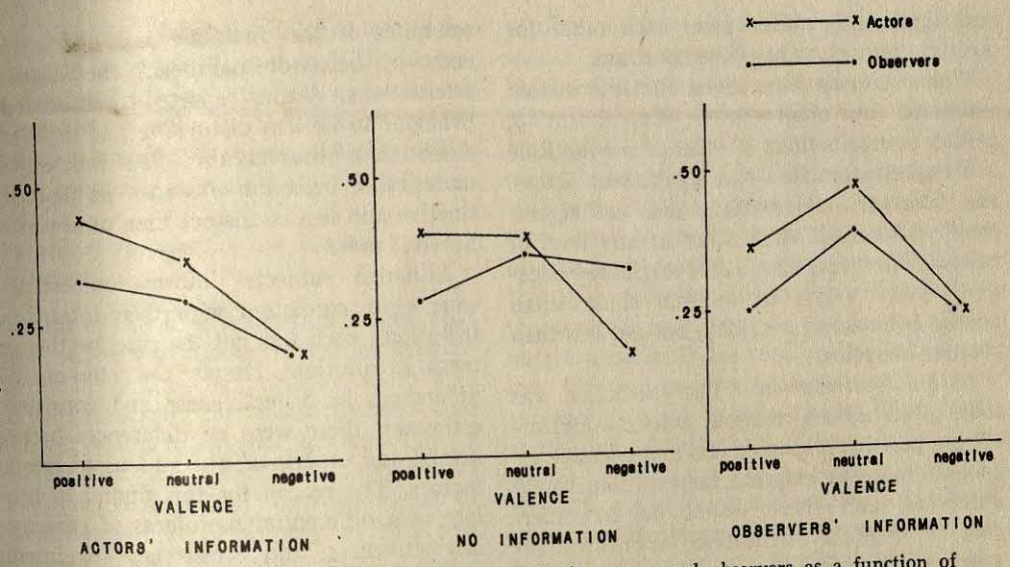


Figure 2. The proportion of internal attributions made by actors and observers as a function of information and behavioral valence.

ervers' internal attributions for positive, neural, and negative responses did not significantly differ from each other (see Figure 2). As predicted, there was no actor–observer difference in internal attribution for neutral responses.

When subjects were given the information generated by the actors in Experiment 1, a significant main effect of valence was obtained, $F(2, 96) = 21.06$, $p < .0$[illegible], but the Role × Valence interaction did not approach significance. As shown in Figure 2, actors' attributions reflected the same pattern as when given no information; the pattern for observers was similar to that of actors. Both groups exhibited the "positivity" bias shown only by actors in the no-information condition. Actors' and observers' attributions did

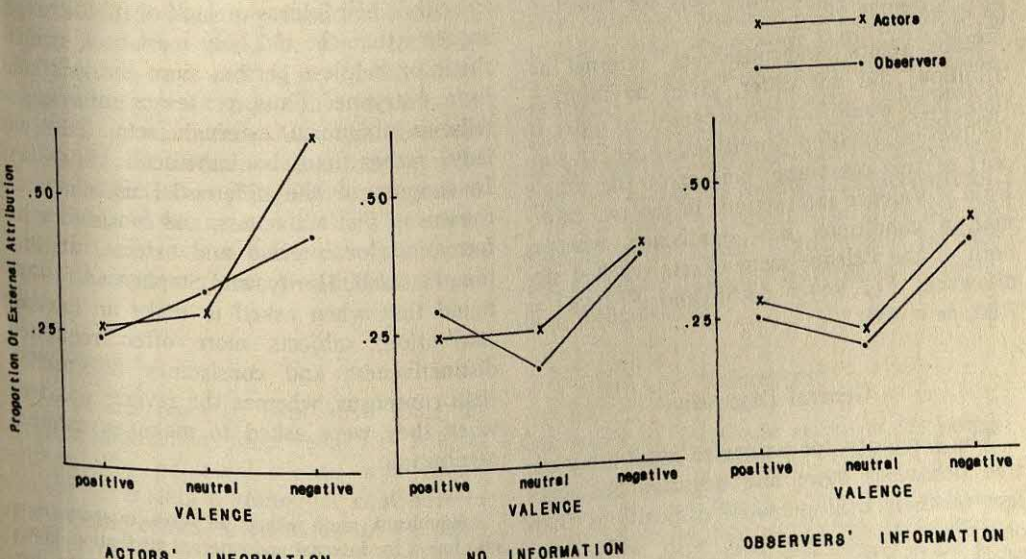


Figure 3. The proportion of external attributions made by actors and observers as a function of information and behavioral valence.

not significantly differ from each other for positive, neutral, or negative behaviors.

When subjects were given the information generated by observers in Experiment 1, neither the main effect of valence nor the Role × Valence interaction was significant. Actors' and observers' attributions did not significantly differ from each other at any level of valence. However, actors still attributed negative behaviors less to internal causes than neutral behaviors ($p < .05$), but not less than positive behaviors.

External attribution. The prediction was that, given no information, actors would attribute positive responses less and negative responses more to external factors than would observers, and there would be no difference in external attribution for neutral responses. As predicted, actors' and observers' external attribution for neutral behaviors did not differ. However, contrary to prediction, there were no actor–observer differences for positive or negative responses either (see Figure 3). The simple Role × Valence interaction for the no-information condition was not significant. However, the simple main effect of valence was significant, $F(2, 96) = 7.21$, $p < .001$, indicating that both actors and observers attributed negative responses more externally than neutral responses, although the difference was significantly only for observers ($p < .05$).

Since actors' and observers' external attributions did not differ, given no information, they would not be expected to differ in the two information conditions either. In support of this reasoning, neither of the simple Role × Valence interactions in the two information conditions was significant, whereas both of the valence main effects reached significance, $F(2, 96) = 33.65$ and $F(2, 96) = 7.80$, ps $< .001$.[3]

General Discussion

The tendency for actors to attribute positive behaviors more and negative behaviors less to their own personal dispositions than observers do is consistent with self-serving biases found in previous research (Bradley, 1978). Furthermore, actors' and observers' estimates of the distinctiveness and consistency of behavior paralleled the disparate attributions. Actors reported their desirable behavior to be less distinctive and more consistent than observers thought it was, whereas undesirable behavior was seen as more distinctive and less consistent than observers believed it to be.

Although subjects' information estimates were quite consistent with their internal attributions, such was not the case for their external attributions. Despite the actor–observer differences in disinctiveness and consistency estimates, there were no differences between actors' and observers' attributions to external factors. The reason for this finding may relate to a differential usefulness of consensus, distinctiveness, and consistency for internal versus external attribution. It has been suggested that distinctiveness and consistency information are more relevant for internal than external attribution because they provide clues as to what sort of person someone is (Anderson, 1974; McArthur, 1976). If we know that Priscilla usually helps the old lady (high consistency) and that she helps out other people, too (low distinctiveness), we can infer that she is a helpful person. In contrast, knowing that other people help the old lady (high consensus) has no direct implications about Priscilla's helpfulness or lack of it. Rather, it suggests that the old lady must be a sympathetic or helpless person, since she gets help from everyone. Thus, consensus information tells us about an external factor (the old lady) rather than about the actor (Priscilla). In support of the differential usefulness of consensus, distinctiveness, and consistency information for internal and external attribution, Garland, Hardy, and Stephenson (1975) found that when asked to make an internal attribution, subjects more often requested distinctiveness and consistency information than consensus, whereas the reverse was true when they were asked to make an external attribution.

---

[3] Significant main effects of verb category were also found for internal and external attribution. However, since no actor–observer differences were predicted or found, the findings pertaining to verb category are not reported.

If consensus information is most useful for making an external attribution, then actors' and observers' external attributions would not be expected to differ in the present experiment, since actors and observers did not differ in their consensus estimates.

Influence of Information on Causal Attribution

Results of the present research suggest that actors' self-serving biases in causal attribution can be reduced by providing them with information about the distinctiveness and consistency of their behavior, generated by an observer. Such information may be viewed as more objective than their own knowledge of distinctiveness and consistency of their behavior, which, like their attributions, appears to reflect self-serving biases. Similarly, observers can be swayed by actors' biased information estimates to attribute positive behavior more and negative behavior less to an actor's personal disposition.

It has been suggested that self-serving biases reflect strategic self-presentation by actors, rather than their private beliefs (Bradley, 1978). Attempts should be made to minimize the likelihood that such an artifact may be responsible for experimental findings. In the present research, subjects were told that their responses to the inventory items and the dependent measures would be kept confidential. No names appeared on any of the forms, and only the experimenter could identify the completed questionnaires. Although observers were shown the inventory-item responses of another subject, that person was not present at the time and was never identified by name. Thus, the influence of self-presentation motives in the present research should have been minimal.

Implications

The results reported here have some interesting implications for a variety of natural settings. The perennial dieter, who claims that he never eats a thing yet continues to gain weight, comes to mind as an example. He may attribute his obesity to an abnormally slow metabolism or some other factor that is out of his control. In accord with his attribution, he reports that he hardly ever snacks between meals—that is, snacking is highly distinctive and inconsistent. On the other hand, the dieter's wife, who attributes her husband's obesity to his eating habits, correctly reports the late afternoon chocolate bars and midnight refrigerator raids. If the dieter were to keep a running record of every bite he took during the day, he would have the more objective information reported by the observer and might be convinced that his weight problem is the result of his excess eating. This self-monitoring technique has been successfully used to help dieters in correctly attributing causality and, ultimately, in losing weight (Mahoney, 1974). Similar methods might be useful in other therapeutic or educational settings where taking responsibility for negative behaviors is a prerequisite for positive change.

On the other hand, there are also situations in which the actor's information may be more accurate or complete than the observer's. A student may attribute his or her lack of attendance in Psychology 101 to the fact that the lectures are boring and uninformative. Consistent with this attribution, that student may know that his or her attendance at other classes is and always has been excellent. The professor may attribute the student's behavior to an unreliable nature, based on consistent absence throughout the semester, regardless of what topics were being covered. In this situation, the actor (student) has more consistency and distinctiveness information than the observer. If the professor learned that the student's behavior in Psychology 101 really was highly distinctive and inconsistent, he or she might be expected to alter the impression of the absent student and perhaps even to realize that the lectures leave something to be desired.

References

Anderson, N. H. Likableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology*, 1968, *9*, 272–279.

Anderson, N. H. Cognitive algebra: Integration theory applied to social attribution. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 7). New York: Academic Press, 1974.

Arkin, R. M., & Duval, S. Focus of attention and causal attributions of actors and observers. *Journal of Experimental Social Psychology,* 1975, *11,* 427–439.

Arkin, R. M., Gleason, J. M., & Johnston, S. Effects of perceived choice, expected outcome, and observed outcome of an action on the causal attributions of actors. *Journal of Experimental Social Psychology,* 1976, *12,* 151–158.

Bradley, G. W. Self-serving biases in the attribution process: A reexamination of the fact or fiction question. *Journal of Personality and Social Psychology,* 1978, *36,* 56–72.

Garland, H., Hardy, A., & Stephenson, L. Information search as affected by attribution type and response category. *Personality and Social Psychology Bulletin,* 1975, *1,* 612–615.

Hansen, R. D., & Donoghue, J. M. The power of consensus: Information derived from one's own and others' behavior. *Journal of Personality and Social Psychology,* 1977, *35,* 294–303.

Hansen, R. D., & Lowe, C. A. Distinctiveness and consensus: The influence of behavioral information on actors' and observers' attributions. *Journal of Personality and Social Psychology,* 1976, *34,* 425–434.

Jones, E. E., & Nisbett, R. E. *The actor and the observer: Divergent perceptions of the causes of behavior.* New York: General Learning Press, 1972.

Kelley, H. H. Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 15). Lincoln: University of Nebraska Press, 1967.

Mahoney, M. J. Self-reward and self-monitoring techniques for weight control. *Behavior Therapy,* 1974, *5,* 48–57.

McArthur, L. A. The how and what of why: Some determinants and consequences of causal attribution. *Journal of Personality and Social Psychology,* 1972, *22,* 171–193.

McArthur, L. Z. The lesser influence of consensus than distinctiveness information on causal attributions: A test of the person–thing hypothesis. *Journal of Personality and Social Psychology,* 1976, *33,* 733–743.

Miller, D. T., & Ross, M. Self-serving biases in the attribution of causality: Fact or fiction? *Psychological Bulletin,* 1975, *82,* 213–225.

Monson, T. C., & Snyder, M. Actors, observers, and the attribution process: Toward a reconceptualization. *Journal of Eperimental Social Psychology,* 1977, *13,* 89–111.

Nisbett, R. E., Caputo, C., Legant, P., & Maracek, J. Behavior as seen by the actor and as seen by the observer. *Journal of Personality and Social Psychology,* 1973, *27,* 154–164.

Orvis, B. R., Cunningham, J. D., & Kelley, H. H. A closer examination of causal inference: The roles of consensus, distinctiveness, and consistency information. *Journal of Personality and Social Psychology,* 1975, *32,* 605–617.

Regan, D. T., & Totten, J. Empathy and attribution: Turning observers into actors. *Journal of Personality and Social Psychology,* 1975, *32,* 850–857.

Sicoly, F., & Ross, M. Facilitation of ego-biased attributions by means of self-serving observer feedback. *Journal of Personality and Social Psychology,* 1977, *35,* 734–741.

Snyder, M. L., Stephan, W. G., & Rosenfield, D. Egotism and attribution. *Journal of Personality and Social Psychology,* 1976, *33,* 435–441.

Stevens, L., & Jones, E. E. Defensive attribution and the Kelley cube. *Journal of Personality and Social Psychology,* 1976, *34,* 809–821.

Storms, M. D. Videotape and the attribution process: Reversing actors' and observers' points of view. *Journal of Personality and Social Psychology,* 1973, *27,* 165–175.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1971.

Received March 9, 1978

Journal of Personality and Social Psychology
1979, Vol. 37, No. 2, 273-285

Sex Differences in Eavesdropping on Nonverbal Cues

Robert Rosenthal and Bella M. DePaulo
Harvard University

Three series of studies investigated the hypothesis that nonverbally, women are more interpersonally accommodating than men. The first series of studies showed that women lost much of their advantage in decoding visual cues when the cues were based on displays too brief to be under good sender control. The second series of studies showed that as nonverbal cues became less intended (more "leaky"), women showed decreasing advantage over men in accuracy of decoding nonverbal cues. There was also a trend for women who were more skilled at eavesdropping on nonverbal cues to be seen as having less successful social outcomes. Women were also more biased to use (the more controllable) visual cues than tone of voice cues and especially so when the video cues were of the face rather than of the "leakier" body. The third series of studies showed that women were more polite in their ascription of characteristics to others, more accurate in decoding of nondeceptive behavior, but substantially more likely to interpret deceptive cues as the deceiver wanted them to be interpreted. Finally, it was shown that women's nonverbal cues were more easily read than men's.

The superiority of women over men in the accuracy of decoding nonverbal cues has long been postulated and investigated. Recently, these investigations have been exhaustively, systematically, and quantitatively summarized so that the issue is no longer in doubt; women are indeed superior to men in the decoding of nonverbal cues (Hall, 1978).

Also based on the results of dozens of studies of sensitivity to nonverbal cues is the general finding that people who are more skillful at decoding nonverbal cues are more effective in their interpersonal relationships (Hall, Rosenthal, Archer, DiMatteo, & Rogers, 1978; Rosenthal, Hall, Archer, DiMatteo, & Rogers, 1979; Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979). Given that women in our culture have traditionally played more expressive, socioemotional, and supportive rather than instrumental roles (Parsons, 1955; Parsons, Bales, & Shils, 1953; Zelditch, 1955), their superiority at decoding nonverbal cues, a concomitant of superior interpersonal relationships, seems quite consistent.

Recently, however, a question has been raised about the conditions under which skill at decoding nonverbal cues might actually be detrimental to smooth social functioning (Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979). The evidence suggests that social relationships may suffer when people are too good at decoding nonverbal messages they were not intended to receive. Thus, for example, people who are very good decoders of very briefly exposed (e.g., 42 msec) facial and body cues may have less satisfactory interpersonal relationships (Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979).

We thank Faye Crosby, Michael Fridkin, Carolyn Rieder Green, Judith A. Hall, Judith Rosenkrantz, Shelley Taylor, and Miron Zuckerman for their helpful comments on an earlier draft of this manuscript.

Preparation of this article was supported by the Milton Fund of Harvard University and Biomedical Sciences Support Grant 5S07 RR07046-12 from the National Institutes of Health to Harvard University. During the conduct of this research, the second author was supported by a National Science Foundation Fellowship.

Requests for reprints should be sent to Robert Rosenthal, Department of Psychology and Social Relations, Harvard University, William James Hall, 33 Kirkland Street, Cambridge, Massachusetts 02138.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3702-0273$00.75

If it is disruptive of smooth interpersonal functioning for a participant to "know too much" about the state of the other, we would expect women to show relatively less advantage over men in decoding nonverbal cues when those cues are under less control of the sender and are more likely to be unintended than intended cues. The plausibility of this hypothesis is strengthened by Weitz's (1976) finding that the nonverbal styles of females are closely adjusted to the personality traits of male partners.

Among the various channels of nonverbal communication that have been studied, the most informative, and also the one most likely to be best controlled by the sender, is the face (Ekman & Friesen, 1969; Izard, 1971; Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979). Compared to the face, each of the following sources of nonverbal cues is likely to be less informative under ordinary circumstances of interaction and less likely to be as controllable by the sender:

1. *Body.* Ekman and Friesen (1969, 1974) have shown that the body is more likely than the face to give off or "leak" deception cues.

2. *Tone of voice.* Several studies (Ekman, Friesen, & Scherer, 1976; Streeter, Krauss, Geller, Olson, & Apple, 1977; Krauss, Geller, & Olson, Note 1) have shown that tone of voice is an additional source of cues to deception or stress. It has also been shown that tone of voice may leak one's true feelings about oneself (Bugental, Henker, & Whalen, 1976; Bugental & Love, 1975; Holzman & Rousey, 1966) or about others (Weitz, 1972).

3. *Very brief exposures.* Very brief exposures to face or body cues appear likely to offer further unintended cues to the other's state without being as controllable as cues (in the same channels) occurring in longer time frames (e.g., 42 msec vs. 2,000 msec) (Ekman & Friesen, 1969; Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979).

4. *Discrepancy.* Except in the special cases of irony, sarcasm, or humor, discrepancies between visual and auditory nonverbal cues may be unintended and difficult to control (DePaulo, Rosenthal, Eisenstat, Rogers, & Finkelstein, 1978).

The purpose of the present series of studies was to investigate the hypothesis that nonverbally, women are interpersonally more accommodating than are men. We wanted to know whether women's greater social civility would be evident in their decoding skills, in their attentional biases, in their interpretive biases, and in their nonverbal expressiveness. More specifically we wanted to know the following:

1. Do women show relatively less advantage over men in their decoding accuracy as the nonverbal cues decoded become less intended and less controllable by the sender (i.e., are women less likely to eavesdrop on the leakages of others, taking into account women's general superiority at decoding nonverbal cues)?

2. Do women, more than men, attend preferentially to the more overt and more controlled nonverbal channels?

3. Do women, more than men, ascribe socially more desirable characteristics to the nonverbal cues they read?

4. Do women, more than men, interpret nonverbal cues in ways in which the sender would like them to be interpreted?

5. Do women, more than men, permit their nonverbal cues to be read by others?

6. Given the degree of sex role stereotyping that still exists, are there social costs to women who are less interpersonally accommodating in these nonverbal ways?

Study Series 1

Very Brief Exposures

A series of 10 studies was conducted to test the hypothesis that for visual cues of very brief exposure, women's advantage over men in accuracy of decoding would be smaller than that established for visual cues available in longer time frames (Hall, 1978).

Method

The Brief Exposure Profile of Nonverbal Sensitivity (PONS), a test of sensitivity to nonverbal visual cues with median exposure lengths of only 250 msec (Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979), was administered to six samples of college students ($N = 352$), three samples of adults ($N = 87$), and one sample of high school students ($N = 111$). For each of the 10 samples, the magnitude of females' superiority was computed by subtracting

Table 1
Stem-and-Leaf Plots of Female Superiority (in σ Units) in Sensitivity to Ordinary Versus High-Speed Visual Cues

Value	Ordinary cues[a]		High-speed cues[b]	
	Stems[c]	Leaves[d]	Stems	Leaves
	1.8	6		
	1.3	1		
	1.0	2		
	.6	0 5 7 9		
	.5	3 4		
	.4	0 0 8		
	.3	0 0	.3	2 8
	.2	6	.2	0 2 2 2 2
	.1	0 1 4 4	.1	
	.0	0 0 2 3 5	.0	
	−.0	2	−.0	6
	−.1	2 5	−.1	2
	−.2		−.2	4
	−.3	1		
	−.6	0		
Maximum	1.86		.38	
Quartile 3	.57		.24	
Mdn	.26		.22	
Quartile 1	.01		−.08	
Minimum	−.60		−.24	
$Q_3 - Q_1$	.56		.32	
$\hat{\sigma}\,[.75(Q_3 - Q_1)]$	.42		.24	
S	.49		.20	
M	.32		.14	

[a] Based on 29 studies (Hall, 1978); $N = 5{,}671$ subjects.
[b] Based on 10 studies (present Study Series 1); $N = 550$ subjects.
[c] Female superiority in nonverbal sensitivity given in σ units to one decimal place.
[d] Second decimal place for each study of female superiority in nonverbal sensitivity.

the mean accuracy of the males from the mean accuracy of the females and dividing the difference by the common standard deviation of the male and female subsamples (Cohen, 1977; Rosenthal & Rosnow, 1975). An analysis of variance showed that the three types of samples (college, adult, and high school) did not differ in the degree to which the females outperformed the males, $F(2, 7) < 1$; accordingly, all subsequent analyses disregarded the age level of the samples.

Results

Table 1 shows the stem-and-leaf plot (Rosenthal & Rosnow, 1975; Tukey, 1977) of our results side by side with the stem-and-leaf plot of the results of the 29 studies reported by Hall (1978), in which she computed the magnitude of female superiority in sensitivity to visual cues exposed for at least several seconds, on the average.[1] The most obvious difference between the two distributions is the percentage of studies of ordinary-length visual cues in which women showed larger advantages over men (41%) than was true in any study of high-speed cues. In contrast, the percentage of studies of ordinary cues in which women showed smaller advantages over men than was true in any study of high-

[1] Tukey (1977) developed the stem-and-leaf plot as a special form of frequency distribution to facilitate the inspection of a batch of data. Each number in the data batch is made up of one stem and one leaf, but each stem may serve several leaves. Thus, the fifth stem of Table 1, a .5, is followed by two leaves of 3 and 4 representing the two numbers .53 and .54. The first digit is the stem; the second is the leaf. The eye takes in a stem-and-leaf plot as it does any other frequency distribution, but the original data are preserved with greater precision in a stem-and-leaf plot than would be the case with ordinary frequency distributions.

speed cues was only 7%, a difference that was significant at $p = .006$ (sign test) or at $p = .08$ by the less focused Kolmogorov-Smirnov test, $\chi^2(2) = 5.09$. The summaries of the two distributions given at the bottom of Table 1 show that the two variances differ substantially, $F(28, 9) = 6.00$, $p = .008$, with the bulk of this heterogeneity due to the studies of ordinary nonverbal cues showing very large advantages of females over males.

In summary, 10 studies of sex differences in accuracy of judgments of briefly exposed cues showed that women were markedly less advantaged at this task than they were at reading visual cues of ordinary duration. These results suggest that women may be more polite than men in their decoding of nonverbal cues (i.e., refraining from decoding too efficiently those nonverbal cues under less control of the encoder).

Study Series 2

Ordered Leakage Channels

Method

Five measures of sensitivity to nonverbal cues derived from the PONS were administered to samples of 148 high school students and 94 college students. Details of the first four measures are given in Rosenthal, Hall, DiMatteo, Rogers, and Archer (1979), and details of the fifth measure are given in DePaulo et al. (1978). (See Appendix.) Briefly, the measures were as follows:

1. *Face.* A 20-item test of sensitivity to facial expressions.
2. *Body.* A 20-item test of sensitivity to body movements.
3. *Tone.* A 40-item test of sensitivity to speech masked by content-filtering (Rogers, Scherer, & Rosenthal, 1971) and random-splicing techniques (Scherer, 1971).
4. *Brief exposures.* A 40-item test of sensitivity to high-speed exposures of the face or of the body with a median exposure length of only 250 msec.
5. *Discrepancies.* A 128-item test of sensitivity to the degree of discrepancy between the tone of voice and either facial expressions or body movements.

The five measures are listed in the order in which we believed them to fall on a dimension of "leakiness." Thus, the face cues were felt to be the best controlled or least "leaky," whereas discrepancy cues were felt to be least well controlled or most "leaky." Ekman and Friesen's (1969, 1974) work convinces us that the body may be a leakage channel, though Feldman (1976) suggests that this may not always be the case. Evidence for the leakiness of tone-of-voice cues is more pervasive (deriving from studies of personality, psychopathology, and attitudes, as well as from studies of deception) and is as yet uncontested. Briefly exposed cues may be similar to the "microaffect displays" discussed by Ekman and Friesen (1969). These microdisplays may be mere fragments of full displays, or they may be complete movements that are greatly reduced in time. In either case, they should be more difficult to control than any of the full-channel messages communicated at ordinary speeds. Finally, we believe that discrepant messages would be most "leaky" or difficult to control because (a) they involve simultaneous communication by two different channels, each of which might itself be difficult to control, and (b) the combined effect of both messages together might add a further element of subtlety or uncontrollability.

The specific ordering of these five measures was necessarily highly speculative, since prior studies had rarely ever compared more than two of these channels simultaneously. However, it should be noted that given only the placement of our most and least leaky channels (face and dicrepancies), our particular ordering of the five measures (with linear contrast weights of +2, +1, 0, −1, −2) correlated very highly ($r = .89$) with the ordering that does not make any discrimination among the middle three measures (with contrast weights of +1, 0, 0, 0, −1). Or, even more conservatively, if we had merely claimed that the face was less leaky than any of the other channels, the resulting ordering (+4, −1, −1, −1, −1) would still correlate very highly ($r = .71$) with the present ordering.

Results

Table 2 shows the mean intercorrelations among the five measures for the two samples. The fact that all correlations were positive, most of them significantly so, shows that these measures were clearly related to each other, but not to as great an extent as we would expect if all the measures were tapping only a single underlying skill. The diagonal of the matrix of Table 2 shows the internal consistency reliability for each of the measures. The median reliability of .40 was substantially greater than the median intercorrelation (.22) among the five measures, supporting the relative independence of the five measures.

The bottom row of Table 2 shows the degree of accuracy exceeding chance (in z units) for each of the five measures. As we would expect, the channels posited as more leaky or less controllable tended to be less accurately decoded, $r(3) = .87$, $p < .03$, one-tailed. However, even the most difficult cu

Table 2
Intercorrelations, Reliabilities, and Accuracy Levels of Five Measures of Sensitivity to Nonverbal Cues

Measure	Face	Body	Tone	Brief exposures	Discrepancies
Face	**.40**				
Body	.21*	**.36**			
Tone	.30***	.08	**.28**		
Brief exposures	.40***	.33***	.27**	**.53**	
Discrepancies	.16	.12	.17*	.22*	**.59**
Accuracy (in σ units)	3.88	2.82	1.82	2.58	1.12

Note. Internal consistency reliabilities are in boldface type in the diagonal. Accuracy is defined by the degree to which performance exceeds chance (in σ units).
* $p \leq .05$.
** $p \leq .01$.
*** $p \leq .001$.

to decode, those of discrepancy, were decoded very much better than chance (1.12σ). Inspection of the reliabilities of the five measures also shows that differences in reliability could not be responsible for differences in accuracy. The most reliable measure was also the one with the lowest accuracy, and the overall correlation between test accuracy and reliability was negative, $r(3) = -.33$, *ns*.

Table 3 shows the results of our analysis for each sample and for both combined. The magnitude of females' superiority over males in decoding decreased in a linear manner, going from less to more leaky channels. Based on the analyses of variance testing the Sex × Linear Trend interaction, the college sample showed the regression ($r = .96$) to a significant degree, the high school sample showed the regression ($r = .72$), but not to a significant degree, and the combined samples showed the regression ($r = .91$) to a significant degree ($p = .004$). As channels became more and more leaky, females lost more and more of their advantage. These results, too, suggest that women may be more polite in their decoding of nonverbal cues (i.e., refraining more

Table 3
Female Superiority in Sensitivity to Five Types of Nonverbal Cues (in σ units)

Decoding skill	Sample: High school[a] ($N = 109$)	Sample: College[b] ($N = 81$)	M[c]
Face	.36	.38	.37
Body	.24	.34	.29
Tone	.32	−.02	.15
Brief exposures	.18	−.18	.00
Discrepancies	.22	−.28	−.03
F (linear contrast)	1.25	7.43	—
df error	276	196	
p	.265	.007	.004
r (predicted with obtained)	.72	.96	.91

Note. Ns vary from row to row; median Ns are given at head of columns.
[a] The mean effect size of .26 was significantly greater than zero ($p = .003$).
[b] The mean effect size of .05 was not significantly greater than zero ($p = .59$).
[c] The mean effect size of .16 was significantly greater than zero ($p = .006$); the correlation over the five skills between the high school and college samples was .87, $p < .03$, one-tailed.

Table 4
Loss of Female Superiority in Sensitivity to Four Types of Nonverbal Cues for Earlier and Present Studies (in σ units)

Difference between face and:	Earlier studies	No. of earlier studies	Present studies (2)	Total studies (weighted)
Body	.10	2	.08	.09
Tone	.14	10	.22	.15
Brief exposures	.19	10	.37	.22
Discrepancies	.26	1	.40	.35
Mdn	.17		.30	.19
Unweighted *M*	.17		.27	.20
Weighted *M*	.16		.27	.19

Note. Loss of female superiority was defined as the difference between female superiority in the face channel (in σ units) and female superiority in each of the leakier channels (in σ units).

and more from decoding effectively those nonverbal cues that are less and less controllable by the encoder).

Additional Evidence

Although we knew of no other studies in which sex differences in sensitivity to nonverbal cues had been specifically examined for a series of channels differing in leakiness, we tried to find studies in which at least one relatively leaky channel could be compared to the least leaky channel, the face, with exposure length not too brief.

We found 2 studies in which female superiority had been measured on the body channels using the same test as employed in the present article (Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979, chapters 6 and 8); 10 studies in which female superiority had been measured in the tone-of-voice channel (Hall, 1978); 10 studies in which female superiority had been measured for brief-exposure visual cues (the present article); and 1 study in which female superiority had been measured for discrepant cues (DePaulo et al., 1978). In the 2 studies measuring female superiority in the body channel, female superiority in the face channel was measured for the same sample. In all other cases, female superiority in the face channel was defined by Hall's visual category data (mean effect size = .32σ).

Table 4 shows the loss (in σ units) of female superiority in going from the face channel to each of the four more leaky channels. Just as was the case for our two samples of Table 3, the earlier studies yielded an ordering of loss of female superiority that went from face to body to tone to brief exposure to discrepant cues. When these earlier studies were combined with our two studies, the degree of linear relationship between the ordering of leakiness of channel and the loss of female superiority was very great, $r(3) = .99$, $p < .001$.

Social Consequences of Eavesdropping

Teacher Ratings

Method. To obtain an indication of whether eavesdropping skill might have social consequences, and more particularly to see whether females might be penalized for being too accurate at reading hidden cues, we examined teachers' ratings of the social effectiveness of the high school students who had been administered the five measures described earlier. Several teachers ($M = 2.4$, range = 1–5) rated the students whom they knew on their popularity with the same sex, popularity with the opposite sex, and degree of social understanding. We correlated all three ratings with degree of skill in each of the five nonverbal tasks, separately for males and females, yielding a total of 30 correlations. Because the patterns of results for the three ratings were so similar (the intercorrelations among the three variables were .69, .79, and .90), we averaged the correlations obtained from the three ratings to yield the summary correlations of Table 5.

Results. Table 5 shows a strong trend, though not a significant one, for social effectiveness to be less and less strongly related to nonverbal sensitivity as we move from less to

Table 5
Correlations Between Sensitivity to Five Types of Nonverbal Cues and Teachers' Ratings of Social Effectiveness

Decoding skill	Males (n = 43)	Females (n = 63)	Difference	M
Face	.36*	.38**	−.02	.37***
Body	.10	.03	.07	.07
Tone	.42**	.23	.19	.33***
Brief exposures	.33*	.16	.17	.25**
Discrepancies	.03[a]	−.06[b]	.09	−.02
Correlation (n = 5) between magnitude of relationship and degree of "leakiness" of channel	−.40	−.69	.60	−.56

Note. ns vary from row to row; median ns are given at head of columns.
[a] The analogous value for a sample of 36 junior high school males was .18.
[b] The analogous value for a sample of 37 junior high school females was −.12.
* $p \leq .05$.
** $p \leq .01$.
*** $p \leq .001$.

more "leaky" channels. This trend was found for both males (r = −.40) and females (r = −.69), but the trend was stronger for females, as shown in the third column of Table 5. Relative to males, correlations between social effectiveness and nonverbal sensitivity dropped off more linearly for females in going to more leaky channels (r = .60). There are indications, therefore (and that is all they are), that eavesdropping skill may not be as socially beneficial as skill at decoding less "leaky" channels. These indications are stronger for females than for males, although we should note that even for females, the average correlation still tends to be positive rather than negative.

Self-Ratings

Method. For both the high school and college samples described earlier, we obtained students' self-ratings of the quality of their relationships with members of their own and of the opposite sex. These ratings were correlated with skill at decoding each of the five types of nonverbal cues, separately for males and females. Because the results for the ratings of same- and opposite-sex relationships were similar, we present the average of the correlations of each rating with each skill in Table 6, separately for the males and females of the two samples.

Table 6
Correlations Between Sensitivity to Five Types of Nonverbal Cues and Self-Ratings of Social Relationships

Decoding skill	High school		College	
	Males (n = 36)	Females (n = 43)	Males (n = 32)	Females (n = 49)
Face	.14	−.13	.09	−.09
Body	−.05	−.15	.22	.18
Tone	.03	−.04	−.23	−.18
Brief exposures	.06	−.17	.12	−.23
Discrepancies	.29	.03	−.03	.06
Correlation (n = 5) between magnitude of relationships and degree of "leakiness" of channel	.50	.56	−.31	−.10

Note. ns vary from row to row; median ns are given at head of columns.

Table 7
Correlations Between Sensitivity to Five Types of Nonverbal Cues and Bias Favoring Positive Affect

Decoding skill	High school		College		
	Males (n = 46)	Females (n = 63)	Males (n = 32)	Females (n = 50)	Interaction[a]
Face	.02	.09	.14	−.12	.33
Body	−.09	.19	−.03	−.20	.45
Tone	−.01	−.14	.12	−.15	.14
Brief exposures	−.03	.12	.22	−.04	.41
Discrepancies	−.17	.11	.01	−.33	.62
Correlation (n = 5) between magnitude of relationship and degree of "leakiness" of channel	−.67	−.04	−.02	−.38	

Note. ns vary from row to row; median ns are given at head of columns.
[a] The difference between the correlations on the diagonals of the 2 × 2 table formed from the factors gender and age, for example, [(.09) + (.14)] − [(.02) + (−.12)] = .33.

Results. Table 6 shows that self-ratings of the quality of relationships with others were not strongly related to sensitivity to nonverbal cues (median r = .00), a finding that is consistent with the results of 10 earlier studies (median r = .06; Hall et al., 1978).

The last row of Table 6 provides an interesting hypothesis for future testing. Among high school males and females, there is a tendency (rs = .50 and .56, respectively) for correlations between nonverbal skill and self-ratings to grow larger as channels increase in leakiness. Thus, for our younger students, skill at eavesdropping on increasingly leaky channels was associated with increasingly better social relationships, as *they* perceived these relationships. These results are exactly opposite to those obtained when teacher ratings had been employed as the criteria of social effectiveness (rs = −.40 and −.69 for males and females, respectively).

When we examine the analogous results for the college males and females of Table 6, we find the trend to be reversed. For our older students, skill at eavesdropping on increasingly leaky channels was associated with increasingly poor social relationships, as they perceived these relationships. Perhaps these students have learned, as have the younger students' teachers, that it may not be beneficial to social relationships to be an effective eavesdropper on leaked nonverbal cues. It may well be that this knowledge is tacit rather than explicit (Polanyi, 1962).

Biases and Nonverbal Skill

The Nonverbal Discrepancy Test (DePaulo et al., 1978), in which conflicting visual and auditory cues are presented, provides several measures other than accuracy at decoding discrepant cues. Two of these measures are indexes of bias in the processing of nonverbal cues. One of these biases, the positivity bias, is the extent to which the subject is more influenced by positive than by negative nonverbal cues. Thus, when either video or audio positive cues are paired with negative cues in the remaining channel, the positivity-biased subject rates the scenario as more positive. These bias scores are corrected for the tendency to rate all scenarios as positive by subtracting ratings of positiveness made to scenarios showing no discrepancy of affect between channels. The other bias, video bias, is the extent to which the subject is more influenced by video than by audio nonverbal cues.

Positivity Bias

Table 7 shows the correlations between subjects' degree of positivity bias and their skill at decoding five types of nonverbal cues. Results are given separately for the male and

female subjects of the high school and college samples we have been discussing.

For male subjects, there is a tendency for these correlations to be more positive at the college level than at the high school level, whereas for the female subjects, there is a tendency for these correlations to be more negative at the college than at the high school level; for interaction of Sex × Sample, $F(1, 2) = 11.85$, $p < .08$. One interpretation of these results is that relative to men, as women mature, those who tend to focus more on the "good" in nonverbal cues tend to become less accurate decoders of nonverbal cues, perhaps in an effort to avoid knowing "more than they should" about others' states. If this interpretation were reasonable, we might expect to find that the interaction described would be especially true for skills of decoding the less controllable, more leaky channels. The statistical test for the systematic growth of a two-way interaction along an ordered dimension is the A × B × Linear Trend in C interaction, which for these data was the interaction of Sex × Sample × Linear Trend in Leakiness of the Nonverbal Skills. Results of this analysis supported our interpretation, $F(1, 3) = 18.25$, $p < .03$. That is, the tendency for women, relative to men, to become less accurate at nonverbal decoding as they mature, when they are more inclined to see the "good" in nonverbal cues, becomes systematically greater when skill at nonverbal decoding would make them more able and likely to eavesdrop.

Video Bias

When people are confronted with contradictory nonverbal cues in the video and audio channels, there is a general tendency for them to be more influenced by the video than the audio cues (DePaulo et al., 1978). People show interesting individual differences, however, in the degree to which they are biased in favor of video cues.

One concomitant of a greater video bias may be the tendency to focus more on intended rather than on leaked nonverbal cues, since video bias is most pronounced for facial cues, and facial cues are felt to be best controlled and least leaky. As we might expect from the other results presented in this article, we found females to be more biased in favor of video nonverbal cues in both the high school and the college samples (mean effect size for sex difference = $.55\sigma$; $df = 161$; $p < .001$). We also found that this sex difference in video bias was significantly greater when the video cues were from the face rather than from the body (mean effect size for Sex × Channel interaction = $.37\sigma$; $df = 161$; $p = .02$).

These results are consistent with the interpretation that women, more than men, are more likely to attend to better controlled rather than to more leaky channels of nonverbal communication. Still greater force is given to this interpretation by the finding of a trend for this sex difference in video bias to be greater when self-reported relationships with the opposite sex were above average in quality (effect size = $.20\sigma$; $df = 161$; $p < .16$). Perhaps women are reinforced with better opposite-sex relationships when they attend to the better controlled channels and learn to avoid eavesdropping on the more poorly controlled, more leaky channels.

Study Series 3

Perception of Others

In a third series of studies, 20 male and 20 female college students were videotaped while describing a person they liked, one they disliked, one they were ambivalent about, one they were indifferent to, a person they really liked as though they disliked him or her, and a person they really disliked as though they liked him or her. The first two encodings were of clear affects, the second two encodings were of mixed affects, and the third two encodings were of deceptive affects. Each of the 40 encoders also served as decoder and rated the sendings of 20 of the encoders on the degree to which the sending reflected liking, disliking, ambivalence, discrepancy, tension, and deception.

There were large overall differences in the average ratings assigned to encodings by decoders. Perhaps as part of the pattern of seeing only what one is supposed to see, females rated all encodings as significantly less tense

(effect size = 1.45σ; $t = 4.10$, $p < .001$), significantly less ambivalent (effect size = 1.48σ, $t = 4.19$, $p < .001$), and significantly less discrepant (effect size = 0.84σ; $t = 2.38$, $p < .02$) than did males. There were no sex differences in ratings on the other three scales ($t < 1$).

Overlooking Deception

Each decoder's ratings could be scored for accuracy of decoding. For example, a decoder was scored as more accurate in decoding liking sendings if higher liking ratings were made to the encodings of liking than to the encodings of disliking. For both the liking and disliking accuracy variables, women were substantially and significantly more accurate than men, with effect sizes of $.92\sigma$ ($p < .001$) for liking and $.68\sigma$ ($p < .001$) for disliking. Thus, at this task, as at so many others, women were more accurate than men at decoding nonverbal cues.

A result of greater interest, however, was that compared to men, women were substantially and significantly more likely to interpret the deceptive encodings as the deceiver wanted them interpreted rather than as the deceiver really felt. That is, when senders were describing the person they liked as though they disliked him or her, females were more likely than males to believe that the sender was communicating disliking; analogously, women were more likely to interpret "dislike as though like" descriptions as communications of liking. Effect sizes were $.78\sigma$ ($p < .001$) for liked as though disliked and $.95\sigma$ ($p < .001$) for disliked as though liked descriptions. In decoding deception, then, women are more likely politely to "read" what they are supposed to read rather than what is true.

Openness in Sending

There is considerable evidence to suggest that women are better encoders than men (Hall, Note 2), and the suggestion has been offered that women have been encouraged to be more emotionally expressive than men (Buck, Miller, & Caul, 1974; Buck, Savin, Miller, & Caul, 1972). One way to view these results, in the context of the present article, is that women not only eavesdrop less than men but that they are also more interpersonally accommodating by making themselves easy to read nonverbally.

In the present person description study, as in the studies reviewed by Hall (Note 2), we found that the nonverbal messages communicated by women were more easily decoded than the nonverbal expressions of men (effect size = $.73\sigma$, $p < .05$).

If women more than men allowed their nonverbal behaviors to be conveniently legible, we might also expect to find a sex difference on Snyder's Self-Monitoring Scale (1974). Persons who score high on this scale are thought to be more controlling of the cues they emit, whereas those who score low are thought much more to "wear their feelings on their sleeves." We administered the Self-Monitoring Scale to 37 males and 44 females of the high school sample discussed in Study Series 2 and to the 20 males and 20 females in the person description study. The sex differences in self-monitoring for these two samples, though modest in size ($.42\sigma$ and $.25\sigma$) were consistent with each other and with our expectations: Women, more than men, were likely to express their true feelings rather than to control or monitor their nonverbal expressions to fit a given situation (mean effect size = $.34\sigma$; combined $p = .03$).

Conclusions

Three series of studies were conducted that addressed the hypothesis that nonverbally, women will be more interpersonally accommodating than will men. The first two series of studies focused on the hypothesis that as nonverbal cues become less intended and less controllable by senders, women will show relatively less advantage over men in their decoding accuracy. Put another way, the hypothesis was that women would be relatively less likely than men to eavesdrop on the nonverbal leakages of others, correcting for women's general superiority at decoding nonverbal cues.

The first series of studies, based on 10 samples, showed that women's general ad-

vantage over men in decoding visual cues was to a large extent lost when the visual cues were of very brief duration (median duration of 250 msec). Exposures so short may be less controllable by the sender and may carry more unintended or leaked messages than exposures in longer time frames.

The second series of studies, based on two samples (and 23 earlier studies), showed that when five measures of skill in decoding nonverbal cues were arranged from most controllable to least controllable (most leaky), women showed a systematic decrease in their superiority over men in going from the less to the more leaky channels. Although our initial ordering of these five measures of skill on a dimension of leakiness was very tentative, the overall results presented in this article provide good support for the construct validity of this particular ordering of the five measures. Most compelling in this regard are the accuracy results, based on over 20 studies, which show that the initial ordering of these five types of cues fits very well with the ordering of the magnitude of women's loss of superiority in decoding these cues.

This additional evidence for the hypothesis that women were relatively less likely than men to eavesdrop on leaky nonverbal channels was accompanied by suggestive evidence that there may be social costs to eavesdropping. The greater one's skill at decoding the leakier channels, the relatively less effective are one's interpersonal relationships as judged by outside observers—a finding that was stronger for women than for men. This result, and the finding that older students show similar results for self-ratings of interpersonal relationships, however, are based on only two samples and should, therefore, be viewed only as hypotheses for considerable further evaluation.

The second series of studies also suggested that as women mature, those who tend to focus on the positive or "good" in the nonverbal cues of others may tend to become less accurate decoders of nonverbal cues, and more especially so as these nonverbal cues become less intended, less controllable, or more leaky. This series of studies also showed that relative to men, women were more biased to utilize visual cues than tone-of-voice cues, and more so when these cues were of the face rather than of the body. Once again, then, women appeared more likely to use the information from channels under better sender control and offering lower degrees of leakiness. This sex difference in video bias tended to be stronger among those with better opposite-sex relationships.

The third series of studies showed that women rated the behavior of others as less tense, less ambivalent, and less channel discrepant—results that fit well with the pattern of seeing only what it is polite to see. This series also showed that women were markedly superior to men in the decoding of clear, nondeceptive behavior. However, when deception cues were being emitted, women were substantially more likely to interpret these cues as the deceiver wanted them to be interpreted, rather than as the deceiver really felt. Finally, this series of studies suggested not only that women were less likely to eavesdrop on leaky nonverbal channels but also that they were more likely to be accurately decoded by others.

Taking together the results of our three series of studies, it seems reasonable to conclude that women are more polite in the nonverbal aspects of their social interactions than are men. They are more guarded in reading those cues that senders may be trying to hide, but more open in the expression of their own affective states. Further, there are indications that women who are less accommodating in these nonverbal ways experience less successful interpersonal outcomes. Perhaps women in our culture have learned that there may be social hazards to knowing too much about other people's feelings. This relative avoidance of eavesdropping by women is consistent with the standards of politeness and social smoothing-over that are part of the traditional sex role ascribed to women in our culture, a sex role that is only now beginning to change.

Reference Notes

1. Krauss, R. M., Geller, V., & Olson, C. *Modalities and cues in the detection of deception.* Paper presented at the annual meeting of the American

Psychological Association, Washington, D.C., September 1976.

2. Hall, J. A. *Gender effects in encoding nonverbal cues*. Paper submitted for publication, 1978.

References

Buck, R., Miller, R. E., & Caul, W. F. Sex, personality, and physiological variables in the communication of affect via facial expression. *Journal of Personality and Social Psychology*, 1974, *30*, 587–596.

Buck, R., Savin, V. J., Miller, R. E., & Caul, W. F. Communication of affect through facial expressions in humans. *Journal of Personality and Social Psychology*, 1972, *23*, 362–371.

Bugental, D. B., Henker, B., & Whalen, C. K. Attributional antecedents of verbal and vocal assertiveness. *Journal of Personality and Social Psychology*, 1976, *34*, 405–411.

Bugental, D. B., & Love, L. Nonassertive expression of parental approval and disapproval and its relationship to child disturbance. *Child Development*, 1975, *46*, 747–752.

Cohen, J. *Statistical power analysis for the behavioral sciences* (Rev. ed.). New York: Academic Press, 1977.

DePaulo, B. M., Rosenthal, R., Eisenstat, R. A., Rogers, P. L., & Finkelstein, S. Decoding discrepant nonverbal cues. *Journal of Personality and Social Psychology*, 1978, *36*, 313–323.

Ekman, P., & Friesen, W. V. Nonverbal leakage and clues to deception. *Psychiatry*, 1969, *32*, 88–106.

Ekman, P., & Friesen, W. V. Detecting deception from the body or face. *Journal of Personality and Social Psychology*, 1974, *29*, 288–298.

Ekman, P., Friesen, W. V., & Scherer, K. R. Body movement and voice pitch in deceptive interaction. *Semiotica*, 1976, *16*, 23–27.

Feldman, R. S. Nonverbal disclosure of teacher deception and interpersonal affect. *Journal of Educational Psychology*, 1976, *68*, 807–816.

Hall, J. A. Gender effects in decoding nonverbal cues. *Psychological Bulletin*, 1978, *85*, 845–857.

Hall, J. A., Rosenthal, R., Archer, D., DiMatteo, M. R., & Rogers, P. L. The Profile of Nonverbal Sensitivity. In P. McReynolds (Ed.), *Advances in psychological assessment* (Vol. 4). San Francisco: Jossey-Bass, 1978.

Holzman, P. S., & Rousey, C. The voice as percept. *Journal of Personality and Social Psychology*, 1966, *4*, 78–86.

Izard, C. E. *The face of emotion*. New York: Appleton-Century-Crofts, 1971.

Parsons, T. The American family: Its relations to personality and to the social structure. In T. Parsons & R. F. Bales, *Family, socialization and interaction process*. New York: Free Press, 1955.

Parsons, T., Bales, R. F., & Shils, E. A. *Working papers in the theory of action*. New York: Free Press, 1953.

Polanyi, M. Tacit knowing: Its bearing on some problems of philosophy. *Reviews of Modern Physics*, 1962, *34*, 601–616.

Rogers, P. L., Scherer, K. R., & Rosenthal, R. Content-filtering human speech. *Behavioral Research Methods and Instrumentation*, 1971, *3*, 16–18.

Rosenthal, R., Hall, J. A., Archer, D., DiMatteo, M. R., & Rogers, P. L. The PONS test: Measuring sensitivity to nonverbal cues. In S. Weitz, *Nonverbal communication* (Rev. ed.). New York: Oxford University Press, 1979.

Rosenthal, R., Hall, J. A., DiMatteo, M. R., Rogers, P. L., & Archer, D. *Sensitivity to nonverbal communication: The* PONS *test*. Baltimore, Md.: The Johns Hopkins University Press, 1979.

Rosenthal, R., & Rosnow, R. L. *The volunteer subject*. New York: Wiley-Interscience, 1975.

Scherer, K. R. Randomized splicing: A note on a simple technique for masking speech content. *Journal of Experimental Research in Personality*, 1971, *5*, 155–159.

Snyder, M. Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology*, 1974, *30*, 526–537.

Streeter, L. A., Krauss, R. M., Geller, V., Olson, C., & Apple, W. Pitch changes during attempted deception. *Journal of Personality and Social Psychology*, 1977, *35*, 345–350.

Tukey, J. W. *Exploratory data analysis*. Reading, Mass.: Addison-Wesley, 1977.

Weitz, S. Attitude, voice, and behavior: A repressed affect model of interracial interaction. *Journal of Personality and Social Psychology*, 1972, *24*, 14–21.

Weitz, S. Sex differences in nonverbal communication. *Sex Roles*, 1976, *2*, 175–184.

Zelditch, M. Role differentiation in the nuclear family: A comparative study. In T. Parsons & R. F. Bales, *Family, socialization and interaction process*. New York: Free Press, 1955.

Appendix

The first four measures of sensitivity to verbal cues were derived from the PONS test, a 47-minute film consisting of 220 2-sec audio and/or visual nonverbal stimuli. In each 2-sec segment, a 24-year-old female acts in one of 20 different emotional situations. The 20 situations are categorized with reference to four different types of emotion, each created by the crossing of two affective dimensions: positivity–negativity and dominance–submission. Hence, there are five positive-dominant situations (e.g., talking to a lost child), five negative-dominant situations (e.g., expressing strong dislike), five positive-submissive situations (e.g., expressing deep affection), and five negative-submissive situations (e.g., asking forgiveness). The situations were originally categorized as either positive or negative and as dominant or submissive according to the ratings of two different samples of judges (Rosenthal, Hall, Archer, DiMatteo, & Rogers, 1979; Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979).

The 220 PONS items consist of a random ordering of these 20 situations, each represented in 11 different "channels" of nonverbal communication. Three channels are pure video channels: face only, body only (neck to knees), and face plus body. Two channels are pure audio channels: "content filtered" (CF) and "randomized spliced" (RS). In both of these channels, verbal messages are rendered incomprehensible. CF preserves sequence and rhythm (RS does not); RS saves pitch and intensity. The other six channels are "mixed" channels, consisting of all audiovisual combinations of the two audio with the three video channels.

From each of the 220 items, subjects select one of two situational labels, one of which correctly describes the situation and one which incorrectly describes it. The incorrect alternative was assigned to each item by randomly choosing one of the 19 situation labels that was not the correct answer.

The fifth measure, the Nonverbal Discrepancy Test (NDT), employs 8 of the 20 situations of the PONS test, 2 from each of the four affective quadrants of the PONS. Half of the scenes in the NDT are represented in the face channel, and half are represented in the body channel. In addition, half of the scenes represented in each of these two video channels are also represented in the content-filtered audio channel, and half of the scenes are represented in the randomized-spliced audio channel.

In the discrepancy test, each of the 8 scenes is paired with every other scene twice. Hence, there are 128 items in the test (8 scenes × 8 scenes × 2 replications). Each item consists of the simultaneous pairing of either a face or a body with a content-filtered or randomized-spliced voice. Every possible audio–video pairing (face–CF, face–RS, body–CF, body–RS) occurs exactly 32 times. For one quarter of the items, the audio and the video segments are from the same affective quadrant (e.g., a positive-dominant face might be paired with a positive-dominant voice). One quarter of the items consist of audio and video segments from exactly opposite quadrants (e.g., a positive-dominant face might be paired with a negative-submissive voice). The audio and video segments of the remaining items differ on only one of the affective dimensions. For example, a positive-dominant face might be paired with a positive-submissive voice. In this case, the discrepancy would be along the dominance dimension, since both evaluative inputs are the same (i.e., both are positive). Or, a positive-dominant face might be paired with a negative-dominant voice. In this case, both inputs assume the same value on the dominance dimension (i.e., both are dominant) but they differ on the evaluative dimension (i.e., one is positive and the other is negative).

Subjects rate each scene on a 9-point scale of discrepancy. Their accuracy (or sensitivity) is a function of the degree to which they rate as more discrepant those scenes in which the video and audio channels are, in fact, more discrepant, compared to their ratings of the scenes in which the video and audio channels are, in fact, less discrepant.

Received July 21, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 2, 286–300

Young Children's Preferred Attentional Strategies for Delaying Gratification

Brian T. Yates
American University

Walter Mischel
Stanford University

In four experiments, children controlled the frequency and duration of self-exposure to sets of stimuli during a delay of gratification. The stimuli included the real rewards for which they were waiting, symbolic (picture) versions of those rewards, or irrelevant objects. Previous research had shown that exposure to the real rewards hinders self-imposed delay, but viewing irrelevant stimuli or pictures of the rewards facilitates delay. In the present study, we recorded how long children viewed each stimulus during the delay and also assessed their verbal viewing preferences. Preschool children spontaneously tended to prefer to view the real rather than the symbolic stimuli, regardless of their relevance to the rewards in the delay contingency, both when the delay was self-imposed and when it was imposed externally. This generalized preference for attending to the "real rewards" rather than to more symbolic representations helps to explain why it is so difficult for young children to tolerate voluntary delay of gratification. By attending to the real rewards, they may make such delay especially frustrative and arousing, thereby defeating their own ability to wait for what they want. In contrast, children in Grades 1, 2, and 3 systematically selected effective attentional strategies during delay.

A great deal of experimental research has focused on the stimulus conditions that facilitate or impede delay of gratification. These studies have manipulated such variables as the presence or absence of the rewards and assessed how they influence the ability and willingness of subjects to tolerate various types of delay of gratification (e.g., Mischel, 1968, 1974; Mischel & Baker, 1975; Moore, Mischel, & Zeiss, 1976). The results provide much evidence about conditions that may facilitate waiting for deferred outcomes. In contrast, remarkably little is known about the subject's own natural attentional strategies for coping with various types of delay. Thus, we know a considerable amount about delay-enhancing stimulus conditions but almost nothing about children's preferred attentional strategies during delay of gratification. To begin to fill this gap, in the present set of studies we explored young children's verbal preferences and actual use of different attentional strategies for sustaining delay of gratification. Our results should help clarify the degree to which young children know and utilize delay-facilitating attentional strategies when faced with situations requiring them to wait for deferred but desired outcomes.

The delay of gratification paradigm used in the earlier research provided children with the dilemma of waiting for a more preferred but delayed reward or terminating their delay and getting a less preferred reward immediately (Mischel, Ebbesen, & Zeiss, 1972). Using this paradigm, several studies examined how attention to the delayed and immediate rewards changes children's ability to continu

This research was supported by National Institute of Mental Health Grant M6830 and National Institute of Child Health and Human Development Grant 09814 to Walter Mischel. We deeply appreciate the assistance in data collection that was provided by David Delatorre, Chris Haggenmacher, Kim Hallam, William Haven, Joanne Julius, Kim Kroener, Margie Meseroll, Sandy Montgomery, Kathy Scott, Beth Schneider, and Scott Whitely.

Requests for reprints should be sent to Walter Mischel, Department of Psychology, Stanford University, Stanford, California 94305.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3702-0286$00.75

to wait for the preferred outcome. Initially, it was predicted that attention to the rewards would facilitate delay of gratification. This prediction was based on Freud's (1911/1959) provocative idea that "cathecting" images of desired but blocked gratifications is a cornerstone for the development of frustration tolerance. Surprisingly, it was found that rather than facilitating delay by helping the subject anticipate the desired but blocked goal, attending to the delayed and immediate rewards appears to increase the child's arousal and frustration to the point where delay becomes too aversive to continue (Mischel, 1974). Specifically, the presence of the actual rewards during the delay period consistently produces shorter delay times than does the absence of the rewards (Mischel & Ebbesen, 1970; Mischel et al., 1972; Schack & Massari, 1973). When the rewards are not visible during delay, instructions to think about them have the same negative effect on delay time as do the rewards themselves (Mischel et al., 1972).

But while attention and thought directed at the real rewards hinder delay of gratification, attention to their symbolic representation, in the form of images (pictures of the rewards), facilitates delay (Mischel & Moore, 1973; Moore, 1974). Taken as a whole, these and related studies (e.g., Mischel & Baker, 1975; Moore et al., 1976) provide a coherent, theoretically meaningful set of rules that describe how attention to different types of stimuli can hinder or facilitate self-imposed delay of gratification. For the present investigation, the gist of these rules is that for preschool children, there are at least two basic strategies for using attention to facilitate delay of gratification and at least two basic attentional strategies that hinder delay. Attention to a symbolic or "cool" representation of the real reward objects that are in the delay contingency ("relevant rewards") as well as attention to comparable real objects that are irrelevant to the delay contingency ("irrelevant rewards") are effective strategies for self-imposed delay of gratification. Attention to the real, relevant rewards and attention to a symbolic representation of irrelevant rewards are both ineffective strategies that hinder delay, at least in the types of situations sampled in these studies.

In a recent series of related studies, Miller and Karniol (1976a, 1976b) have distinguished between self-imposed delay (SID) of the sort studied in the paradigm of Mischel and his associates and externally imposed delay (EID), in which the subject cannot voluntarily terminate the delay in favor of an alternative, immediate, but less preferred reward. Miller and Karniol believe that the findings of Mischel and associates apply to SID but not to EID. They argue that when individuals have no choice but to endure the delay, it is adaptive to focus on the rewards, and they have provided some empirical support for their view. Miller and Karniol's conclusions should still be seen as tentative, but their work does underline the need to distinguish between the two types of delay, and therefore, in the first study to be reported here, we included both the SID and EID paradigms.

Experiment 1

To investigate children's preferred attentional strategies for delaying gratification, the delay paradigm developed by Mischel et al. (1972) was modified. Instead of placing fixed stimuli in front of the child for the entire delay, we designed equipment that allowed the subject to self-regulate the presentation of stimuli throughout the delay period. Each child had available for self-presentation one of the two types of stimuli that facilitate SID and one of the two types of stimuli that hinder it. Different groups of children had available for viewing one of the four possible pairings of the two SID-facilitating and two SID-hindering stimuli. We recorded the amount of time that the children actually viewed the different types of stimuli during either a self-imposed delay of gratification (SID) or an externally imposed delay (EID) of equal duration and also assessed their preferences verbally after completion of the delay.

Table 1
SID-Facilitating and SID-Hindering Stimuli Available for Presentation in Different Stimulus-Pair Conditions

Stimulus pair		
SID-facilitating stimulus set	vs.	SID-hindering stimulus set
Picture of rewards (Relevant symbolic)		Actual rewards (Relevant real)
Irrelevant objects (Irrelevant real)		Irrelevant picture (Irrelevant symbolic)
Picture of rewards (Relevant symbolic)		Irrelevant picture (Irrelevant symbolic)
Irrelevant objects (Irrelevant real)		Actual rewards (Relevant real)

Note. SID = self-imposed delay.

Method

Overview and Design

Each child had one SID-facilitating and one SID-hindering set of stimuli available for viewing during the delay period. Two sets of stimuli exemplified the attentional strategies that have been found to hinder SID: (a) the actual rewards in the delay contingency (relevant real stimuli) and (b) a life-size color photograph of desirable objects that were irrelevant to the delay (irrelevant symbolic stimuli). Two other sets of stimuli exemplified the attentional strategies that have been found by previous research to facilitate SID: (c) a life-size color photograph of the rewards involved in the delay (relevant symbolic stimuli) and (d) a set of desirable objects that were irrelevant to the delay contingency (irrelevant real stimuli). All possible pairings of one set of delay-facilitating stimuli and one set of delay-hindering stimuli were tested. Table 1 lists the four stimulus-pair conditions used in this and the following studies.

In a 2 (SID versus EID type of delay) × 4 (stimulus pairs) factorial design, children were exposed to a delay of gratification while being able to self-present stimuli from one of the four types of stimulus pairs shown in Table 1. Duplicating the procedure used by Mischel and his colleagues for self-imposed delay of gratification (Mischel, 1974), we instructed children in the SID condition that they could terminate the delay at any time by pushing a button that rang a bell and signaled the experimenter to return. The children were also told that if they terminated the delay, they would receive the nonpreferred reward instead of the preferred reward. Children in the EID condition could not signal the experimenter to return. Instead, these children were yoked to children in the SID condition for duration of delay.[1] That is, if a child in the SID condition terminated the delay before the maximum time had elapsed, a child of the same sex in the EID condition was required to wait the same amount of time. The maximum delay period was 10 minutes.

Subjects and Experimenters

The subjects were 96 preschoolers (48 girls and 48 boys) who attended Bing Nursery School at Stanford University. Two male and two female experimenters tested equal numbers of male and female children in each cell of the design. Children were randomly assigned to conditions so that 6 males and 6 females participated in each cell of the design.[2] The children ranged in age from 3.0 to 5.2 years, with a mean of 4.4 years.

Apparatus

The experimental room resembled the one diagrammed in Mischel and Moore (1973). A visual barrier concealed a box of attractive toys in one corner. A small table and chair were set against the wall on the opposite side of the room. The chair was positioned inside a square that was outlined on the floor with black tape. Electromechanical equipment on the table enabled the child to control the frequency and duration of two stimulus sets. The buttons that the child pushed to present alternative stimuli for viewing required extremely little pressure to operate and were spaced 2 inches (5.1 cm) apart on the surface of a prominent metal box. The location of these buttons in relation to other apparatus is shown in Figure 1.

Two devices that contained the stimuli faced the child on his or her immediate right and left. Pushing the left button exposed the stimulus set on the child's left, and pushing the right button exposed the stimulus set on the child's right. An opaque cover was hinged to the back edge of the lower part of each device so that a solenoid held the cover open, presenting the set of stimuli for viewing only as long as the child continued to press the button. Stimuli were mounted at a 45° angle so that they were easily viewed by the child. When the child was not pressing the button, another solenoid held the cover closed, preventing the child from viewing stimuli without pushing a button. Only one device could be operated at a given time; pushing both buttons simultaneously exposed neither set of stimuli. To control for possible position preferences, half of the children in each cell

[1] Of course, the children were observed throughout the delay (through a one-way mirror), and the session would have been stopped immediately if the child appeared upset.

[2] Nine additional children began the experiment but were not included in the final sample because of problems with equipment or with the child's comprehension of instructions.

of the design had a stimulus set on their right, and the other children had the same stimulus set on their left.

Electromechanical equipment recorded the duration and frequency of the child's self-presentation of each stimulus set. Duration was recorded in units of .005 minutes. Relay circuitry initiated data recording automatically when the experimenter left the room and terminated data recording after 10 minutes.

Reward Stimuli

The alternative rewards used in the present study differed only in quantity; they were (a) three stick pretzels versus one stick pretzel (each pretzel was 7.6 cm × .3 cm) and (b) three regular-size poker chips (white, red, and blue) versus one regular-size poker chip (white). Half of the children in each cell chose between (and waited for) the food rewards, and the other children in each cell chose between (and waited for) the nonfood rewards. Thus, for the children who waited for food rewards, the stick pretzels were the relevant reward stimuli, and the poker chips were irrelevant. For the children who waited for poker chips, the poker chips were the relevant rewards, and the pretzels were the irrelevant stimuli, providing a complete crossover design. The stimulus pairs available for self-presentation included real and symbolic (picture) versions of the relevant and irrelevant items in the four combinations shown in Table 1. The four sets of stimuli from which these combinations were made were (a) three stick pretzels, (b) three poker chips, (c) a life-size color photograph of the three pretzels, and (d) a life-size color photograph of the three poker chips.

Procedure

Upon entering the room, the child was shown the toys that would be played with later. The experimenter then concealed the toys and asked the child to sit at a table on the other side of the room. The child was instructed to remain seated inside the square that was taped on the floor. The experimenter took a box from underneath the table, exposed the scheduled reward, and asked the child which alternative he or she would like to have. The experimenter explained that he or she would leave the room soon but would return later to give the chosen items to the child. These instructions were repeated until the child understood them. Afterwards, the rewards were returned to the box, which was then concealed under the table.

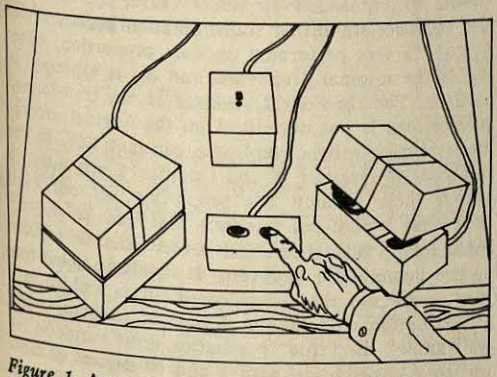


Figure 1. Apparatus for self-presentation of alternative stimulus sets, viewed from about 2 feet (61.5 cm) above child's eye level.

The experimenter proceeded to show the child how to use the stimulus presentation apparatus. So that the children would be less inclined to push both buttons simultaneously (a problem noted in pretesting), the children were asked to fold their hands together, to extend their forefingers, and to keep their forefingers together throughout the session. In essence, this procedure prevented the children from pushing both buttons simultaneously. After folding his or her hands and fingers together, the child was encouraged to push the buttons. Whenever the child pushed a button, the experimenter named the presented stimulus set, noting whether it was a set of real objects or a picture of real objects (e.g., "Oh, look! That button shows you a real _____ and a real _____!" or "Oh, look! That button shows you a picture of _____ and a picture of _____!"). The child was instructed to look closely at whichever stimuli were exposed, and the experimenter modeled intent viewing of whichever stimuli the child self-presented.

After the child had become familiar with the stimuli and the apparatus by presenting each stimulus set at least three times, he or she was asked not to push both buttons at the same time and not to touch the devices or stimuli. The child was then quizzed on which button produced which set of stimuli, to be certain that the child knew which stimuli were available for viewing and how to present each of the available stimulus sets. In this study, the children were given no specific instructions about how to structure their viewing of the alternative stimuli during the delay; they were simply asked to "push the buttons." The children also were not given specific information about when the experimenter would return but simply were reassured that the delay would not be "too long."

The experimenter left the room after reminding the child that he or she would play with the toys later. During the delay, the experimenter observed the child through a one-way mirror and returned to the room after 10 minutes. The child was then again shown the reward items and was asked which he or she was supposed to receive. All children remembered which one they previously had chosen to receive, were given that object, and next were asked the following questions in random order:

Question 1. If you could have just one of these (stimulus sets) in front of you while you were

Table 2
Stimulus Preferences in Each Pair on Each Measure (Experiment 1)

Stimulus pair	Measure: Question 1 (choice)	Measure: Question 2 (liking)	Measure: Viewing time
Relevant Real vs. **symbolic**	.83***	.75**	.57**
Irrelevant **Real** vs. symbolic	.83***	.71*	.59***
Symbolic **Relevant** vs. irrelevant	.79**	.62	.53
Real Relevant vs. **irrelevant**	.58	.62	.49

Note. The data are proportions: The larger the proportion, the greater the preference for the top stimulus in each pair. $n = 24$ for all measures. Stimuli that facilitate self-imposed delay (SID) are in boldface type. In all tables, proportions for Question 1 and Question 2 are not directly comparable to proportions for viewing time (see text). For Questions 1 and 2, binomial tests compared the data to chance levels; for viewing time, t tests compared the means to an ideal neutral proportion of .50. All tests are two-tailed throughout this article.

* $p = .06$.
** $p < .025$.
*** $p < .002$.

waiting, which one would you want in front of you all the time while you were waiting?
Question 2. Which one (of the stimulus sets) did you like to see the most while you were waiting?

After recording the child's answer, the experimenter led the child to the toys and, after they had played with them for 5 to 10 minutes, returned the child to the nursery school classroom.

Results

Children's answers to the postdelay questions and the relative amount of time they spent viewing alternative stimuli showed a consistent pattern: They preferred to view real stimuli regardless of their relevance to the delay (see Table 2) and regardless of the type of delay (EID or SID).

Answers to the postdelay questions were analyzed in terms of the proportion of children in a given cell of the design who answered the question by pointing to the delay-facilitating stimulus in the stimulus pair.[3] The relative amount of time spent viewing the delay-facilitating stimulus in a given pair was calculated for each child separately as the duration of viewing the delay-facilitating stimulus divided by the total duration of viewing both stimuli.

Preliminary analyses found no significant effects of food versus nonfood rewards. Separate analyses of variance conducted on the three measures found no significant main effects or interactions of SID versus EID: both Fs < 1 for the two questions; $F(1, 88) = 2.47$, $p > .10$, for relative time spent viewing stimuli.

To examine the effects of stimulus pairs, we combined data for types of delay within each stimulus-pair condition. These analyses showed a consistent pattern on all three measures. More specifically, the proportions in Table 2 express the relative preference for the stimulus listed first in each pair. When given a choice between relevant real versus relevant symbolic stimuli, the children preferred the real stimuli: $p < .002$ by a two-tailed binomial test for Question 1; $p = .02$ by a two-tailed binomial test for Question 2; $t(88) = 2.71$, $p < .01$, two-tailed, for time

[3] As recommended by Winer (1971, pp. 399–401), the variance-stabilizing transformation $\phi_{jk} = 2 \arcsin [(p_{jk})^{1/2}]$ was performed on each proportion, where p_{jk} is the original proportion and ϕ_{jk} is expressed in radians. The theoretical variance of the transformed proportions is not dependent on the original proportions and instead is simply the constant $1/N$. In the analyses of Question 1 and Question 2 responses, in which there is only one proportion per cell (and hence no within-cell variance that can be directly calculated), this theoretical error variance is used as the denominator in F tests. It should be noted that just as we use the normal distribution (which is a t distribution with infinite degrees of freedom) when we "know" the "true" population error variance, we use an F distribution with k and ∞ degrees of freedom when we know the true or "theoretical" error variance (cf. Langer & Abelson, 1972, for an example and extended discussion of this procedure).

spent viewing stimuli.[4] When given a choice between irrelevant real and irrelevant symbolic stimuli, children again preferred the real stimuli: $p < .002$ by a binomial test for Question 1; $p = .06$ by a binomial test for Question 2; $t(88) = 3.48$, $p < .002$, for time spent viewing stimuli. In contrast, children who could not choose between real and symbolic stimuli, but only between relevant and irrelevant stimulus sets that were both real or both symbolic, did not exhibit a consistently significant preference for either relevant or irrelevant stimuli: $p < .01$ in favor of relevant symbolic as opposed to irrelevant symbolic stimuli for Question 1; but $p > .30$ for Question 2 in the same stimulus pair; and $p > .30$ for both Question 1 and Question 2 in the relevant–real versus irrelevant–real stimulus pair. In addition, children did not exhibit a significant preference when actually viewing relevant versus irrelevant stimuli during the delay, regardless of whether both stimuli were symbolic or real ($ts < 1.2$; $ps > .20$).

Fourfold point correlations (Hays, 1963) showed that the three measures were significantly and substantially correlated with one another: $\phi = .43$, $p < .001$, between Question 1 and viewing; $\phi = .34$, $p < .005$, between Question 2 and viewing; $\phi = .55$, $p < .001$, between Question 1 and Question 2.

Before discussing the significance and implications of the results of the first study, we will describe two additional experiments; then the meaning of the total pattern of data will be examined.

Experiment 2

Because important changes in ideation and imagery may occur during the preschool years (e.g., Paivio, 1971; Piaget & Inhelder, 1966), our second study was developmental and systematically included children from three age groups: 3.00 to 4.25 years, 4.50 to 5.75 years, and 6.00 to 7.00 years. Thus, we used a 4 (stimulus pairs) × 3 (age groups) factorial design. Given the finding in Experiment 1 that preference patterns were highly similar for SID and EID, in our second experiment (and also in the third one to be reported in this article) we used only the EID paradigm, because it controlled length of delay time, thus preventing stimulus-viewing proportions from being confounded with duration of delay. Except for this change and for differences in the specific relevant and irrelevant reward stimuli used, the basic procedures and measures were essentially identical with those in Experiment 1.

Method

Reward Stimuli

To extend the generalizability of findings, somewhat different sets of reward stimuli were used. Specifically, the stimulus sets used in Experiment 2 were (a) one stick pretzel and one miniature marshmallow, (b) one small red poker chip (.75 inches, or 1.9 cm, in diameter) and one penny, (c) a life-size color photograph of the pretzel and the marshmallow, and (d) a life-size color photograph of the poker chip and the penny.

Half of the children in each cell of the design chose between receiving one pretzel and one marshmallow, and the other children in each cell chose between receiving one poker chip and one penny. Thus, for the former groups of children, the pretzel and marshmallow were the relevant reward items, and the poker chip and penny were the irrelevant items. For the other group of children, the poker chip and penny were the relevant reward items, and the pretzel and marshmallow were the irrelevant items. To avoid differences in stimulus viewing that might result from different durations of delay or from different durations of exposure to alternative stimuli, all children were exposed to a fixed period (12 minutes) of EID. Children were tested in random order in terms of stimulus-pair condition, age group, gender, and experimenter.

Subjects and Experimenters

The subjects were 96 preschool children (48 girls and 48 boys) who attended Bing Nursery School at Stanford University. Each cell of the design included 4 male and 4 female children. Two children of each sex delayed food gratification (pretzel/marshmallow), and two children of each sex delayed nonfood gratification (poker chip/penny). Two female and two male experimenters tested equal proportions of the children in each cell of the design.

---

[4] These t tests used the error term from the preceding analysis of variance as the pooled estimate of standard deviation, $(MS_e/n)^{1/2}$. These and all t tests in this article are two-tailed.

Table 3
Stimulus Preferences in Each Pair on Each Measure (Experiment 2)

Stimulus pair	Measure: Question 1 (choice)	Question 2 (liking)	Viewing time
Relevant			
Real vs. **symbolic**	.83***	.71*	.57**
Irrelevant			
Real vs. symbolic	.83***	.71*	.58***
Symbolic			
Relevant vs. irrelevant	.67	.67	.53
Real			
Relevant vs. **irrelevant**	.67	.58	.51

Note. The data are proportions: The larger the proportion, the greater the preference for the top stimulus in each pair. Stimuli that facilitate self-imposed delay (SID) are in boldface type. $n = 24$ for all measures.
* $p = .06$.
** $p < .025$.
*** $p < .002$.

Results

Children's answers to the postdelay questions and the relative amount of time that they actually spent viewing the alternative stimuli again showed the same consistent pattern. As summarized in Table 3, both verbally and behaviorally, children preferred to view real stimuli regardless of their relevance or irrelevance to the delay contingency.[5]

Separate analyses of variance conducted on the three measures revealed no significant main effect or interactions of age group (all Fs < 1). To examine the effects of stimulus pairs, subsequent analyses combined data for age groups within each stimulus-pair condition.

These analyses showed a consistent pattern on all three measures that closely matched the results of Experiment 1. Specifically, when given a choice between relevant real versus relevant symbolic stimuli, the children preferred real stimuli. Two-tailed binomial tests yielded $p < .002$ for Question 1 and $p = .06$ for Question 2; $t = 2.90$, $p < .01$, for time spent viewing stimuli. Similarly, when given a choice between irrelevant real and irrelevant symbolic stimuli, children preferred the real stimuli: $p < .002$ for Question 1; $p = .06$ for Question 2; $t(84) = 3.31$, $p < .002$, for time spent viewing stimuli. When the children could not choose between real and symbolic stimuli, but only between relevant and irrelevant stimulus sets that were both real or both symbolic, no significant preferences were found on any measure. Children who had these stimulus sets available showed a nonsignificant tendency to prefer relevant stimuli (ps $= .15$ for Question 1; ps $\geq .15$ for Question 2; ts ≤ 1.3 for time spent viewing stimuli).

Fourfold point correlations measured the strength of intermeasure relationships. The three measures were again significantly correlated with each other: $\phi = .35$, $p < .001$, between Question 1 and stimulus viewing; $\phi = .39$, $p < .001$, between Question 2 and stimulus viewing; $\phi = .42$, $p < .001$, between Question 1 and Question 2. The strength of intermeasure relationships did not change systematically as a function of age group. Between Question 1 and stimulus viewing, the correlations were .27, .23, and .32 for the 3.00 to 4.25-year-old, the 4.50 to 5.75-year-old and the 6.00 to 7.00-year-old groups, respectively (all ps $>$ .05). Between Question 2 and stimulus viewing, correlations were .62 ($p < .001$), .15, and .44 ($p < .025$) for the three age groups respectively. The correlations between the two questions were .22 (ns), .56 ($p < .005$), and .47 ($p < .01$) for the three age groups respectively.

Experiment 3

The results of Experiment 2 strengthen the generality of the findings from Experiment 1 and are completely consistent with it. Regardless of age, preschool children tend to prefer

[5] As in Experiment 1, analyses of variance revealed no significant effects due to the food–nonfood variation in Experiment 2 or in Experiment 3.

view whichever stimuli are real, during delay of gratification. Before reaching any firm interpretations and conclusions, however, we conducted a third experiment designed to overcome some possible ambiguities. Namely, in the first two experiments, the viewing instructions to the children simply requested that they push buttons exposing stimuli during the delay period. Likewise, the postdelay questions may have been somewhat ambiguous. Can we be sure that the children's responses reflected attentional strategies that they believed would help them wait? Instead, might they have viewed whatever they found more attractive or desirable (i.e., real objects rather than pictures of them)? Might the children's viewing and verbal preferences for the real stimuli possibly have been the result of trying to "tell" the experimenter that they wanted to receive one of the sets of real objects after the delay, instead of receiving a picture of those objects? Conceivably, the children could have thought that the experimenter would somehow know which stimuli they viewed most during the delay. In that case, they might have conjectured that what they viewed during the delay would influence what the experimenter gave them when he or she returned. Likewise, the children may have self-presented real rather than symbolic stimuli so that when the returning experimenter opened the door to the room, he or she would see the children looking at the real stimuli and might be more likely to give them the real objects rather than pictures of them.

To identify the meaning of the children's responses more precisely, in Experiment 3 we explicitly asked half the subjects to view those stimuli that would best facilitate delay (efficacy strategy) and half to view those stimuli that they wanted to see (wish strategy). Moreover, we carefully arranged the situation and the postdelay questions to avoid alternative interpretations.

Method

In an EID situation, half of the children in each of the four stimulus-pair conditions were instructed to view those stimuli which, in their opinion, would best help them wait (efficacy strategy). The remaining children in each stimulus-pair condition were asked to view those stimuli which they wanted to see (wish strategy). Thus, we used a 2 (viewing instructions) × 4 (stimulus pairs) factorial design.

The same rewards and stimuli used in Experiment 1 were used in Experiment 3. As in the first two studies, rewards and relevancy of stimuli were completely crossed within the design. Because Experiment 1 found no significant difference on any measure between SID and EID, in Experiment 3 we again used only EID to provide maximum control over duration of delay.

Subjects

A total of 80 children (40 girls and 40 boys) who attended Bing Nursery School were the subjects. Four female experimenters tested equal proportions of the 5 girls and 5 boys who were randomly assigned to each cell in the design. The children ranged in age from 2.7 to 5.4 years, with a mean age of 4.3 years.

Procedure

The procedure used in the first two studies was modified as follows: The children were assured that the experimenter would not know anything about what they did during the delay. The children also were told that the experimenter would knock on the door three times before reentering the room and at that time they should stop looking at either stimulus set and, instead, should place their hands on top of their heads as soon as the experimenter knocked on the door. Furthermore, a tall (approximately 8 feet, or 2.44 m, high) divider was placed between the door and the table at which the children sat. The children were told that in case they forgot to put their hands on their heads when they heard the experimenter knock, the divider would prevent the experimenter from seeing which stimuli they might be looking at.

Just before leaving the room, the experimenter consulted a previously concealed portion of the schedule sheet to find out which predelay stimulus-viewing instructions the child was to receive. Half the children were given the *efficacy strategy* and were told the following:

> Push the button that shows you what *helps you wait* the most. . . . Look at the one (stimulus set) which makes it easiest for you to wait.

The other half were given the *wish strategy* and were told to:

> Push the button that shows you what you *like to see* the most while you're waiting. . . . Look at the one which you like to see the most.

Upon returning to the room after the delay, the experimenter gave the child the promised reward. She

Table 4
Stimulus Preferences in Each Pair on Each Measure (Experiment 3)

Stimulus pair	Question 1 (help)			Question 2 (liking)			Viewing time		
	Viewing strategy			Viewing strategy			Viewing strategy		
	Wish	Efficacy	Total	Wish	Efficacy	Total	Wish	Efficacy	Total
Relevant									
Real vs. **symbolic**	1.00***	.50	.75*	.80	.70	.75*	.67**	.50	.58
Irrelevant									
Real vs. symbolic	.70	.70	.70	.70	.90**	.80**	.56	.57	.56
Symbolic									
Relevant vs. irrelevant	.40	.40	.40	.70	.40	.55	.60	.54	.57
Real									
Relevant vs. **irrelevant**	.40	.90**	.65	.60	.70	.65	.47	.60	.53

Note. The data are proportions: The larger the proportion, the greater the preference for the top stimulus in each pair. $n = 10$. Stimuli that facilitate self-imposed delay (SID) are in boldface type.
* $p < .05$.
** $p < .025$.
*** $p < .002$.

then asked two questions. The first question of the previous experiment was revised to make it more direct:

> Please tell me which one (of the two stimulus sets) showed you what *helped you wait* the most. Which one made it easiest for you to wait?

The second question remained quite similar to Question 2 in Experiments 1 and 2:

> Please tell me which one showed you what you *liked to see* the most. Which one did you want to look at?

Both questions were asked of all children in a forced-choice format and in random order.

Results

Separate analyses of variance for the three measures yielded no significant main effects of predelay viewing instructions (all Fs < 1). The Instruction × Stimulus Pair interaction did not reach significance for time spent viewing stimuli, $F(3, 72) = 2.12$, $p > .10$, or for Question 2, $F(3, \infty) = 1.23$, $p > .20$. However, there was a significant interaction of Instructions × Stimulus Pairs on Question 1, $F(3, \infty) = 6.15$, $p < .001$.[6] As Table 4 indicates, this interaction reflects the fact that children who had been given a wish strategy overwhelmingly chose the relevant real (rather than relevant symbolic) stimuli while showing no strong preference for relevant real versus irrelevant real stimuli. In contrast, subjects given the efficacy strategy showed no systematic preference for relevant real versus relevant symbolic stimuli, but did choose relevant real over irrelevant real stimuli.

The overall data are summarized in Table 4. Let us consider each measure separately.

During the delay, children tended to actually view the relevant real more than the relevant symbolic stimuli when given a predelay wish strategy (i.e., when asked to self-present the stimuli they would like to see the most while waiting); $t(72) = 2.75$, $< .01$. But as Table 4 shows, children who

[6] Recall (Footnote 3) that since the theoretical error variance of transformed proportions was used in analyses of Questions 1 and 2, an F distribution with k and ∞ degrees of freedom was used for all significance tests.

had been asked to self-present the stimuli that would help them wait the most (efficacy strategy) showed no such systematic preferences; they viewed the relevant symbolic stimuli as much as they did the relevant real rewards.

On Question 2, the postdelay measure asking which stimuli they had liked to see most, children significantly preferred real stimuli to symbolic ones ($p < .05$ for both stimulus pairs). In addition, children who were given the efficacy viewing strategy showed a significant preference for irrelevant real over irrelevant symbolic stimuli when asked which they "liked" to view ($p < .025$), but not when asked which *helped* them wait best ($p > .10$). Finally, consider Question 1, the postdelay question asking children which stimuli had helped them wait the most. Here the answers again depended on the predelay viewing instructions (as the previously noted significant interaction indicated). Specifically, when children had been given the wish strategy instructing them to view the stimuli they liked to see the most, they consistently preferred relevant real over relevant symbolic stimuli ($p < .002$), paralleling the results found in Experiments 1 and 2. But when they had been instructed to view the stimuli that would help them wait most (efficacy strategy), they did not report any significant preference for relevant real over relevant symbolic stimuli, choosing each equally often.

Note also that on Question 1, children who had been given the predelay efficacy strategy preferred relevant real to irrelevant real stimuli. Since these children were waiting in an EID situation, their belief that attention to relevant real rewards would help them wait more than attention to irrelevant real stimuli may appear to support Miller and Karniol's (1976a, 1976b) view that while attention to the actual rewards impedes self-imposed delay, it facilitates externally imposed delay. Before reaching that conclusion, however, we must take into account some other considerations. First, Miller and Karniol's subjects were several years older than ours. Second, contradictory evidence comes from a fourth study that we conducted with subjects of comparable age to those used by Miller and Karniol, as reported next.

Experiment 4

In this final study, we examined the stimulus preferences of older (elementary school) children. We concentrated on their preference for relevant versus irrelevant real stimuli, since that is the most theoretically interesting choice, but continued to distinguish between EID and SID situations and between food and nonfood delay. We assumed that these older, more developmentally mature subjects would have a better awareness of the attentional strategies that would facilitate delay and that this awareness would be reflected in their stimulus preferences.

Method

In an EID or SID situation, half of the children waited for food rewards and half waited for nonfood rewards. Thus, we used a 2 (EID vs. SID) × 2 (food vs. nonfood rewards) design with 12 children in each cell. Rewards and relevancy of stimuli again were crossed completely in the design. The child's choice in each condition was between viewing the relevant real stimuli (the actual rewards in the delay contingency) or the irrelevant real stimuli (real stimuli irrelevant to the delay contingency).

Subjects

The data came from 48 children (24 girls and 24 boys) who attended Grades 1, 2, or 3 in one of two local elementary schools (Kirk and Valley View Elementary Schools in San Jose, California). One male experimenter conducted the predelay procedures; an experimentally blind assistant administered the postdelay questions about the delay-facilitating characteristics of stimuli. A male and two female adults served as assistants, for the same proportion of children in each cell.

Procedure

The same procedure was employed in Experiment 4 as in the other studies, with the following modifications made because of the age of this subject population. The rewards used in earlier studies were changed to be more appropriate for older children. Specifically, these children waited for either two versus six regular-size cat's-eye marbles or for one half a cookie versus three cookies (half a sugar cookie vs. a chocolate chip, an oatmeal, and a sugar cookie, each 3 inches, or 7.6 cm, in diam-

Table 5
Stimulus Preferences on Each Measure in Each Condition (Experiment 4)

Condition	Food		Nonfood	
	Question 1 (choice)	Viewing time	Question 1 (choice)	Viewing time
SID	.21**	.42	.71*	.73***
EID	.25**	.43	.79**	.69***
Total	.23***	.42*	.75***	.71***

Note. SID = self-imposed delay; EID = externally imposed delay. The larger the proportions, the greater the preference for relevant real versus irrelevant real stimuli. $n = 24$ for Question 1; $n = 12$ for viewing time.
* $p < .05$.
** $p < .025$.
*** $p < .002$.

eter). To collect more data on stimulus preferences during the delay, the apparatus was altered slightly so that children could alternate stimulus sets, determining how long each set was presented. The child chose which stimulus set was viewed, but always viewed one of the two stimulus sets. As in the previous studies, the children were assured that the experimenter would not know what they were doing during the delay, but no tall divider was used, and no hands-on-head instructions were given. As in Experiments 1 and 2, the children were not given specific viewing instructions. The delay period was increased to 20 minutes, in view of the children's age and to collect more extended data on stimulus viewing.

Verbal stimulus preferences (Question 1) were obtained after the delay by having the assistant ask the child which stimulus would be best to view for food delay and, separately, for nonfood delay, regardless of whether the child had experienced a food or nonfood delay. For subjects in the SID condition, this question was phrased as follows:

> Suppose that you would get these (food or nonfood rewards) if you waited for (experimenter's name) to come back on his own. Also, suppose that you would get (this) (these) if you pushed this (the termination) button and called (the experimenter) back. While you're waiting, would looking at what's under this (assistant points to left) box lid help you wait? Or, while you're waiting, would it help you wait more if you looked at what's under this (experimenter points to right) box lid? Looking at which would help you wait better?

In the EID condition, the question was rephrased to describe an EID rather than SID situation. Half of the children in each experimental cell were asked the question first for food delay, and half were asked the question first for nonfood delay.

Results

An analysis of variance for proportions of time spent viewing the alternative stimuli found no significant main effect for SID versus EID situations ($F < 1$) and no Delay Situation × Rewards interaction ($F < 1$). There was, however, a significant main effect for food versus nonfood rewards, $F(1, 44) = 28.85$, $p < .001$.

Verbal stimulus preferences also were not significantly affected by SID versus EID situations ($F < 1$ for all children's answers to the food delay question; $F < 1$ for answers to nonfood delay questions). The data for Question 1 in Table 5 show that these children said that they preferred viewing irrelevant real stimuli significantly more than relevant real stimuli during externally imposed, as well as during self-imposed, delays of food gratification (ps $< .025$, binomial tests). When SID and EID preferences are pooled, the preference for viewing irrelevant rather than relevant real stimuli is even more significant ($p < .001$). Consistent with these verbal preferences, the children spent significantly more time actually viewing the irrelevant stimuli during delay of food rewards, $t(44) = 2.11$, $p < .05$, two-tailed, for SID and EID pooled.

In contrast, these subjects significantly preferred relevant real over irrelevant real stimuli during delays of *non*food gratification, whether EID ($p < .025$ for verbal pref-

erence); $t(44) = 3.55$, $p < .002$, for viewing preference, or SID ($p < .05$) and $t(44) = 4.30$, $p < .002$. Possible reasons for this effect are explored in the discussion.

Further analyses examined the role of age in the children's viewing preferences. We divided the children into three age groupings: (a) less than or equal to 7 years ($n = 9$), (b) greater than 7 years but less than or equal to 8 years ($n = 17$), and (c) greater than 8 years ($n = 22$). Chi-square tests of the number of children in each age group who preferred to view irrelevant or relevant real rewards revealed that verbal stimulus preferences were not affected systematically by age, $\chi^2(2) = 1.99$, $p > .30$, for food delay; $\chi^2(2) < 1$ for nonfood delay. Moreover, the proportion of the delay spent viewing relevant real stimuli was not significantly different across age groupings. Specifically, in a three-way analysis of variance for viewing proportions, the only significant effect or interaction was the main effect for food versus nonfood delay, $F(1, 36) = 25.96$, $p < .001$; all other Fs < 1.5. Correlations between age and viewing preferences (verbal and nonverbal) were all nonsignificant and low for food and nonfood delay, both for EID and SID separately and pooled.

General Discussion and Conclusions

Taken as a whole, the results of the present series of experiments lead to several conclusions concerning the nature and consequences of young children's preferred attentional strategies during delay of gratification. To recapitulate briefly, Experiment 1 showed that during both self-imposed delay (SID) and externally imposed delay (EID), preschool children prefer to view real stimuli rather than symbolic (pictorial) representations of them, regardless of their relevance to the rewards in the delay contingency. No differences in preference patterns were found in SID versus EID. The same pattern of preferences was replicated in Experiment 2, which also found no significant effects of age within the three age categories sampled (ages 3 to 4.25 years, 4.50 to 5.75 years, and 6.00 to 7.00 years).

Experiment 3 allowed us to distinguish clearly between children's attentional preferences during delay when they are oriented to attend to the stimuli that they want to view (wish strategy) as opposed to those stimuli that they believe would most help them wait (efficacy strategy). The results revealed that when following a wish strategy, children attended significantly more to the real rewards than to their more abstract (pictorial) representations, further replicating the findings of Experiments 1 and 2. But when following an efficacy strategy, the children did not show any systematic preference for viewing relevant real as opposed to relevant symbolic stimuli: Presumably, they simply did not know which of these two types of stimuli would help more and chose quite randomly.

The young subjects of the first three experiments attended as much to relevant real as to irrelevant real stimuli during delays and preferred the two equally, again indicating a lack of awareness of delay-facilitating attentional strategies. This lack of discrimination was found consistently in all three experiments, and on all measures, with only one exception: Namely, on Question 1 in Experiment 3, children who had been given the predelay efficacy strategy significantly preferred relevant real to irrelevant real stimuli.

The data from the older children (Grades 1, 2, 3) obtained in Experiment 4 were dramatically different and suggest a change in delay strategies in the course of development, beginning around the age of 7 years. Specifically, when waiting for food, these older children systematically preferred viewing nonfood stimuli and were aware that this strategy would help them delay gratification best. This was true in externally imposed as well as in self-imposed delay situations. Thus, the older children, in sharp contrast to the younger ones in Experiments 1, 2, and 3, preferred to view the irrelevant stimuli during food delay. This finding supports the view that in the course of development, the child becomes aware of effective delay strategies and realizes that attention to the rele-

vant reward stimuli impedes effective delay in EID as well as in SID situations.

The fact that the older children also preferred to view the irrelevant stimuli during externally imposed food delay raises important theoretical questions about the attentional strategies that actually facilitate or impede such EID. A great deal of direct evidence supports the conclusion that self-imposed delay (SID) is impeded by attention to the relevant real rewards (e.g., Mischel, 1974; Mischel et al., 1972; Moore et al., 1976; Schack & Massari, 1973). In contrast, the conclusion of Miller and Karniol (1976a, 1976b) that externally imposed delay is helped by attention to the rewards was based entirely on indirect evidence, such as indications of differential frustration during reward absence versus reward presence in the course of EID, inferred from such data as subjective time estimates. The present data therefore raise the possibility that firm conclusions about the facilitating effects of attention to the rewards during EID may be premature. If an attentional strategy that focuses on the rewards really facilitates EID, it seems unlikely that the relatively mature subjects sampled in Experiment 4 would systematically tend to reject it during food delay, while younger subjects were more likely to prefer it (in Experiments 1, 2, and 3). Unfortunately, a direct test of the efficacy of different attentional strategies on actual duration of EID is not possible. By definition, EID requires that the subject cannot control the length of actual waiting time, thus preventing such a direct test of the effects of any stimulus manipulations upon duration of externally imposed delay.

So far we have discussed the preferred attentional strategy of the older children only during food delay. But recall that when waiting for nonfood rewards, the older children preferred to view the actual rewards in the delay contingency (i.e., the relevant real stimuli), during both EID and SID. This preference might seem puzzling and contradictory. However, it becomes understandable when one examines more closely the children's specific choice in this nonfood delay condition. Namely, in this condition of the cross-over design, food became the irrelevant stimulus. Thus, while waiting for a nonfood object, the subject's choice was either to view that nonfood stimulus or a food treat not relevant to the delay contingency. It seems likely that the older children realized that if they chose to view the food during the delay, they would be frustrating themselves (by arousing their desire for a consumable but unavailable object) more than they would if they viewed the relevant nonfood object. That is, while waiting for marbles, it may well be less frustrating to view those objects than to view cookies. In future research, tests of preference for relevant versus irrelevant stimuli probably should include both relevant and irrelevant food stimuli or both relevant and irrelevant nonfood stimuli, in order to avoid conflicts between the two different types of categories.

Taken as a whole, the present series of studies suggests a clear shift in preferred attentional strategies at the age of approximately 7 years. Specifically, the first three experiments unequivocally show that preschool children do not systematically prefer or use attentional strategies that have been found to optimize self-imposed delay of gratification (SID).[7] Indeed, the observed tendency of preschool children in Experiments 1, 2, and 3 to prefer attending to real stimuli (rather than to pictorial or abstract representations) may make it especially frustrative and difficult for them to engage in sustained voluntary waiting for preferred but delayed gratifications.

Our finding that the younger children (Experiments 1, 2, and 3) spontaneously prefer to view the real stimuli during delay of gratification seems to be congruent with

---

[7] Follow-up studies using simpler choices to explore knowledge of delay suggest that awareness of at least one basic delay rule—that it helps to cover rather than expose the rewards while waiting for them—may begin to emerge even earlier. In one study, most children knew that it would help to cover rather than expose the rewards by the time they were 5 years old (Mischel, Mischel, & Hood, Note 1). Even simpler or more ingenious methods might of course reveal important knowledge about self-control at still younger ages.

Freud's (1911/1959) classical theory of wish-fulfilling ideation during delay: When the desired object is blocked, the frustrated child tries to self-present it, to "have it." But our previous research suggests that in so doing, young children are making self-imposed delay even more difficult for themselves by increasing their desire and hence enhancing the frustrativeness of the delay (e.g., Mischel, 1974)—especially if this ideation is consummatory rather than more abstract (Mischel & Baker, 1975; Moore et al., 1976). The youngsters are then trapped in a delay-defeating cycle, attending to the consummatory qualities of what they really want and becoming increasingly frustratively aroused, thereby making it even harder to wait successfully. These interpretations are supported by the fact that, at least in the SID paradigm, it has been demonstrated repeatedly that attention to the real stimuli increases the frustrativeness of the delay and reduces the length of voluntary waiting (Mischel, 1974; Schack & Massari, 1973; Toner & Smith, 1977; also see Yates, 1977).

The young children's preferences for real stimuli over abstract or symbolic ones in the delay situation, and their inability to discriminate effective and ineffective delay strategies, probably reflect their cognitive-developmental immaturity. With greater cognitive development, the child comes to recognize and prefer attentional strategies that avoid frustrative arousal (Experiment 4). As children increase their ability to deal with stimuli more abstractly, they can transform them in delay-facilitating ways (Mischel, 1974; Moore et al., 1976). Specifically, we suggest that the older child can focus more on the abstract rather than consummatory qualities of incentives, thereby avoiding excessive frustration while remaining oriented to, and guided by, preferred but delayed goals.

Reference Note

1. Mischel, H. N., Mischel, W., & Hood, S. Q. *The development of knowledge about self-control.* Unpublished manuscript, Stanford University, 1978.

References

Freud, S. Formulations regarding the two principles in mental functioning. In *Collected papers* (Vol. 4). New York: Basic Books, 1959. (Originally published, 1911.)

Hays, W. L. *Statistics.* New York: Holt, Rinehart & Winston, 1963.

Langer, E. J., & Abelson, R. P. The semantics of asking a favor: How to succeed in getting help without really dying. *Journal of Personality and Social Psychology,* 1972, *24,* 26–32.

Miller, D. T., & Karniol, R. Coping strategies and attentional mechanisms in self-imposed and externally imposed delay situations. *Journal of Personality and Social Psychology,* 1976, *34,* 310–316. (a)

Miller, D. T., & Karniol, R. The role of rewards in externally and self-imposed delay of gratification. *Journal of Personality and Social Psychology,* 1976, *33,* 594–600. (b)

Mischel, W. *Personality and assessment.* New York: Wiley, 1968.

Mischel, W. Processes in delay of gratification. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 7). New York: Academic Press, 1974.

Mischel, W., & Baker, N. Cognitive appraisals and transformations in delay behavior. *Journal of Personality and Social Psychology,* 1975, *31,* 254–261.

Mischel, W., & Ebbesen, E. B. Attention in delay of gratification. *Journal of Personality and Social Psychology,* 1970, *16,* 329–337.

Mischel, W., Ebbesen, E. B., & Zeiss, A. R. Cognitive and attentional mechanisms in delay of gratification. *Journal of Personality and Social Psychology,* 1972, *21,* 204–218.

Mischel, W., & Moore, B. Effects of attention to symbolically presented rewards on self-control. *Journal of Personality and Social Psychology,* 1973, *28,* 172–179.

Moore, B. Cognition and delay of gratification (Doctoral dissertation, Stanford University, 1973). *Dissertation Abstracts International,* 1974, *34,* 4639-B. (University Microfilms No. 74-6517, 42)

Moore, B., Mischel, W., & Zeiss, A. R. Comparative effects of the reward stimulus and its cognitive representation in voluntary delay. *Journal of Personality and Social Psychology,* 1976, *34,* 419–424.

Paivio, A. *Imagery and verbal processes.* New York: Holt, Rinehart & Winston, 1971.

Piaget, J., & Inhelder, B. *L'image mentale chez l'enfant.* Paris: Presses Universitaires de France, 1966.

Schack, M. L., & Massari, D. J. Effects of temporal aids on delay of gratification. *Developmental Psychology,* 1973, *8,* 168–171.

Siegel, S. *Nonparametric statistics of the behavioral sciences*. New York: McGraw-Hill, 1956.

Toner, I. J., & Smith, R. A. Age and overt verbalization in delay-maintenance behavior in children. *Journal of Experimental Child Psychology*, 1977, *24*, 123–128.

Winer, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1971.

Yates, B. T. Children's awareness and use of attentional strategies for delaying gratification (Doctoral dissertation, Stanford University, 1976). *Dissertation Abstracts International*, 1977, *37*, 5414-B. (University Microfilms No. 77-7192, 99)

Received May 10, 1978

Ad Hoc Reviewers

The Editor acknowledges with thanks the advice and consultation provided by the following persons, who served recently as ad hoc reviewers of manuscripts. (The number of manuscripts reviewed is in parentheses, if more than one.)

John R. Anderson
Robert S. Baron
James Blascovich
Eugene Borgida
Clifford Bradley
Marilyn Brewer
Philip Brickman
Bruce Brown
John Cacioppo
J. Merrill Carlsmith
John S. Carroll
Richard Christie
Harris Cooper
James Dabbs
Robyn Dawes
Kay Deaux (2)
William deJong
Bela DePaulo
Anthony N. Doob (3)
Norman T. Feather (2)
Irene Frieze
David C. Glass
Alan Gross
David Hamilton
Victor A. Harris
Elaine Hatfield
Ernest Hilgard
Irving Janis
Robert Krauss
Robert Kraut
Jim Laird
Ellen Langer
Darwyn Linder
Melvin Manis
Hazel Markus
Charles McClintock
Norman Miller
Lawrence Pervin (2)
Dean Pruitt
John Pryor (3)
Judith Rodin
Robert Rosenthal
Michael Ross
Michael Saks
Michael Scheier (3)
Ronald Schlegel
John Schopler
Steven J. Sherman
Dean Keith Simonton
Jerome E. Singer
Eliot Smith
Mark Snyder
Yacov Trope
Bernard Weiner
Gary L. Wells
Paul Werner
Ladd Wheeler
Timothy Wilson

Journal of Personality and Social Psychology
1979, Vol. 37, No. 2, 301-313

Strategy Signals in Face-to-Face Interaction

Starkey Duncan, Jr., Lawrence J. Brunner, and Donald W. Fiske
University of Chicago

The notion of "strategy signal" is proposed as an element in the description of face-to-face interaction. Strategy signals are actions by a participant that have a distinct and consistent effect on specified subsequent actions by the partner, but that have been rejected as signals within the organization of interaction. Data are presented on three hypothesized strategy signals in eight two-person conversations. Both the speaker's head direction toward the auditor and the number of turn cues have the effect of increasing the auditor's attempts to take the turn under certain circumstances. Speaker smiling has the effect of increasing auditor smiles. The role of strategy signals within the broader description of face-to-face interaction is discussed.

We are proposing a new component of face-to-face interaction, the strategy signal: one or more actions by a participant that have a distinct and consistent effect on a specified subsequent action by the partner. Because the full definition of a strategy signal requires the prior notions of organization (or structure) and of strategy in interaction, as well as a certain approach to research method, we turn first to these matters.

Describing the organization of interaction is not unlike describing language in terms of grammar. A hypothesized organization for interaction would contain elements such as signals (composed variously of actions in language, paralanguage, body motion, and the like) and rules for appropriately using and responding to these signals. The organization as a whole provides a structure for the proper conduct of the interaction. A program of research in this laboratory (Brunner, in press; Duncan, 1972, 1974; Duncan & Fiske, 1977; Duncan & Niederehe, 1974) has attempted to describe some aspects of the organization of certain two-person conversations between adults. Because the research is concerned with the organization of interaction (as opposed to individual messages), analyses are based on sequences of action involving both participants.

A complete description of face-to-face interaction, however, would include consideration of interaction strategy in addition to organization (Duncan & Fiske, 1977). The notion of interaction strategy is a basically simple one. A close analogy is found in games such as chess or baseball. If the organization of face-to-face interaction is analogous to the rules of a game, then interaction strategy is analogous to ways of playing the game. For example, one rule in tennis is that the player must return the ball into the opponent's court, this rule leaving to the player's strategy the decision where to place the ball within that court. And within the rules for baseball, a batter may choose to swing or not at a given pitch. In this sense, our use of the term, though confined to describing face-to-face interaction, is congruent with one of its major meanings in common usage. Of course, the notion of strategy is common in social science.

This research was supported in part by National Science Foundation Grant SOC74-24084, National Institute of Mental Health Grant MH-30654, and Biomedical Research Support Grant PHS 5 SO7 RR-07029-12. The authors gratefully acknowledge the assistance of Barbara Kanki.

Requests for reprints should be addressed to Starkey Duncan, Jr., Committee on Cognition and Communication, Department of Behavioral Sciences, The University of Chicago, 5848 S. University Avenue, Chicago, Illinois 60637.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3702-0301$00.75

In general, interaction strategy arises whenever there are options attaching to the rule-based organization of interaction. We may think of two broad types of option. One type has to do with legitimate alternatives within the rules. In the organization hypothesized for aspects of some two-person conversations (Duncan & Fiske, 1977), an example of a legitimate option would be the auditor's choice to take the turn or not upon activation of the turn signal by the speaker. Each such choice would constitute an element of the auditor's strategy in the interaction.

Not all rules for interaction involve legitimate alternatives. For example, it would seem that if someone offers to shake hands with another, the recipient of that offer does not have a legitimate option to refuse, except under highly specialized circumstances (such as demonstrably dirty hands), for which apologies are given. In such cases where the rules require the taking or avoiding of certain actions, there remains the option simply to violate the rules. The refusal to shake hands without extenuating circumstances is presumably a rule violation; similarly, an interruption is a violation of the turn-signal rules. Thus, rule violation—or its absence—is another aspect of interaction strategy.

One component of strategy—rule violation—is always present in interaction, whereas the other—legitimate alternatives—is present only when it is provided for by the rules. Of course, rules providing for legitimate alternatives may be prevalent in many interactions. In any event, because of the presence of optionality, engaging in rule-based interaction requires a participant to choose a course of action continuously with respect to rule violation, and also at each point at which there are legitimate alternatives.

Organization, Strategy, and Action Sequences

To this point, distinctions have been made (a) between organization and strategy in interaction and (b) between rules that provide for legitimate alternatives and those that do not. These distinctions may be directly related to research on action sequences. Consider a sequence A–B, in which A is a action by one participant, and B is a sub sequent action by the partner. Let us assum that a sufficient number of A's and B's hav been observed in one or more face-to-fac interactions and that an A is sometimes, bu not always, followed by a B. In this case, i a relationship between A and B exists withi the organization of interaction, then B an not-B would have to be considered legitimat alternatives upon the occurrence of A. A example would be the hypothesized turn sig nal, which was sometimes, but not always followed by an auditor's attempt to take th turn.

Given these data, there are two types o questions we may ask of the relationship be tween A and B. First, given an occurrenc of A, what is the probability that it is fol lowed by B? We shall call this the prob ability of a consequent—certainly the most common perspective in studies of action sequences. But of course we may also ask the reverse: Given an occurrence of B, what is the probability that it is preceded by A? We shall call this the probability of an antecedent.

The distinction between the probability of a consequent and the probability of an antecedent provides us with an analytic framework for implementing the distinction between organization and strategy. When rules providing for legitimate alternatives are at issue (as they are in this example), evidence relating to the organization of interaction is presented in terms of probabilities of antecedents. That is, to support a hypothesized rule for the action sequence A–B in which B and not-B are considered legitimate alternatives, the data would optimally show that each B was preceded by an A; the probability of that antecedent is 1.00. As with the turn signal, the hypothesized rule would state that B may appropriately occur only after A, but need not so occur. The partner apparently exercises the choice of whether to do B upon the occurrence of the participant's A.

In contrast, evidence relating to interaction strategy is presented in terms of probabilities of consequents. That is, given the

choice to do B or not-B after the occurrence of A, some partners may do B more than others. For example, upon activation of the speaker turn signal, some auditors may attempt to take the turn (leading to smooth exchanges) more frequently than others. Thus, the probability of the consequent (turn attempt) for one auditor may be .10, whereas for another auditor it may be .40, even though each attempt by either auditor is preceded by a turn signal (probability of antecedent = 1.00).

This example illustrates how there may be strong individual differences in the probability of a consequent in action sequences when the probability of an antecedent is essentially uniform across participants. When results of this sort are obtained, we would interpret the uniform antecedent probabilities as reflecting the organization of the observed interactions; the variations in consequent probabilities would be interpreted as reflecting participants' differences in interaction strategy within that organization.

Given a hypothesized organization for observed interactions, it seems natural that some studies of interaction strategy would investigate whether some participants more than others tend to exercise certain legitimate alternatives (such as acting to take the turn in the presence of a turn signal) or tend to violate certain rules (such as interrupting).

When the rule expresses an obligatory relationship between A and B—not providing for legitimate alternatives—we would expect B to follow A in virtually 100% of the occurrences of A (with possible allowances for rule violation and noise in the data). In this case, the probability of a consequent and the probability of an antecedent would both be close to 1.00. The ensuing discussion will be concerned only with action sequences involving legitimate alternatives. Obligatory sequences will not be further considered.

Strategy Effect of Organization Signals

Consider the case in which optimal results are obtained for an organization relationship between two actions (all B's are preceded by A's), and the relationship involves legitimate alternatives (not all A's are followed by B's). In this case, a certain type of strategy effect will be found: The occurrence of an A will raise the probability of a B from zero to some higher number. For example, if virtually all auditor attempts to take the speaking turn occur when the speaker turn signal is displayed, then we shall also find for every auditor that the speaker's displaying the turn signal increases the probability that the auditor will attempt to take the turn, when compared to those points at which there is no turn signal.

Of course, even when this sort of built-in strategy effect is found, there may be wide individual differences between participants in the probability of a B following an A. That is, upon display of a turn signal, some auditors may attempt to take the turn more than others. Nevertheless, all auditors will take the turn more when the turn signal is displayed than when it is not.

Thus, when the legitimate alternatives of doing B or not-B are involved, a specific type of strategy effect may be expected for all organization signals: The probability of B will be increased following the signal. But it should be noted that this same effect may be observed for other actions that regularly precede B, but that, for any reason, are not included in the hypothesized organization signal. How might such actions be excluded from an organization signal?

Further Criteria for Organization

Evidence for a hypothesized organization signal may not rest entirely on the associated probability of an antecedent. Other criteria may be invoked, depending on the nature of the hypothesized relationship and on the results of analyses. In some cases, a criterion will require that an action, A, precede all B's, but not precede instances of some other event, C. For example, the turn signal was hypothesized to precede smooth exchanges of the turn, but not instances of simultaneous turns in the conversation. That is, it was hypothesized that auditor attempts to take the turn when the turn signal is active would result in smooth exchanges of the turn and

that auditor attempts when the turn signal is not active would result in instances of simultaneous turns. Similarly, the speaker-state signal (Duncan & Fiske, 1977; Duncan & Niederehe, 1974) was hypothesized to occur at the beginnings of turn attempts by the auditor, but not at the beginnings of auditor back channels (Yngve, 1970), such as head nods and shakes and vocalizations such as "yeah," and "m-hm." The speaker-state signal was hypothesized to differentiate turn beginnings from auditor back channels. Of course, this sort of differentiation will not be a criterion for every signal, but it is useful when available.

Parsimony may also be used as a criterion. For example, an action, X, may not be incorporated as part of a hypothesized organization signal, O, if O already precedes virtually all instances of B, and the addition of X would serve merely to increase the number of displays of O. When an organization signal is hypothesized to have multiple cues, results may permit parsimony in the rules defining signal display. For example, the turn signal was hypothesized to have six distinct cues (described in more detail in the Results section). Strong results were obtained when a signal display was defined as a display of at least one cue, as opposed to more complex displays involving some specified combination of cues or some larger number of cues.

Thus, hypotheses on the organization of interaction will typically include criteria in addition to the probability of an antecedent. For this reason, an action that regularly precedes some subsequent Action B may not become part of the organization signal for B. It immediately follows that, because of these additional criteria, some actions may show the strategy effect described above for organization signals but not be part of any hypothesized signal, having been rejected on the basis of these additional criteria.

Strategy Signals

We are now in a position to consider the subject of this article: strategy signals. An action, S, will be considered a strategy signal in some observed interaction when it does not meet criteria set for the relevant organization signal, but it does show a distinctive relation to the probability of a subsequent action B. It also seems reasonable to expect that it show the high degree of consistency over participants in the interactions that was required for organization signals.

Evidence will be presented on three hypothesized strategy signals in the interactions we have observed. Although preliminary analyses suggest that additional strategy signals may be hypothesized in these interactions, space limitations prohibit their full description. In any event, such a listing is unnecessary for the purpose of introducing the notion and a general research approach.

Method

Analyses presented below were based on data drawn from videotapes of eight two-person conversations between adults. All participants were aware of the video camera and had given permission for the taping. Conversations are identified by the same numbering system used in Duncan and Fiske (1977). Interaction 1 was a psychotherapy intake interview between a male therapist about 40 years old and a female client in her early 20s. Interaction 2, an informal consultation about a client, was between the same therapist and another therapist, also about 40 years old; the two were good friends. Interaction 3 was between two male college students, recruited as they emerged from another psychological experiment; they were previously unacquainted and were paid $2 for their participation. Interaction 4 was between two female graduate students who worked as research assistants in the same office. They were continuing their discussion of shopping. Interactions 5–8 were between unacquainted professional-school students taking part in a larger study of interaction. These participants were instructed to have a conversation, to get acquainted, or to talk about anything else that interested them. Each of these participants was paid $3. These four interactions were between a male and a female. Interactions 1 and 2 each lasted 30 to 40 minutes, of which the first 19 minutes were transcribed and analyzed. All of the videotaped interaction was transcribed and analyzed in the remaining interactions. Interactions 3 and 4 were about 5 minutes long; and Interactions 5–8 were about 7.5 minutes long. Thus, about 78 minutes of interaction were analyzed.

Individual participants in each interaction are identified in terms of their interaction number and their sex (M,F). When the two participants are of the same sex, "a" and "b" will be appended to

their identifications. Thus, the participants in Interaction 1 are 1F and 1M; the participants in Interaction 2 are 2Ma and 2Mb.

For purposes of analysis, each interaction was subdivided into a series of units, defined in terms of the actions of the speaker. Briefly, a boundary of a unit of analysis was drawn at the onset of the first syllable following the end of a phonemic clause (Trager & Smith, 1957), during or immediately after which at least one of the following actions occurred: (a) a deviation in the intonation from an essentially level pattern (22| in the Trager–Smith system); (b) the utterance of a sociocentric sequence (Bernstein, 1962) such as "you know"; (c) the completion of a grammatical clause, involving a subject–predicate combination; (d) paralinguistic drawl (Trager, 1958) on the final syllable or stressed syllable of a phonemic clause; (e) termination of a hand gesticulation or the relaxation of a tensed hand position (e.g., a fist); (f) a decrease in paralinguistic pitch and/or loudness, either across an entire phonemic clause or during its final syllable or syllables; (g) an audible inhalation by the speaker; (h) an unfilled pause (Maclay & Osgood, 1959); (i) a false start (Maclay & Osgood, 1959); (j) turning the speaker's head direction toward the partner; and (k) a relaxation of the foot or feet from a strong dorsal flexion. Counts of occurrences and nonoccurrences of actions in the transcription were made using these units of analysis.

The data analyzed for this article had been transcribed previously at several different times. No interjudge agreement was tested for these particular data, because earlier analyses had yielded highly satisfactory results. For example, 96% agreement (Cohen's kappa = .80) was found for transcribing head direction on a syllable-by-syllable basis (Duncan & Fiske, 1977, p. 342; more detailed descriptions of the conversations and procedures are given in that book).

Results

Speaker's Head Direction

The speaker turn signal is defined as the display of at least one of six cues: the first five actions listed above in the drawing of units of analysis, plus one additional cue—a decrease of paralinguistic loudness or pitch on a sociocentric sequence. In the observed conversations, the turn signal was said to operate as follows: If the auditor acted to take the turn when the signal was active, the speaker was obliged to yield the turn. However, the auditor was not obliged to take the turn each time the signal was active; taking the turn in the presence of the signal is considered to be an option that the auditor may exercise.

In addition, a speaker gesticulation signal was hypothesized. This signal was composed of a single cue: one or both of the speaker's hands being engaged in gesticulation or in a tensed hand position. The gesticulation signal serves to inhibit any turn signal concurrently being displayed. That is, when the gesticulation signal is displayed concurrently with the turn signal, the auditor's attempts to take the turn are strikingly reduced.

Two criteria were used for evaluating the effectiveness of a possible turn signal: (a) The signal must precede a maximum number of smooth exchanges of the speaking turn in the observed conversations; (b) the signal must differentiate smooth exchanges of the turn from instances of simultaneous claimings of the turn by the two participants. Ideally, the signal would precede all instances of smooth turn exchanges and no instances of simultaneous turns. Notice that the signal is defined as the display of any one of its constituent cues. Therefore, each of these cues when occurring alone is evaluated by these two criteria.

Speaker's head direction was tested as a possible turn-signal cue. The cue was considered to be activated when the speaker's head turned toward the auditor. The potential head-direction cue was considered to be active from the moment of the actual head turning until the third boundary of a unit of analysis following the head turn (provided the head was still toward the auditor). Head direction was used in this case because it was not always possible to describe on our videotapes the direction of the speaker's gaze. However, head direction is taken as a surrogate for gaze direction; results for head direction are understood as most likely reflecting the effect of gaze.

Speaker's head direction was rejected as a possible turn cue for two reasons. First, it was not necessary. Virtually all smooth exchanges (251 out of 254; 98.8%) were preceded by at least one of the six turn cues described above. Second, inclusion of head direction would have substantially impaired the ability of the turn signal to differentiate

Table 1
Auditor Attempts to Take the Turn in Response to Speaker Turn-Signal Displays With and Without Concurrent Speaker Head-Direction Cue

Participant	No head-direction cue		Head-direction cue present		
	n[a]	Proportion of attempts	n[a]	Proportion of attempts	Ratio of proportions
1F	60	.116	75	.360	3.10
1M	13	.000	87	.367	—
2Ma	58	.034	54	.148	4.35
2Mb	28	.035	41	.219	6.25
3Ma	11	.000	50	.240	—
3Mb	34	.058	51	.117	2.01
4Fa	11	.090	36	.222	2.46
4Fb	10	.100	30	.333	3.33
5F	24	.041	49	.265	6.46
5M	18	.055	55	.272	4.94
6F	17	.235	36	.527	2.24
6M	37	.108	71	.281	2.60
7F	62	.096	64	.265	2.76
7M	21	.095	52	.365	3.84
8F	6	.000	71	.281	—
8M	27	.111	49	.346	3.11
Total	437	.080	871	.289	3.61

Note. F = female; M = male; a and b = first and second members of a same-sex dyad.
[a] Number of units having a speaker turn-signal display, in the absence of a speaker gesticulation signal, and not followed by an auditor back channel; this number is the denominator for the proportion (e.g., for the first entry, first column, 7/60 = .116).

smooth exchanges from simultaneous turns. For example, in the eight conversations, there were 26 occurrences of simultaneous turns in which neither turn signal nor gesticulation signal was active. Nineteen of these occurrences (73%) were preceded by the potential head-direction cue. Thus, in the absence of the turn signal, the head-direction cue was associated with a large proportion of simultaneous turns in these data. When it occurred alone, the cue functioned quite poorly in differentiating smooth exchanges from simultaneous turns.

However, it is possible that the speaker's head direction has an effect on the auditor's subsequent response. Specifically, when the turn signal was active in the observed conversations, the auditor was more likely to attempt to take the turn when the speaker's head was toward the auditor than when it was not. (In these analyses, head direction was considered active for its full duration.) Table 1 shows data on the instances in which the turn signal was active and the gesticulation signal was not active. (The indicated participant is the speaker.) Rate of auditor attempts to take the turn when the head-direction cue was present is contrasted with that rate when the head-direction cue was not present. In this situation, 254 smooth exchanges and 33 simultaneous turns occurred. Each auditor showed a higher proportion of turn attempts when the speaker's head-direction cue was present. Summing across all participants, the auditor was about 3.6 times more likely to attempt to take the turn when the head-direction cue was present. In three cases, all of the auditor's turn attempts occurred when the head-direction cue was present. Although there was variation among the participants both in overall rate of turn attempts and in degree to which attempts were increased by the head-direction cue, the basic strategy-signal effect was present for each participant. (For all 16 participants showing an effect in the same

direction, $p < .0001$ in a two-tailed sign test.)

Speaker Smiling

Data on smiling were derived from the four conversations (2, 5, 6, 7) in which there was an appreciable number of smiles. In a recent study of these four conversations, Brunner (in press) found evidence that the beginnings of auditor smiles were distributed very much like auditor back channels (e.g., head nods and shakes and vocalizations such as "m-hm" and "yeah"). Like auditor back channels, auditor smile beginnings located at or just after the boundaries of units of analysis were systematically preceded by the speaker within-turn signal. Briefly, the latter signal is defined as the occurrence of at least one of two cues: completion of a grammatical clause by the speaker and the speaker's head facing in the auditor's direction.

In Duncan and Fiske (1977) this head-direction cue was considered to be deactivated three units of analysis after the speaker had turned his or her head toward the auditor. On the basis of Brunner's (in press) analyses of the original back-channel data, the head-direction cue was redefined as being present as long as the speaker's head was facing in the direction of the auditor. This new definition was also used for analyses of auditor smile beginnings.

We expected that within the organization of the observed conversations, speaker smiles would be part of a signal for auditor smile beginnings. However, the existing speaker within-turn signal was quite adequate in accounting for auditor smile beginnings: 39 of the 40 smile beginnings were preceded by the speaker within-turn signal. While the one remaining auditor smile was preceded by speaker smiling, there seemed little reason to redefine the signal on the basis of this single event. Further, including speaker smiling with various combinations of the two within-turn cues did not improve the statistical relationship between within-turn signal and subsequent auditor smile beginnings. Thus, it was concluded that speaker smiling was unnecessary to account for auditor smile beginnings in the observed conversations, and that speaker smiling did not add to the relationship between speaker signal and auditor smiling. Brunner's study led to the conclusions that auditor smiling was one form of auditor back channel and that the speaker within-turn signal marked the appropriate points for auditor smile beginnings located at or just after the boundaries of units of analysis, in the same manner that the signal functioned for the other, similarly placed auditor back channels previously studied.

The following hypothesis is advanced for speaker smiling as a strategy signal: When speaker smiling is present in addition to the speaker within-turn signal, the auditor is more likely to smile than when the speaker within-turn signal alone is present. That is, speaker smiling will increase the probability of a subsequent auditor smile beginning in response to the within-turn signal. Of course, auditor smiling is not mutually exclusive of other auditor back channels. The auditor might smile in conjunction with a head nod or "m-hm." The hypothesis merely states that there is an increased probability that smiling will be part of the response.

For each participant in the four conversations studied, Table 2 shows the percentage of displays followed by an auditor smile beginning (a) for within-turn-signal displays without smiling and (b) for within-turn-signal displays with concurrent speaker smiling. (In the table, the indicated participant is the speaker.) For each participant, the probability of auditor smiling is increased when speaker smiling is present ($p < .01$, two-tailed sign test).

Number of Turn Cues Displayed

Duncan and Fiske (1977) reported a positive, linear relationship between number of turn cues displayed conjointly by a speaker and the probability of a smooth exchange of the turn following the display. (This relationship was found only for turn-signal displays not accompanied by a gesticulation signal.) Two correlations between number of turn cues displayed and probability of smooth exchange were calculated, one pooling data

Table 2
Auditor Smiling in Response to Speaker Within-Turn-Signal Displays With and Without Concurrent Speaker Smiling

Participant	No speaker smiling		Speaker smiling		Ratio of proportions
	n[a]	Proportion of auditor smiling	n[a]	Proportion of auditor smiling	
2Ma	102	.009	1	1.000	—[b]
2Mb	273	.014	11	.090	6.42
5F	38	.053	8	.375	7.07
5M	34	.000	5	.600	—
6F	11	.090	20	.150	1.66
6M	16	.062	16	.250	4.03
7F	28	.178	19	.263	1.47
7M	19	.105	5	.400	3.80
Total	521	.030	85	.258	8.60

Note. M = male; F = female; a and b = first and second members of a same-sex dyad.
[a] Number of units having speaker within-turn-signal display and not followed by an auditor attempt to take the turn; denominator for proportion.
[b] Insufficient data.

from Conversations 1 and 2 and the other pooling data from Conversations 3 through 8. The correlations were .99 and .98, respectively. These pooled data suggest that the probability of a smooth exchange increases as the number of turn cues displayed increases. This relationship stems from interaction strategy, as it is based on the probability of a consequent: the probability of a smooth exchange following a cue display.

Table 3
Rate of Smooth Turn Exchanges as a Function of Number of Active Turn Cues (Gesticulation Signal Absent)

Participant	1 or 2 active turn cues		3 to 6 active turn cues		Ratio of proportions
	n	Proportion of smooth exchanges	n	Proportion of smooth exchanges	
1F	75	.160	58	.344	2.15
1M	70	.200	27	.555	2.77
2Ma	80	.050	31	.161	3.22
2Mb	40	.075	29	.241	3.21
3Ma	46	.152	14	.285	1.87
3Mb	68	.073	15	.066	.90
4Fa	30	.100	16	.312	3.12
4Fb	27	.222	11	.272	1.22
5F	61	.163	11	.272	1.66
5M	54	.111	14	.357	3.21
6F	32	.312	20	.600	1.92
6M	77	.220	30	.200	.90
7F	99	.161	25	.200	1.24
7M	62	.225	7	.428	1.90
8F	60	.150	13	.538	3.58
8M	64	.203	9	.444	2.18
Total	945	.157	330	.318	2.02

Note. F = female; M = male; a and b = first and second members of a same-sex dyad.

Previous analyses had led to the rejection f number of turn cues displayed as part of ie definition of the turn signal within the rganization of the observed conversations. ince the display of any single cue was found o differentiate between smooth exchanges nd simultaneous talking, it was considered o constitute a display of the signal as a hole. In addition, it will be apparent that he reported correlations and the results be- ow are not simply a product of the strategy ffect of the turn signal. That strategy effect ould be to raise the probability of a smooth xchange from near zero following no turn ignal to some higher number following a ignal display.

Table 3 shows data for each participant in the eight conversations studied. Data for individual participants are insufficient to support analysis of each of the six data points (one through six cues). Therefore, two data points were calculated for each participant: the probability of a smooth exchange following (a) the display of one or two cues and (b) the display of three, four, five, or six cues. The data were divided between two and three cues because this division would permit sufficient observations in the two categories for each of the participants.

Table 3 shows that for 14 of the 16 participants, the probability of a smooth exchange following the display of one or two cues was lower than the probability of a smooth exchange following the display of three, four, five, or six cues. This result is significant at the .004 level (two-tailed test). These data disregard the presence or absence of the head-direction signal. When the data are analyzed only for instances in which the head-direction signal is present, the same results are obtained: 14 of the 16 participants showed the strategy-signal effect.

Discussion

The purposes of this article are to propose the strategy signal as a phenomenon of interest in face-to-face interaction, to suggest an approach to providing evidence supporting hypotheses of such phenomena, and to present evidence on three hypothesized strategy signals operating in some transcribed interactions.

We have suggested four criteria for hypothesizing an action to be a strategy signal: (a) that the strategy-signal action not meet criteria for hypothesizing it as an element in the relevant part of the organization of the interaction(s) studied; (b) that the action have an effect on the probability of occurrence of a subsequent action; (c) that this effect be highly consistent across the participants for which it is hypothesized; and (d) that the effect apply to action sequences involving legitimate alternatives—that the signal not be invariably followed by the subsequent action in question. By "effect" we mean a shift in the probability that the subsequent action will occur. We do not specify the exact nature or direction of this shift. Thus, it need not be of the kind described in the Results section.

Strategy Signals and Organization Signals

The strategy-signal criteria rely on the parallel distinctions between organization and strategy in interaction and between the probability of an antecedent and the probability of a consequent in analyses of action sequences. In general, we have identified the probability of an antecedent (the probability that A will precede B) with information on the organization or structure of interaction. This is because organization has to do with regularities in the placement of actions in interaction, given that they occur. More broadly, organization has to do with the appropriate placement of actions in the stream of interaction.

We have identified the probability of a consequent (the probability of B following A) with information on the strategy used in the interaction. This is because strategy is concerned with patterns of influence (or patterns of reaction) in interaction: one or more actions changing the probability that a subsequent action will occur.

Of course, hypothesizing A as an organization signal for B will almost certainly involve criteria in addition to the finding that A consistently precedes B. Investigators will

attempt to formulate the simplest and most effective definition of the hypothesized Signal A. And in some cases, the ability of the signal to differentiate different classes of action will be used as a criterion. An example would be the differentiating of smooth exchanges from simultaneous turns by the turn signal.

Of the four criteria proposed for hypothesizing a strategy signal, the last three seem relatively straightforward, whereas the first —that the action be excluded from the relevant part of the interaction organization— may bear further discussion. Strategy signals are seen as supplementing organization signals in describing interactions. If an action is an organization signal, we gain nothing by also labeling it as a strategy signal. Hence, a careful formulation of organization should precede the investigation of strategy signals. Furthermore, each organization signal has its own strategy effect. In the optimal case, when a given organization signal is not displayed, the probability of the consequent action associated with it is zero; but when the organization signal is displayed, the probability is some value above zero. This strategy effect of the organization signal stems from the fact that in the optimal case, the signal precedes all instances of the consequent action associated with it. That is, given the consequent action, the probability of the antecedent signal is 1.00. If an investigator finds, for example, that the occurrence of Action A raises the probability of a consequent Action B, A may be an organization signal or a strategy signal. Action A should be hypothesized as a strategy signal for B only after it has been rejected as part of an organization signal for B. This active rejection was part of the evaluation of each of the strategy signals hypothesized in the Results section.

This discussion indicates that the occurrence of a legitimate alternative, B, may be affected both by a preceding organization signal and by a preceding strategy signal. The hypothesis of a strategy signal enables the investigator to account for more of the variance in the occurrence of B than does an organization signal alone. However, as a legitimate alternative, the occurrence of B can never be entirely predicted on the basis of preceding actions. The partner is not in any sense "caused" to do B on the basis of preceding actions. Rather, in choosing to do B, the partner may be influenced to a greater or lesser extent by the participant's preceding organization and strategy signals.

Speaker's Head-Direction Signal

In the case of head direction and the turn signal, it is clear that the speaker's head being turned toward the auditor preceded many auditor attempts to take the turn; otherwise, the strategy-signal results (cast in terms of consequents) could not have been obtained. It was in differentiating smooth exchanges from simultaneous turns that head direction was markedly worse than the proposed turn-signal cues. Thus, even though the probability that the speaker's head direction precedes the auditor's turn attempt may be relatively high for many participants, head direction fails to meet the criteria set for turn-signal cues. The finding of a strong and consistent strategy effect, in the absence of the required organization effect, led us to propose a new interaction component: the strategy signal.

Despite the extensive literature on gaze direction (see Davis, 1972; Key, 1977; and Argyle & Cook, 1976, for references), there has not been intensive study of the relation of gaze direction to turn taking. Relevant exceptions would be an early study by Kendon (1967) and two more recent studies further exploring the phenomena Kendon described (Beattie, 1978; Rutter, Stephenson, Ayling, & White, 1978; see also an accompanying commentary by Kendon, 1978). Unfortunately, none of these studies provides data on the effect reported here: given a turn signal, an increased probability that the auditor will attempt to take the turn when the speaker is gazing. These investigations studied other parameters, such as the effect of speaker gaze on length of the pause between turns and the proportion of turn exchanges in which the speaker or auditor is looking or not looking.

The evidence presented here focuses on head direction (presumably gaze) with respect to a specific phenomenon: the exchange of speaking turns. Ellsworth and Langer (1976) take a broader perspective, suggesting that staring "elicits attention, arousal, and a sense of interpersonal involvement. The type of involvement inferred and the response perceived to be appropriate depend upon contextual cues" (p. 122). Although we are not certain that Ellsworth and Langer would consider staring to include the intermittent gazing typically found in conversations, we believe that the results presented here are entirely consistent with their hypothesis. The increased arousal resulting from speaker gaze may have prompted in part the higher rates of turn attempts by the auditor. From the perspective of Ellsworth and Langer, our results bear on some of the "contextual cues" relevant to the appropriate interpretation of gaze in the observed conversations.

Speaker Smiling Signal

As for smiling, the claim that speaker smiling increases the probability of auditor smiling is perhaps less surprising than the claim that speaker smiling is not part of an organization signal for auditor smiling in the observed conversations. Yet, in our conversations, it was true both that some speaker smiles were not reciprocated by the auditor and that some auditor smiles were not preceded by speaker smiles. (For an interesting account of auditor failure to reciprocate speaker smiling, see Rosenfeld, 1972.) Thus, the relationship between auditor and speaker smiling is not necessarily a simple one. In our conversations, the proposal to include auditor smiling within the more general auditor back-channel class of actions seems intuitively acceptable, is supported by a series of data analyses (Brunner, in press), and has the advantage of extending the province of the speaker within-turn signal (Duncan, 1974; Duncan & Fiske, 1977), while avoiding the necessity of hypothesizing an additional speaker organization signal.

Several studies have reported correlations between interacting subjects in the number or rate of smiles (Duncan & Fiske, 1977; Rosenfeld, 1966, 1967). Such correlations provide indirect support for the results reported here. But because data based on frequencies or rates do not contain information on the location of smiling in the interactions, they do not allow us to evaluate the degree to which smiles by a participant were actually reciprocated by the partner.

Number of Speaker Turn Cues as a Signal

As a strategy signal, number of turn cues is somewhat different from the two just discussed. Both head direction and smiling are actions that frequently precede their respective subsequent actions. In contrast, number of turn cues employs the same actions as the turn signal, but differs in the rule describing the way in which turn cues are combined. Effective results were obtained for the turn signal when number of turn cues displayed was ignored. The occurrence of any single cue constituted a full organization signal display. Number of turn cues displayed was considered and rejected as a possible element of the turn-signal definition. Analyses suggested that number was quite irrelevant to the criteria of (a) preceding smooth exchanges and (b) differentiating smooth exchanges from instances of simultaneous turns.

However, it seems additionally true for the conversations we observed that the probability of a smooth exchange increases in some manner as the number of turn cues increases. More specifically, for most participants the probability of smooth exchanges after the display of three, four, five, or six turn cues is greater than after the display of one or two turn cues.

Individual Differences

As mentioned in the Results section, Tables 1, 2, and 3 provide data on two types of strategy effects. On the one hand, there is evidence of consistency: Each speaker's head direction or smiling may be seen to change in the same direction the probability

of the relevant response by each auditor. This is the strategy-signal effect. On the other hand, variation between partners is evident in rates of response and in the degree to which that response is increased following the hypothesized strategy signal. This variation reflects the individual differences one may expect to find in interaction strategy, an area of research that extends well beyond strategy signals.

Generality of Results

How general are these strategy-signal results? We hold that generality in research of this sort is entirely an empirical matter, not a logical one. The organization and strategy signals discussed in this article are elements of social practices or conventions hypothesized to be operating in the observed interactions. The cultural boundaries of the use of these practices is an issue that must be determined empirically. A strict analogy would be work on language dialects. An investigator may interview one or more informants, gathering sufficient information to describe their dialect, but only dialect survey research can establish the geographical and demographic area in which the dialect is spoken. In dealing with social practices, we do not believe that it is appropriate to generalize to a postulated a priori domain, as is commonly done in some other types of behavioral research.

Nevertheless, we may cite circumstantial evidence that the strategy effects we report are not entirely confined to the conversations studied. With the exception of Conversations 2 and 4, the participants were unacquainted with each other prior to their interaction. The participants came from a rather wide variety of social and educational backgrounds and geographical areas. Although the settings were not completely natural, they were naturalistic and, hence, seem fairly representative of everyday interactions. In terms of the three dimensions of naturalness identified by Tunnell (1977), the behavior was natural and the treatment was also natural, because the stimuli for each participant were provided by the partner, and there were no experimental manipulations. For these reasons, findings from our studies may generalize more readily than results from contrived laboratory investigations or from responses in testing rooms.

Regardless of the domain to which a set of results applies, in the process of face-to-face interaction, a hypothesized organization implies strategy, and vice versa. Descriptions of strategy must be coordinated with descriptions of organization. Research on interaction requires careful examination of each side of an action sequence.

References

Argyle, M., & Cook, M. *Gaze and mutual gaze.* Cambridge, England: Cambridge University Press, 1976.

Beattie, G. W. Floor apportionment and gaze in conversational dyads. *British Journal of Social and Clinical Psychology,* 1978, *17,* 7–15.

Bernstein, B. Social class, linguistic codes, and grammatical elements. *Language and Speech,* 1962, *5,* 221–240.

Brunner, L. J. Smiles can be back channels. *Journal of Personality and Social Psychology,* in press.

Davis, M. *Understanding body movement.* New York: Arno Press, 1972.

Duncan, S. D., Jr. Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology,* 1972, *23,* 283–292.

Duncan, S. D., Jr. On the structure of speaker-auditor interaction during speaking turns. *Language in Society,* 1974, *2,* 161–180.

Duncan, S. D., Jr., & Fiske, D. W. *Face-to-face interaction: Research, methods, and theory.* Hillsdale, N.J.: Erlbaum, 1977.

Duncan, S. D., Jr., & Niederehe, G. On signalling that it's your turn to speak. *Journal of Experimental Social Psychology,* 1974, *10,* 234–247.

Ellsworth, P. C., & Langer, E. J. Staring and approach: An interpretation of the stare as a nonspecific activator. *Journal of Personality and Social Psychology,* 1976, *33,* 117–122.

Kendon, A. Some functions of gaze-direction in social interaction. *Acta Psychologica,* 1967, *26,* 22–63.

Kendon, A. Looking in conversation and the regulation of turns at talk: A comment on the papers of G. Beattie and D. R. Rutter et al. *British Journal of Social and Clinical Psychology,* 1978, *17,* 23–24.

Key, M. R. *Nonverbal communication: A research guide and bibliography.* Metuchen, N.J.: Scarecrow, 1977.

Maclay, H., & Osgood, C. E. Hesitation phenomena in spontaneous English speech. *Word*, 1959, *19*, 19–44.

Rosenfeld, H. M. Approval-seeking and approval-inducing functions of verbal and nonverbal responses in the dyad. *Journal of Personality and Social Psychology*, 1966, *4*, 597–605.

Rosenfeld, H. M. Nonverbal reciprocation of approval: An experimental analysis. *Journal of Experimental Social Psychology*, 1967, *3*, 102–111.

Rosenfeld, H. M. The experimental analysis of interpersonal influence processes. *Journal of Communication*, 1972, *22*, 424–442.

Rutter, D. R., Stephenson, G. M., Ayling, K., & White, P. A. The timing of looks in dyadic conversation. *British Journal of Social and Clinical Psychology*, 1978, *17*, 17–21.

Trager, G. L. Paralanguage: A first approximation. *Studies in Linguistics*, 1958, *13*, 1–12.

Trager, G. L., & Smith, H. L., Jr. *An outline of English structure*. Washington, D.C.: American Council of Learned Societies, 1957.

Tunnell, G. B. Three dimensions of naturalness: An expanded definition of field research. *Psychological Bulletin*, 1977, *84*, 426–437.

Yngve, V. H. On getting a word in edgewise. *Papers from the sixth regional meeting, Chicago Linguistic Society*. Chicago: Chicago Linguistic Society, 1970.

Received May 15, 1978 ■

Editors Named for New *Journal of Personality and Social Psychology*

The Publications and Communications Board has named Melvin Manis, Ivan D. Steiner, and Robert Hogan to edit the *Journal of Personality and Social Psychology* (JPSP), which will appear in a new format in January 1980. The affiliations of the new editors are, respectively, University of Michigan and Veteran's Hospital, Ann Arbor; University of Massachusetts; and Johns Hopkins University. Manis will edit the section on ATTITUDES AND SOCIAL COGNITION, Steiner the section on INTERPERSONAL RELATIONS AND GROUP PROCESSES, and Hogan the section on PERSONALITY PROCESSES AND INDIVIDUAL DIFFERENCES.

An announcement on the restructuring of JPSP appeared on p. 668 of the June 1978 issue of JPSP. As indicated at that time, the three editors will work independently with separate groups of advisory editors. Their acceptances will appear as distinct sections in monthly issues of JPSP. Detailed descriptions of each section's coverage will be published in JPSP next month.

Manuscripts for JPSP are currently being received by Acting Editor Clyde Hendrick through March 31, 1979. As of April 1, 1979, authors should submit manuscripts to the appropriate section editor as follows:

ATTITUDES AND SOCIAL COGNITION

April 1 through June 15:

Melvin Manis
c/o Anita DeVivo
Suite 700
1400 North Uhle Street
Arlington, Virginia 22201

After June 15:

Melvin Manis
Research Center for Group Dynamics
Institute for Social Research
University of Michigan
Ann Arbor, Michigan 48104

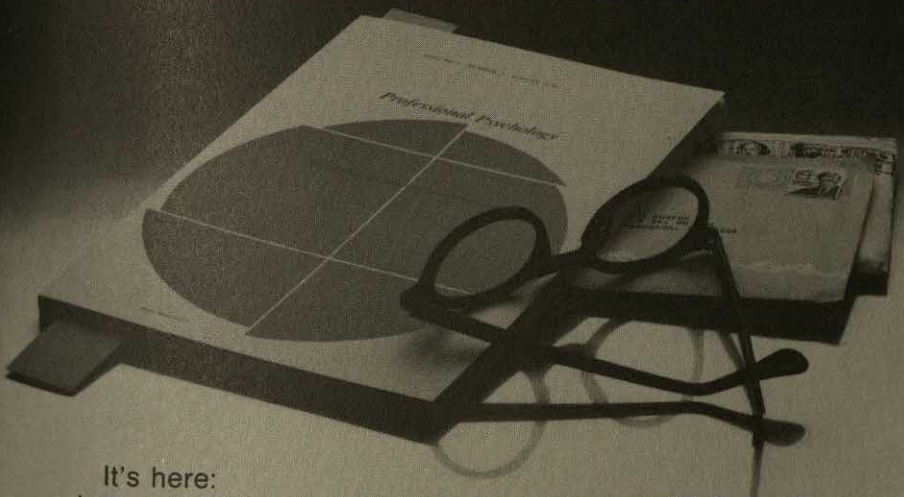
INTERPERSONAL RELATIONS AND GROUP PROCESSES

Ivan D. Steiner
Department of Psychology
University of Massachusetts
Amherst, Massachusetts 01003

PERSONALITY PROCESSES AND INDIVIDUAL DIFFERENCES

Robert Hogan
Johns Hopkins University
Department of Psychology
33rd and Charles Streets
Baltimore, Maryland 21218

NOW: INFORMATION ABOUT THE LAW AND PSYCHOLOGY.



It's here:
Law and Professional Psychology. This special August 1978 issue of **Professional Psychology** brings you thirteen original articles full of fresh thought and constructive recommendations on the issues that vitally affect your day-to-day practice. You'll find precautions you may take to protect yourself as a professional and a plan for improving investigations of malpractice claims. Expert witness testimony, jury selection, civil commitment, confidentiality and privilege, minor's consent to treatment, and the use of psychological devices are critically examined.

Send your order today. And soon you'll receive this special issue.

Single copies of the issue are available for $5 each. All orders $25 or less must be prepaid. Make and send checks payable to:

American Psychological Association
Order Department
1200 17th Street, N.W.
Washington, D.C. 20036

Enclosed is $_____for__copies of **Professional Psychology's** special issue—Law and Professional Psychology—at $5.00 each.

NAME _____

ADDRESS _____

CITY _____

STATE _____ ZIP CODE _____

PP-8

E 11
79

Worchel & Cooper

Understanding Social Psychology

Revised Edition

Just Published

Stephen Worchel
University of Virginia

Joel Cooper
Princeton University

The **Revised Edition of Understanding Social Psychology** covers the traditional areas of social psychology—both theory and research. The authors show how social psychology can and cannot be utilized to better understand real events. They begin with a discussion of the individual—how attitudes are formed and socialization. They then focus on dyadic interaction—attribution, liking and love, helping, and aggression. Finally they discuss the individual in the group—conflict, conformity, leadership, and group effects on performance. The last chapter covers environmental psychology.

New to this edition is a glossary. A list of suggested readings is included at the end of the chapters. A **Student Study Guide (new for this edition)** by **Carl I. Greenberg,** University of Nebraska at Omaha is a supplemental learning device which provides an integrated summary.

Liebert & Spiegler

Personality: Strategies & Issues, Third Edition

Published in 1978

Robert M. Liebert
SUNY at Stony Brook

Michael D. Spiegler
Providence College

This volume provides a widely respected and highly readable introduction to the major strategies and issues in the psychological study of personality. The material is organized so that students can see clearly both the strengths and weaknesses of each viewpoint discussed. The four basic strategies covered are: 1) psychoanalytic, 2) dispositional, 3) phenomenological, and 4) behavioral.

Examination copies for adoption consideration available on request; please indicate course title and text presently used.

THE DORSEY PRESS • **Homewood, Illinois 6043**

JOURNAL OF

Personality AND Social Psychology

VOLUME 37 NUMBER 3 · MARCH 1979

Attitudes Cause Behaviors: A Cross-Lagged Panel Analysis 315
Lynn R. Kahle and John J. Berman

A cross-lagged panel correlation analysis of four issues shows that attitudes have causal predominance over behaviors.

Egocentric Biases in Availability and Attribution 322
Michael Ross and Fiore Sicoly

Subjects remembered best their own contributions to group products and took more credit for outcomes than other participants accorded them.

Ambient Temperature and the Occurrence of Collective Violence: A New Analysis 337
J. Merrill Carlsmith and Craig A. Anderson

The conditional probability of a riot increases with maximum ambient temperature, a result that conflicts with previously published conclusions.

Affective Space is Bipolar 345
James A. Russell

The Generalizability of Salience Effects 357
Shelley E. Taylor, Jennifer Crocker, Susan T. Fiske, Merle Sprinzen, and Joachim D. Winkler

Interpersonal Consequences of Person Perception Processes in Two Social Contexts 369
Lawrence A. Messé, Gary E. Stollak, Rita W. Larson, and Gerald Y. Michaels

Affective States, Expressive Behavior, and Learning in Children 380
John C. Masters, R. Christopher Barden, and Martin E. Ford

Stimulus Recognition May Mediate Exposure Effects 391
Michael H. Birnbaum and Barbara A. Mellers

A Psychological Taxonomy of Trait-Descriptive Terms: The Interpersonal Domain 395
Jerry S. Wiggins

Self-Awareness, Psychological Perspective, and Self-Reinforcement in Relation to Personal and Social Standards 413
Ed Diener and Thomas K. Srull

Inferior Performance as a Selective Response to Expectancy: Taking a Dive to Make a Point 424
Roy F. Baumeister, Joel Cooper, and Bryan A. Skib

(Continued on inside back cover)

Published monthly by the

AMERICAN PSYCHOLOGICAL ASSOCIATION

ISSN 0022-3514

Editor

Anthony G. Greenwald, *Ohio State University*

Topic Editors

Irwin Altman, *University of Utah*
Robert A. Baron, *Purdue University*
Daryl Bem, *Stanford University*
Eugene Burnstein, *University of Michigan*
Robert B. Cialdini, *Arizona State University*
Reid Hastie, *Harvard University*
Douglas N. Jackson, *University of Western Ontario*
Leslie Z. McArthur, *Brandeis University*

Consulting Editors

Icek Ajzen, *University of Massachusetts—Amherst*
Robert S. Baron, *University of Iowa*
Gordon Bear, *Ramapo College of New Jersey*
Ellen Berscheid, *University of Minnesota*
Michael Birnbaum, *University of Illinois at Urbana-Champaign*
Philip Brickman, *Institute for Social Research*
J. Merrill Carlsmith, *Stanford University*
John S. Carroll, *Loyola University of Chicago*
Jerry L. Cohen, *University of Illinois at Urbana-Champaign*
Sheldon Cohen, *University of Oregon*
Rick Crandall, *Texas Christian University*
William D. Crano, *Michigan State University*
Robyn M. Dawes, *University of Oregon*
Kay K. Deaux, *Purdue University*
Marshall Dermer, *University of Wisconsin—Milwaukee*
Edward Diener, *University of Illinois at Urbana-Champaign*
Anthony Doob, *University of Toronto, Toronto, Canada*
Susan V. Eisen, *McLean Hospital, Belmont, Massachusetts*
Baruch Fischhoff, *Decision Research, Eugene, Oregon*
Irene H. Frieze, *University of Pittsburgh*
David Funder, *Stanford University*
David L. Hamilton, *University of California, Santa Barbara*
Ranald D. Hansen, *Oakland University*
Richard J. Harris, *University of New Mexico*
Edward Helmes, *University of Western Ontario, London, Canada*
Donald S. Hiroto, *VA Hospital, Brentwood, Los Angeles, CA*
Edward E. Jones, *Princeton University*
Martin F. Kaplan, *Northern Illinois University*
Douglas T. Kenrick, *Montana State University*
Norbert L. Kerr, *University of California, San Diego*
Samuel Komorita, *University of Illinois at Urbana-Champaign*
Vladimir J. Konečni, *University of California, San Diego*
Robert M. Krauss, *Columbia University*
Ellen J. Langer, *Harvard University*
Mark R. Lepper, *Stanford University*
Melvin Manis, *University of Michigan, (Ann Arbor)*
Hazel Markus, *University of Michigan (Ann Arbor)*
Margaret Matlin, *SUNY College at Geneseo*
Dale T. Miller, *University of British Columbia, Vancouver*
Thomas M. Ostrom, *Ohio State University*
Paul B. Paulus, *University of Texas at Arlington*
Richard Petty, *University of Missouri—Columbia*
Harry T. Reis, *University of Rochester*
Judith Rodin, *Yale University*
Seymour Rosenberg, *Livingston College, Rutgers*
Ralph L. Rosnow, *Temple University*
Lee Ross, *Stanford University*
Michael Ross, *University of Waterloo, Waterloo, Canada*
Diane Ruble, *Princeton University*
Michael J. Saks, *Nat'l Center for State Courts, Williamsburg, VA*
Michael F. Scheier, *Carnegie-Mellon University*
Barry R. Schlenker, *University of Florida*
Philip Shaver, *New York University*
Steven J. Sherman, *Indiana University*
Jerome E. Singer, *University of the Health Services, Bethesda, Maryland*
Eliot R. Smith, *University of California, Riverside*
Mark Snyder, *University of Minnesota*
Melvin L. Snyder, *University of Texas at Austin*
Lawrence Stricker, *Educational Testing Service*
Shelley E. Taylor, *Harvard University*
Abraham Tesser, *University of Georgia*
Bernard Weiner, *University of California, Los Angeles*
Gary L. Wells, *University of Alberta, Edmonton, Canada*
Stephen Worchel, *University of Virginia*
Robert S. Wyer, *University of Illinois at Urbana-Champaign*
Mark Zanna, *University of Waterloo, Waterloo, Canada*
Miron Zuckerman, *University of Rochester*

Assistant to the Editor
Holly Wong

APA Journal Staff

Anita DeVivo, *Executive Editor*
Ann I. Mahoney, *Manager, Journal Production*
Barbara R. Richman, *Production Supervisor*
Marjorie M. Henry, *Production Editor*
Robert J. Hayward, *Advertising Representative*
Juanita Brodie, *Subscription Manager*

The Journal of Personality and Social Psychology publishes original reports of research, methodology, theory, criticism, and literature review in the fields of personality and social psychology of normal processes and populations. Submitted manuscripts should delineate an important question and provide a clear answer to it. In the case of research reports, this will often require demonstration that the reported results are generalizable beyond a single research setting. Consult the editorial in the January 1976 issue (Vol. 33, pp. 1–7) for a detailed statement of policies and instructions to authors.

Manuscripts: Submit manuscripts in quadruplicate, prepared according to the *Publication Manual of the American Psychological Association* (2nd ed.). Copies should be highly legible and will not be returned to authors unless they are original typewritten copies. Mail manuscripts according to instructions on pp. 468–469.

Required supplementary information: Manuscripts cannot be reviewed for publication unless the author provides supplementary information in the following categories: (a) a cover letter that indicates the relation (or the lack of relation) of the submitted manuscript to other papers by the same author(s) that are currently submitted, in press, or recently published; (b) one copy of each such related paper; (c) one copy of a statement of the procedures of the research reported in the submitted manuscript in sufficient detail to serve as a guide for conduct of an exact replication; and (d) one copy of the summary tables (e.g., correlation tables, analysis of variance summaries) for the *major* analyses reported. If it appears that the supplementary materials in categories (c) or (d) might be bulky, the author should exercise judgment and restraint in selecting what to send. It is not necessary to send supplementary material that repeats verbatim manuscript content.

After a manuscript is accepted, authors will be expected to provide assurance that their data are available in unanalyzed form for reanalyses by other investigators and that their research procedures were consistent with the American Psychological Association's published principles of research ethics, except insofar as may be detailed otherwise in the manuscript.

Copyright and Permission: All rights reserved. Written permission must be obtained from the American Psychological Association for copying or reprinting text of more than 500 words, tables, or figures. Permission is normally granted contingent upon like permission of the author, inclusion of the APA copyright notice on the first page of reproduced material, and payment of a fee of $10 per page, table, or figure. Abstracting is permitted with credit to the source. Permissions and fees are waived for authors who wish to reproduce their own material for personal use; fees only are waived for authors who wish to use their own material commercially. Permission and fees are waived for the photocopying of isolated articles for nonprofit classroom or library reserve use by instructors and educational institutions. Libraries are permitted to photocopy beyond the limits of U.S. copyright law: (1) those post-1977 articles with a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through the Copyright Clearance Center, P.O. Box 765, Schenectady, NY 12301; (2) pre-1978 articles, provided that the per-copy fee stated in the Publishers' Fee List is paid through the Copyright Clearance Center, P.O. Box 765, Schenectady, NY 12301. Address requests for reprint permission to the Permissions Office, APA, 1200 Seventeenth Street, N.W., Washington, D.C. 20036.

Subscriptions: Subscriptions are available on a calendar year basis only (January through December). Nonmember rates for 1979: $60 domestic, $64 foreign, $6 single issue. APA member rate: $20. Write to Subscription Section, APA.

Change of Address: Send change of address notice and a recent mailing label to the attention of the Subscription Section, APA, 30 days prior to the actual change of address. APA will not replace undelivered copies resulting from address changes; journals will be forwarded only if subscribers notify the local post office in writing that they will guarantee second-class forwarding postage.

Reprints: Authors may order reprints of their articles from the printer when they receive proofs.

Back Issues and Back Volumes: For information regarding back issues or back volumes write to Order Department, APA.

Microform Editions: For information regarding microform editions write to any of the following: Johnson Associates, Inc., P.O. Box 1017, Greenwich, Connecticut 06830; University Microfilms, Ann Arbor, Michigan 48106; or Princeton Microfilms, Princeton, New Jersey 08540.

Published monthly in one volume per year by the American Psychological Association, Inc., 1200 Seventeenth Street, N.W., Washington, D.C. 20036. Printed in the U.S.A. Second-class postage paid at Washington, D.C., and at additional mailing offices.

JOURNAL OF
Personality
AND
Social Psychology

Volume 37 March 1979 Number 3

Attitudes Cause Behaviors: A Cross-Lagged Panel Analysis

Lynn R. Kahle
University of Michigan (Ann Arbor)

John J. Berman
University of Nebraska—Lincoln

Cross-lagged panel correlations were computed between attitudes and behaviors with respect to college students' ratings of four issues: Jimmy Carter's presidential candidacy, Gerald Ford's candidacy, drinking, and religion. For all four issues, attitudes showed causal predominance over behaviors. These results are interpreted as supporting McGuire's views on the attitude–behavior relationship but as being contrary to Wicker's situational determinism and self-perception theory. Kelman's view of reciprocal causation between attitudes and behaviors is also discussed.

The question of the causal relationship between personal factors such as attitudes, on the one hand, and behaviors, on the other, has a long and illustrious history. William James (1884) was one of the first psychologists to raise the question: "Common sense says, we lose our fortune, are sorry and weep; we meet a bear, are frightened and run; we are insulted by a rival, are angry and strike. The hypothesis here to be defended says that this order of sequence is incorrect" (p. 190).

At the most basic level, one could assume four types of relationships between attitudes and behavior: Attitudes cause behaviors; behaviors cause attitudes; there is reciprocal causation; or the two are causally unrelated. Each of these positions has a contemporary proponent. McGuire (1976) has argued that attitudes generally lead to behavior. D. Bem (1972) has proposed that when internal cues are weak or ambiguous, situationally determined behaviors will lead to attitudes. Kelman (1974) maintains that attitudes guide behaviors but that behaviors also guide attitudes. Wicker (1969) reviewed a number of studies on the relationship between attitudes and behaviors and concluded that "taken as a whole, these studies suggest that it is considerably more likely that attitudes will be unrelated or only slightly related to overt behaviors than that attitudes will be closely related to actions" (p. 65). Hence, the basic

A version of this article was presented at the annual meeting of the Midwestern Psychological Association, Chicago, May 6, 1978. A preliminary report of these data is also contained in the first author's doctoral dissertation (Kahle, 1977). He gratefully acknowledges the aid of his dissertation committee cochairmen, Monte M. Page and John J. Berman, and committee members Richard A. Dienstbier, Harry P. Shelley, and Bruce Dennis Sales. Also deserving of special thanks is David Kenny for his helpful comments on cross-lagged panel correlation.

Requests for reprints should be sent to Lynn R. Kahle, Room 5130, Institute for Social Research, University of Michigan, Ann Arbor, Michigan 48106.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3703-0315$00.75

Diary No. 67
Date 2/4/1929

question that James raised has yet to be resolved. Nevertheless, the results of many laboratory studies supporting Bem and/or Wicker have created some degree of pessimism about the predictive utility of attitudes (cf. McGuire, 1976).

Although discriminating between these four hypotheses in terms of which is more accurate is important for attitude theory and research, it has been difficult to contrast these hypotheses within one laboratory study, in part because, typically, only attitudes or only behaviors are used as independent variables. Thus, the reviewer must rely on only cross-study comparisons to reach any conclusions. This reliance is always risky, because of the numerous differences unrelated to theory that exist between different studies. Furthermore, random assignment of individuals to attitudes may be misleading, since the laboratory-induced attitudes either may be trivial or may interact with prior attitudes.

An entirely new methodology may be necessary before we can progress to a new level of understanding empirically the relationship between attitudes and behaviors. Kenny (1975) has developed a different methodology, cross-lagged panel correlation, which may prove useful for a better understanding of these issues. This article reports on the results of a cross-lagged panel correlation study of the relationship between attitudes and behaviors.

Method

Subjects

The 463 introductory psychology students who filled out questionnaires at both of the times they were administered were the subjects for this research. Each subject was asked questions about the candidacy of Jimmy Carter and Gerald Ford for the presidency, about drinking, and about religion. After deletion of subjects for whom at least one data point was missing for a particular set of variables, the sample size for the political issues was 362, for the drinking issues was 326, and for the religion issues was 394.

Cross-Lagged Panel Correlation

Because some readers may be unfamiliar with cross-lagged panel correlation (CLPC), a brief overview of the method is given here. A comprehensive overview has been given by Kenny (1975).

CLPC is a quasi-experimental design useful for testing causal hypotheses from correlational data. Within a panel (longitudinal) study, there may be both *panel variables,* which are relatively fluctuating (although they should be reliably measured) variables measured at more than one time, and *control variables,* which are relatively stable variables measured at only one point in time (e.g., sex). In the simplest case in which CLPC can be used, two panel variables, A and B, are measured at two separate times, Time 1 and Time 2. There are six possible correlations that can be computed with these data: two synchronous correlations (r_{A1B1} and r_{A2B2}), two autocorrelations (r_{A1A2} and r_{B1B2}), and two cross-lagged correlations (r_{A1B2} and r_{A2B1}). When the cross-lagged correlations are different, as indicated by a Pearson-Filon z test for differences between correlated correlations (Kenny, 1975), and when various other assumptions have been satisfied, one may state that there is a causal relationship between the variables. For example, if $r_{A1B2} > r_{B1A2}$, if both are positive, and if other assumptions have been satisfied, one may state that increases in A cause increases in B, as opposed to stating that increases in B cause increases in A.

In true experiments, randomization is used to rule out the plausibility of spuriousness (third-variable effects) as an alternative explanation to experimental effects. CLPC rules out spuriousness as an alternative explanation to research effects by means of two assumptions about the data: synchronicity and stationarity. *Synchronicity* means that the scores that comprise the synchronous correlations measure processes that occurred at the same time, and *stationarity* means that the causal structure for the variables did not change over time. Kenny (1973) has formulated a mathematical model that given synchronicity, stationarity, and spuriousness, the cross-lagged correlations do not differ significantly.

Stationarity may be of three types: perfect stationarity, proportional stationarity, or quasi stationarity. When the structural equation of the variables does not change over time, perfect stationarity is obtained. A lack of change in synchronous correlations is consistent with perfect stationarity. Proportional stationarity refers to the case in which the structural equations of the variables change by some proportional constant over the time lag. Quasi stationarity refers to the case in which the structural equation of each variable changes over time by a constant unique for each variable (Kenny, 1975). The implication of the assumption of quasi stationarity is that the synchronous correlations would equal each other if corrected for attenuation due to measurement unreliability. If a set of data meets the assumptions of quasi stationarity but not perfect or proportional stationarity, it is necessary to adjust the cross-lagged correlations prior to testing for spuriousness. This adjustment is necessary because variables with decreasing reliability would tend to appear to be causes, and variables with increasing reliability would tend to appear to be effects. It is

possible to test the various assumptions of stationarity statistically (Kenny, 1973, 1975, Note 1) when one has more than two panel variables.

Advantages and Limitations of CLPC

CLPC has an obvious advantage over traditional correlations in that it can be used for causal inferences. It also has some advantages over traditional analysis of variance designs, since it does not necessarily assume randomization and variable manipulation. As a correlational design, it has the further advantage that one need not discriminate between independent and dependent variables a priori. In the attitude–behavior controversy, for example, both attitudes and behaviors could have equal probability of showing up as a cause or as an effect in a CLPC design.

Although CLPC has several advantages over analysis of variance designs, it also has several limitations or weaknesses in comparison to analysis of variance designs. The first limitation is that CLPC is a low-power statistical procedure. Because of this low power, the alpha level for this research was set at .1 for the Pearson-Filon tests of differences between cross-lagged correlations.

Another problem with CLPC is specifying the appropriate time lag. Nothing about the procedure specifies what the appropriate time lag might be with respect to a particular research question. Quite often social science theory also fails to guide researchers to an appropriate time lag. This problem also exists in other longitudinal research and in laboratory research. In the area of attitude change, for example, McGuire (1960) argued that laboratory cognitive-consistency manipulations may not show their full effect for a week. The time lag for noticeable attitude change in the absence of an experimental manipulation is probably often much longer. One possible solution to this problem in the ideal case would be to sample numerous time lags. In the present study, the time lag was identical for all issues, but the issues did differ with respect to the pace at which they were expected to change. During an election year, one may assume that political perspectives are under a great deal of pressure to change, whereas religious and drinking perspectives may tend to change more slowly. Further advantages and limitations of CLPC are given elsewhere (Kahle, 1977; Kenny, 1975, in press).

Hypotheses

In this study, two panel variables, attitude and behavior, were measured for four issues—Jimmy Carter's presidential candidacy, Gerald Ford's candidacy, drinking, and religion.[1] It was hypothesized that, for each of the four issues, attitudes would lead to behaviors.

Diary No.....
Date.....

Materials

The items for all of the panel variables were developed specifically for the purposes of this research, in order to test the constructs of interest: attitudes and behaviors. An initial pool of items was generated by consulting some previous tests of the relevant issues (e.g., Manson, 1963; Robinson & Shaver, 1973) and by using the authors' knowledge of the issues. These 89 items were pilot tested on a group of 41 introductory psychology students. Topics of concern in this pilot study included whether scale points on the questionnaire were appropriately spaced and whether the questionnaire contained demand characteristics or other factors that could potentially bias the responses. Tests of internal consistency (Burns, Note 2) generally guided decisions about what items would be included in the variables for the actual research, but in some cases intuitions overruled psychometrics, because the Republican party had not yet selected a presidential candidate at the time of the pilot study.

Each subject in the actual research responded to all of the panel-variable items twice and all of the control-variable items once. Before any CLPC analyses were completed, the items for each panel variable were analyzed to test whether they were detracting from the internal consistency of the variable. Three political behavior items were eliminated because they detracted from the internal consistency of the variable that they were intended to represent. The eliminated items were ones that had highly skewed distributions and very small variances (e.g., "How many times in the past 14 days have you put up a sign or bumper sticker supporting Jimmy Carter?").

Subjects rated each attitude item on a scale ranging from strongly agree (1) to strongly disagree (9). For the behavior items, subjects were asked how many times in the past 14 days they had engaged in the activity mentioned in the item. The items for each variable (i.e., attitude and behavior) within each issue were summed to obtain one score for each variable.

Table 1 presents an example of an item from each set of questions used to measure the panel variables. The item that correlated most highly with the total score for all of the other items in each panel-variable scale was selected for inclusion in Table 1.

The scales appear to be adequate psychometrically. Table 2 presents two measures of internal consistency, Cronbach's (1951) alpha and Wollin's index of single-factoredness (Burns, Note 2), for each panel variable from both times of measurement. Each

---

[1] A third panel variable, stimulus condition self-selection, was also included in Kahle (1977). This third panel variable was used in the data reported here for tests of stationarity, since it is necessary to have at least three panel variables to test for stationarity.

Table 1
Representative Item From Each Set of Questions

Variable	Item
Carter issues	
Attitude	I strongly support Jimmy Carter for president. (5)
Behavior	How many times in the past 14 days have you made a favorable comment about Jimmy Carter? (7)
Ford issues	
Attitude	I strongly support Gerald Ford for president. (5)
Behavior	How many times in the past 14 days have you made a favorable comment about Gerald Ford? (7)
Drinking issues	
Attitude	I like to drink. (14)
Behavior	How many times in the past 14 days have you had at least one drink? (9)
Religion issues	
Attitude	Being active in church is important. (9)
Behavior	How many times in the past 14 days have you prayed in private?
	How many times in the past 14 days have you been with people from your church? (9)

Note. The number in parentheses after each item is the total number of items in that item's variable scale.

variable was also correlated with the Bem Sex Role Inventory (BSRI) Social Desirability Scale (S. Bem, 1974). All correlations were quite low, with the mean $r^2 = .01$ and the median $r = -.008$.

A number of control variables were measured along with the panel variables. At Time 1, subjects provided demographic information and completed the Bem Sex Role Inventory. At Time 2, subjects answered questions about their religious affiliation, interest in the presidential campaign, self-perceptions of the importance of their attitudes ("Among all of your attitudes, how important are your attitudes toward alcoholic beverages?"), and their attitudinal entailment ("How many attitudes do you have which are related to your attitudes toward alcoholic beverages?") (Kahle, in press). A principal-components factor analysis with a varimax rotation was performed on the Bem Sex Role Inventory,[2] and all items with a factor loading greater than .55 after rotation were summed for each factor with an eigenvalue greater than 1.6. These factors were used as control variables.

The purpose of control variables in CLPC is to increase the stationarity of the data (Kenny, 1975). Three criteria were employed to select which of the many control variables would in fact be used for the cross-lagged panel analysis. To be included, a control variable had (a) to account for some noticeable variance of several of the panel variables within a given analysis (Kenny, 1975), (b) to have some theoretical relevance, and (c) not to reduce sample size excessively because of missing data.

Control variables for each set of panel variables were partialed from all panel variables in that set before any cross-lagged panel tests were performed (Kenny, Note 1). The control variables that met all three of the above criteria for inclusion with respect to political variables included independence (a factor from the Bem Sex Role Inventory), hometown size, importance of political attitudes, and number of debates observed. The drinking control variables were year in college, sex, importance of drinking, and entailment (Kahle, in press) of drinking. The religion control variables were hours per week of work during the school year, extent of self-support for education, and importance of religion.

Procedure

The potential pool of respondents was students enrolled in a large introductory psychology course.

[2] Other researchers at the same university were administering the BSRI to the same subjects for different purposes. Because the BSRI Social Desirability Scale was judged to be psychometrically adequate and in order to avoid overwhelming the subjects with yet another rating scale, a more traditional measure of social desirability was not administered. Given the availability of the data from the rest of the BSRI, it was decided to use those data, too. S. Bem's original masculinity, femininity, and androgyny scores did not account for major amounts of variance in the attitude and behavior data here. In general, the factor analysis showed about the same factor structure that one would expect based on S. Bem's (1974) reported internal consistency measures, with some minor qualifications probably due to differences in subject populations.

At the first meetings of this class, the instructors introduced the course and then asked the students to complete a "Background Information Questionnaire." Students attended these sessions in groups of approximately 75 during the week of August 30, 1976. The date was less than 2 weeks after the Republican National Convention had nominated Gerald Ford.

The students were offered half an hour of research participation credit toward their 3-hour requirement to complete the questionnaire again during the last complete week prior to the presidential election (October 25, 1976). Thus, the time lag was selected to be appropriate for the political issues, and it may or may not have been appropriate for the drinking issues or the religion issues.

Results

Table 3 displays the results from the tests of differences among the cross-lagged correlations. For all four issues, the "attitudes lead to behaviors" hypothesis was significantly supported in the direction predicted.

For those readers familiar with CLPC, several additional details may be useful in interpreting Table 3. All sychronous correlations and autocorrelations were positive and were larger than .30. The degrees of freedom for the tests were 360 for the Carter variables, 360 for the Ford variables, 392 for the religion variables, and 324 for the drinking variables.

Table 2
Internal Consistency of Scales: Cronbach's Alpha[a] and Wollin's Index[b] of Single-Factoredness

Variable	Cronbach's alpha		Wollin's index	
	Time 1	Time 2	Time 1	Time 2
Carter issues				
Attitude	.82	.86	.88	.90
Behavior	.72	.68	.92	.93
Ford issues				
Attitude	.84	.86	.94	.93
Behavior	.84	.75	.97	.92
Drinking issues				
Attitude	.90	.92	.85	.91
Behavior	.92	.89	.93	.95
Religion issues				
Attitude	.85	.86	.94	.96
Behavior	.79	.80	.82	.82

[a] Cronbach (1951).
[b] Burns (Note 2).

Table 3
Attitude–Behavior Cross-Lagged Correlation Test Results

Issue	r_{A1B2}	r_{B1A2}	Difference	z^a
Carter's candidacy	.53	.18	.35	6.29**
Ford's candidacy	.57	.13	.44	7.73**
Religion	.57	.49	.07	1.84*
Drinking	.58	.48	.10	2.20**

[a] Pearson-Filon test for the differences between correlated correlations, following Kenny (1975).
* $p < .1$.
** $p < .05$.

Only the political data have been corrected to meet the assumptions of quasi stationarity. This correction was statistically necessary (probably because the salience of politics intensified during the time lag), in order to rule out changing reliability as an alternative explanation to obtained effects. Uncorrected variables with increasing reliability could otherwise spuriously appear to be effects (Kenny, 1975). The disadvantage of this correction is that it may increase sampling error and thereby make the Pearson-Filon test more approximate. In the present case, in which the political data do not yield borderline results, the approximate nature of the significance test is probably not a problem. For the religion data, the assumption of perfect stationarity was tenable, and for the drinking data, the assumption of proportional stationarity was tenable; thus, it was unnecessary to adjust the religion and drinking data to meet the assumptions of quasi stationarity.

Kenny's weights (Note 1) for estimating the relative sampling error for the Carter, Ford, drinking, and religion data were, respectively, 2.2, 1.9, 8.2, and 7.9. In general, weights of less than 1 should not be interpreted, and weights between 1 and 2 should be interpreted with caution (Kenny, Note 1). These weights, then, imply that a slight amount of caution is called for in interpreting the Ford data. Kenny's (Note 1) overall reliability (communality) ratios for the Carter, Ford, drinking, and religion data

were 1.178, 1.264, 1.044, and 1.008, respectively. Reliability ratios greater than 1 indicate increasing reliability, whereas ratios less than 1 indicate decreasing reliability. Here reliabilities increased slightly. The larger of the cross-lagged correlations tended to be slightly larger than the synchronous correlations. Since Kenny (1975) argues that a cross-lag larger than the synchronous correlations is indicative of a causal effect, this result provides additional (although minor) evidence consistent with the inferences.

Discussion

The results are straightforward. Attitudes had causal predominance over behaviors for all four issues. The weak or nonexistent relationship between attitudes and behaviors that one would expect in the absence of extensive situational information, if one were to accept Wicker's (1969) analysis, was not supported here. This study provided ample evidence for the utility of assuming that attitudes lead to behaviors. In all cases the correlation between attitudes at Time 1 and behaviors at Time 2 was greater than .5.

This study clearly implies that knowledge of attitudes has an important degree of predictive utility. For the politicians, theologians, and alcoholism therapists who may be interested in knowing whether their efforts to change attitudes will help them to achieve their ultimate goals of changing behaviors, the answer appears to be yes. Furthermore, knowledge about present attitudes will help predict future behaviors (in the absence of change). The pessimism about the predictive utility of attitudes is to some extent diminished. With McGuire (1976) we can call into question Wicker's emphasis:

> It is currently fashionable to stress situational determinants of behavior and question whether there are any personal determinants at all (Endler and Hunt, 1968; Mischel, 1973; Alker, 1972; Endler, 1973; [D.] Bem and Allen, 1974). I might seem, therefore, to be violating the current orthodoxy in the yin and yang of conventional wisdom on this issue when I stress the role of attitudes in predicting behaviors that purportedly are determined primarily by situational factors. (p. 9)

The attitude–behavior cross-lagged correlation differences also contradict at least the starting point of self-perception theory (D. Bem, 1972). Although these results do not "disprove" self-perception theory, since Bem theorizes that behaviors will lead to attitudes only "to the extent that internal cues are weak, ambiguous, or uninterpretable" (p. 2), the results do suggest that with respect to topics such as politics, religion, and drinking, this qualification to self-perception theory may need to be invoked more often than the theory itself. And with respect to renditions of self-perception theory that minimize the importance of this qualification (e.g., Kleinke, 1978), the present data are quite damaging. The preponderance of attention given to artificial laboratory studies during the past several decades of attitude research may have distorted our ability to discriminate between major and minor naturally occurring processes involved in attitude change. By way of qualification, however, it is important to note that the data reported here have at best only slight implications concerning issues such as the heuristic utility of self-perception theory in areas such as attribution research. Even if the starting point of a theory is called into question, the total theory may nevertheless make other valuable contributions (Kahle, 1978).

Kenny's (1975) contribution in developing CLPC may have as many implications for theory as it has methodological implications, because of the way it allows researchers to ask questions. In CLPC research it is not necessary to determine irrevocably which of several factors is causal prior to conducting the research. Kenny's methodology also allows the investigation of reciprocal causation and cyclical causation (which may be characteristic of attitudes and behaviors) through the sampling of multiple time lags.

With respect to reciprocal causation, the present data do not suggest that behaviors are causally void. Certainly, a number of laboratory studies suggest that in certain circumstances, behaviors can lead to attitudes. A particularly fruitful area to explore in this vein may be the differences in attitudes about objects, the typical topic of interest, and attitudes about behaviors (Fishbein & Ajzen, 1974). Even when behaviors lead to attitudes, however, it may be that the re-

sultant attitudes themselves will lead to new behaviors (Kelman, 1974). McGuire's (1976) discussion of reciprocal causation is especially sophisticated.

Although the results of the present study clearly show that attitudes have causal predominance over behaviors on these issues, a few words of caution are in order. In some cases, a difference in levels of specificity between attitudes and behaviors may lead to inconsistency (Weigel, Vernon, & Tognacci, 1974). People may often hold contradictory attitudes (Kelman, 1974). Attitudes may be better predictors of groups of behaviors than of specific behaviors (Fishbein & Ajzen, 1974). What researchers view as attitude–behavior consistency may not be what research subjects view as consistency (Bernstein & Kahle, Note 3; cf. also Bem & Allen, 1974). And, of course, other things besides attitudes influence behaviors.

Reference Notes

1. Kenny, D. A. *PANAL: Panel data analysis of the first year of "Sesame Street."* Research report, Harvard University, 1975.
2. Burns, J. A. *A manual to accompany program* ATSCALE *evaluation of the unidimensionality and internal consistency of responses to a series of questions.* Unpublished manuscript, Northwestern University, 1973.
3. Bernstein, D. J., & Kahle, L. R. *An idiographic approach to cognitive consistency.* Paper presented at the meeting of the Midwestern Psychological Association, Chicago, May 1977.

References

Alker, H. A. Is personality situationally specific or intrapsychically consistent? *Journal of Personality,* 1972, *40,* 1–16.

Bem, D. J. Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6). New York: Academic Press, 1972.

Bem, D. J., & Allen, A. On predicting some of the people some of the time: The search for cross-situational consistencies in behavior. *Psychological Review,* 1974, *81,* 506–520.

Bem, S. L. The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology,* 1974, *42,* 155–162.

Cronbach, L. Coefficient alpha and the internal structure of tests. *Psychometrika,* 1951, *16,* 297–334.

Endler, N. S. The person versus the situation: A pseudo issue? A response to Alker. *Journal of Personality,* 1973, *41,* 287–303.

Endler, N. S., & Hunt, J. McV. S-R inventories of hostility and comparison of the proportions of variance from persons, responses, and situations for hostility and anxiousness. *Journal of Personality and Social Psychology,* 1968, *9,* 309–315.

Fishbein, M., & Ajzen, I. Attitudes toward objects as predictors of single and multiple behavioral criteria. *Psychological Review,* 1974, *81,* 59–74.

James, W. What is an emotion? *Mind,* 1884, *9,* 188–205.

Kahle, L. R. *Stimulus condition self-selection in person–situation interactionism.* Unpublished doctoral dissertation, University of Nebraska—Lincoln, 1977.

Kahle, L. R. Dissonance and impression management as theories of attitude change. *Journal of Social Psychology,* 1978, *105,* 53–64.

Kahle, L. R. Attitudinal entailment and resistance to change. *Journal of Social Psychology,* in press.

Kelman, H. C. Attitudes are alive and well and gainfully employed in the sphere of action. *American Psychologist,* 1974, *29,* 310–324.

Kenny, D. A. Cross-lagged and synchronous common factors in panel data. In A. S. Goldberg & O. D. Duncan (Eds.), *Structural equation models in the social sciences.* New York: Seminar Press, 1973.

Kenny, D. A. Cross-lagged panel correlation: A test for spuriousness. *Psychological Bulletin,* 1975, *82,* 887–903.

Kenny, D. A. *The application of causal models to correlational data.* New York: Wiley, in press.

Kleinke, C. L. *Self-perception: The psychology of personal awareness.* San Francisco: Freeman, 1978.

Manson, M. P. *The Western Personality Inventory.* Los Angeles: Western Psychological Services, 1963.

McGuire, W. J. Cognitive consistency and attitude change. *Journal of Abnormal and Social Psychology,* 1960, *60,* 345–353.

McGuire, W. J. The concept of attitudes and their relations to behaviors. In H. W. Sinaiko & L. A. Broedling (Eds.), *Perspectives on attitude assessment: Surveys and their alternatives.* Champaign, Ill.: Pendleton, 1976.

Mischel, W. Toward a cognitive social learning reconceptualization of personality. *Psychological Review,* 1973, *80,* 252–283.

Robinson, J. P., & Shaver, P. R. *Measures of social psychological attitudes.* Ann Arbor: Institute for Social Research, University of Michigan, 1973.

Weigel, R. H., Vernon, D. T. A., & Tognacci, L. N. The specificity of the attitude as a determinant of attitude–behavior congruence. *Journal of Personality and Social Psychology,* 1974, *30,* 724–728.

Wicker, A. W. Attitude versus action: The relationship of verbal and overt behavioral responses to attitude objects. *Journal of Social Issues,* 1969, *25*(4), 41–78.

Received November 15, 1977 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 3, 322-336

Egocentric Biases in Availability and Attribution

Michael Ross and Fiore Sicoly
University of Waterloo, Waterloo, Canada

Five experiments were conducted to assess biases in availability of information in memory and attributions of responsibility for the actions and decisions that occurred during a previous group interaction. The subject populations sampled included naturally occurring discussion groups, married couples, basketball teams, and groups assembled in the laboratory. The data provided consistent evidence for egocentric biases in availability and attribution: One's own contributions to a joint product were more readily available, that is, more frequently and easily recalled; individuals accepted more responsibility for a group product than other participants attributed to them. In addition, statements attributed to the self were recalled more accurately and the availability bias was attenuated, though not eliminated, when the group product was negatively evaluated (Experiment 2). Finally, when another participant's contributions were made more available to the individual via a selective retrieval process, the individual allocated correspondingly more responsibility for the group decisions to the coparticipant (Experiment 5). The determinants and pervasiveness of the egocentric biases are considered.

One instance of a phenomenon examined in the present experiments is familiar to almost anyone who has conducted joint research. Consider the following: You have worked on a research project with another person, and the question arises as to who should be "first author" (i.e., who contributed more to the final product?). Often, it seems that both of you feel entirely justified in claiming that honor. Moreover, since you are convinced that your view of reality must be shared by your colleague (there being only one reality), you assume that the other person is attempting to take advantage of you. Sometimes such concerns are settled or prevented by the use of arbitrary decision rules, for example, the rule of "alphabetical priority"—a favorite gambit of those whose surnames begin with letters in the first part of the alphabet.

We suggest, then, that individuals tend to accept more responsibility for a joint product than other contributors attribute to them. It is further proposed that this is a pervasive phenomenon when responsibility for a joint venture is allocated by the participants. In many common endeavors, however, the participants are unaware of their divergent views, since there is no need to assign "authorship"; consequently, the ubiquity of the phenomenon is not readily apparent. The purpose of the current research was to assess whether these egocentric perceptions do occur in a variety of settings and to examine associated psychological processes.

In exploring the bases of such differential perceptions, we are not so naive as to suggest that intentional self-aggrandisement never occurs. Nonetheless, it is likely that perceptions can be at variance in the absence of deliberate deceit; it is from this perspective that we approach the issue.

To allocate responsibility for a joint endeavor, well-intentioned participants presum-

This research was supported by a Canada Council grant to the first author. We are grateful to the following people for their comments on earlier versions of this article: Dick Bootzin, Leslie McArthur, Hildy Ross, Lee Ross, Shelley Taylor, Amos Tversky, and Mark Zanna. We thank Vicky Vetere for collecting the data in Experiment 5.

Requests for reprints should be sent to Michael Ross, Department of Psychology, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3703-0322$00.75

ably attempt to recall the contributions each made to the final product. Some aspects of the interaction may be recalled more readily, or be more available, than others, however. In addition, the features that are recalled easily may not be a random subset of the whole. Specifically, a person may recall a greater proportion of his or her own contributions than would other participants.

An egocentric bias in availability of information in memory, in turn, could produce biased attributions of responsibility for a joint product. As Tversky and Kahneman (1973) have demonstrated, people use availability, that is, "the ease with which relevant instances come to mind" (p. 209), as a basis for estimating frequency. Thus, if self-generated inputs were indeed more available, individuals would be likely to claim more responsibility for a joint product than other participants would attribute to them.

There are at least four processes that may be operating to increase the availability of one's own contributions: (a) selective encoding and storage of information, (b) differential retrieval, (c) informational disparities, and (d) motivational influences.

Selective Encoding and Storage

For a number of reasons, the availability of the person's own inputs may be facilitated by differential encoding and storage of self-generated responses. First, individuals' own thoughts (about what they are going to say next, daydreams, etc.) or actions may distract their attention from the contributions of others. Second, individuals may rehearse or repeat their own ideas or actions; for example, they might think out their position before verbalizing and defending it. Consequently, their own inputs may receive more "study time," and degree of retention is strongly related to study time (Carver, 1972). Third, individuals' contributions are likely to fit more readily into their own cognitive schema, that is, their unique conception of the problem based on past experience, values, and so forth. Contributions that fit into such preexisting schemata are more likely to be retained (Bartlett, 1932; Bruner, 1961).

Differential Retrieval

The availability bias could also be produced by the selective retrieval of information from memory. In allocating responsibility for a joint outcome, the essential question from each participant's point of view may be, "How much did *I* contribute?" Participants may, therefore, attempt to recall principally their own contributions and inappropriately use the information so retrieved to estimate their *relative* contributions, a judgment that cannot properly be made without a consideration of the inputs of others as well.

Informational Disparities

There are likely to be differences in the information available to the contributors that could promote egocentric recall. Individuals have greater access to their own internal states, thoughts, and strategies than do observers. Moreover, participants in a common endeavor may differ in their knowledge of the frequency and significance of each other's independent contributions. For example, faculty supervisors may be less aware than their student colleagues of the amount of time, effort, or ingenuity that students invest in running subjects, performing data analyses, and writing preliminary drafts of a paper. On the other hand, supervisors are more cognizant of the amount and of the importance of the thought, reading, and so on that they put into the study before the students' involvement begins.

Motivational Influences

Motivational factors may also mediate an egocentric bias in availability. One's sense of self-esteem may be enhanced by focusing on, or weighting more heavily, one's own inputs. Similarly, a concern for personal efficacy or control (see deCharms, 1968; White, 1959) could lead individuals to dwell on their own contributions to a joint product.

The preceding discussion outlines a number of processes that may be operating to render one's own inputs more available (and more likely to be recalled) than the contributions of others. Consequently, it may be dif-

ficult to imagine a disconfirmation of the hypothesis that memories and attributions are egocentric. As Greenwald (Note 1) has observed, however, the egocentric character of memory "is not a necessary truth. It is possible, for example, to conceive of an organization of past experience that is more like that of some reference work, such as a history text, or the index of a thesaurus" (p. 4). In addition, we were unable to find published data directly supportive of the hypothesized bias in availability. Finally, recent developments in the actor–observer literature seem inconsistent with the hypothesis that memories and attributions are egocentric. Jones and Nisbett (1971) speculated that actors are disposed to locate the cause of their behavior in the environment, whereas observers attribute the same behavior to stable traits possessed by the actors. Though a variety of explanations were advanced to account for this effect (Jones & Nisbett, 1971), the recent emphasis has been on perceptual information processing (Storms, 1973; Taylor & Fiske, 1975). The actor's visual receptors are aimed toward the environment; an observer may focus directly on the actor. Thus, divergent aspects of the situation are salient to actors and observers, a disparity that is reflected in their causal attributions. This proposal seems to contradict the thesis that actors in an interaction are largely self-absorbed.

Two studies offer suggestive evidence for the present hypothesis. Rogers, Kuiper, and Kirker (1977) showed that trait adjectives were recalled more readily when subjects had been required to make a judgment about self-relevance (to decide whether each trait was descriptive of them) rather than about a number of other dimensions (e.g., synonymity judgments). These data imply that self-relevance increases availability; however, Rogers et al. did not contrast recall of adjectives relevant to the self with recall of adjectives relevant to other people—a comparison that would be more pertinent to the current discussion. Greenwald and Albert (1968) found that individuals recalled their own arguments on an attitude issue more accurately than the written arguments of other subjects. Since the arguments of self and other were always on opposite sides of the issue, the Greenwald and Albert finding could conceivably reflect increased familiarity with, and memory for, arguments consistent with one's own attitude position rather than enhanced memory for self-generated statements (although the evidence for attitude-biased learning is equivocal, e.g., Greenwald & Sakumura, 1967; Malpass, 1969).

We conducted a pilot study to determine whether we could obtain support for the hypothesized bias in availability. Students in an undergraduate seminar were asked to estimate the number of minutes each member of the seminar had spoken during the immediately preceding class period. An additional 26 subjects were obtained from naturally occurring two-person groups approached in cafeterias and lounges. The participants in these groups were asked to estimate the percentage of the total time each person had spoken during the current interaction.

It was assumed that subjects would base their time estimates on those portions of the conversation they could recall readily. Thus, if there is a bias in the direction of better recall of one's own statements, individuals' estimates of the amount of time they themselves spoke should exceed the average speaking time attributed to them by the other member(s) of the group.

The results were consistent with this reasoning. For seven of the eight students in the undergraduate seminar, assessments of their own discussion time exceeded the average time estimate attributed to them by the other participants ($p < .05$, sign test). Similarly, in 10 of the 13 dyads, estimates of one's own discussion time exceeded that provided by the other participant ($p < .05$, sign test). The magnitude of the bias was highly significant over the 13 dyads, $F(1, 12) = 14.85$, $p < .005$; on the average, participants estimated that they spoke 59% of the time. These data provide preliminary, albeit indirect, evidence for the hypothesized availability bias in everyday situations.

The principle objectives of the current research were (a) to assess the ocurrence of egocentric biases in availability and attributions of responsibility in different settings; (b)

to examine factors that were hypothesized to influence these biases; and (c) to offer preliminary evidence of a relation between a bias in availability and a bias in attributions of responsibility. Experiment 1 assessed the occurrence of egocentric biases in availability and allocations of responsibility in a natural setting and examined the relation between the two biases. Next, a laboratory experiment was conducted to address the issue of whether the quality of the group's performance affects the availability bias: Is the tendency for one's own inputs to be more available reduced substantially when the group's performance is poor, as a *motivational* interpretation would suggest? Experiment 3 further examined the effects of success and failure in a natural setting. The experimental manipulations in Experiments 4 and 5 were designed to influence availability, and changes in attributions of responsibility were assessed. The manipulation in Experiment 4 induced *differential encoding;* the manipulation in Experiment 5 varied the *retrieval* cues provided to the subjects.

Experiment 1

In this experiment, we wished to examine egocentric biases in naturally occurring, continuing relationships. Married couples appeared to represent an ideal target group. Spouses engage in many joint endeavors of varying importance. This circumstance would appear to be rife with possibilities for egocentric biases.

Accordingly, the first experiment was conducted (a) to determine if egocentric biases in allocations of responsibility occur in marital relationships; (b) to replicate, using a different dependent measure, the egocentric bias in availability obtained in the pretest; and (c) to correlate the bias in availability with the bias in responsibility. If the bias in responsibility is caused by a bias in availability, the two sets of data should be related.

Method

Subjects

The subjects were 37 married couples living in student residences. Twenty of the couples had children. The subjects were recruited by two female research assistants who knocked on doors in the residences and briefly described the experiment. If the couple were willing to participate, an appointment was made. The study was conducted in the couple's apartment; each couple was paid $5 for participating.

Procedure

A questionnaire was developed on the basis of extensive preliminary interviews with six married couples. In the experiment proper, the questionnaire was completed individually by the husband and wife; their anonymity was assured. The first pages of the questionnaire required subjects to estimate the extent of their responsibility for each of 20 activities relevant to married couples by putting a slash through a 150-mm straight line, the endpoints of which were labeled "primarily wife" and "primarily husband."[1] The twenty activities were making breakfast, cleaning dishes, cleaning house, shopping for groceries, caring for your children, planning joint leisure activities, deciding how money should be spent, deciding where to live, choosing friends, making important decisions that affect the two of you, causing arguments that occur between the two of you, resolving conflicts that occur between the two of you, making the house messy, washing the clothes, keeping in touch with relatives, demonstrating affection for spouse, taking out the garbage, irritating spouse, waiting for spouse, deciding whether to have children.

Subjects were next asked to record briefly examples of the contributions they or their spouses made to each activity. Their written records were subsequently examined to assess if the person's own inputs were generally more "available." That is, did the examples reported by subjects tend to focus more on their own behaviors than on their spouses'? A rater, blind to the experimental hypothesis, recorded the number of discrete examples subjects provided of their own and of their spouses' contributions. A second rater coded one third of the data; the reliability (Pearson product-moment correlation) was .81.

Results

The responses of both spouses to each of the responsibility questions were summed,

[1] In the preliminary interviews, we used percentage estimates. We found that subjects were able to remember the percentages they recorded and that postquestionnaire comparisons of percentages provided a strong source of conflict between the spouses. The use of the 150-mm scales circumvented these difficulties; subjects were not inclined to convert their slashes into exact percentages that could then be disputed.

so that the total included the amount that the wife viewed as her contribution and the amount that the husband viewed as his contribution. Since the response scale was 150 mm long, there were 150 "units of responsibility" to be allocated. A sum of greater than 150 would indicate an egocentric bias in perceived contribution, in that at least one of the spouses was overestimating his or her responsibility for that activity. To assess the degree of over- or underestimation that spouses revealed for each activity, 150 was subtracted from each couple's total. A composite score was derived for the couple, averaging over the 20 activities (or 19, when the couple had no children).

An analysis of variance, using the couple as the unit of analysis, revealed that the composite scores were significantly greater than zero, $M = 4.67$, $F(1, 35) = 12.89$, $p < .001$, indicating an egocentric bias in perceived contributions. Twenty-seven of the 37 couples showed some degree of overestimation ($p < .025$, sign test). Moreover, on the average, overestimation occurred on 16 of the 20 items on the questionnaire, including negative items—for example, causing arguments that occur between the two of you, $F(1, 32) = 20.38$, $p < .001$. Although the magnitude of the overestimation was relatively small, on the average, note that subjects tended to use a restricted range of the scale. Most responses were slightly above or slightly below the halfway mark on the scale. None of the items showed a significant underestimation effect.

The second set of items on the questionnaire required subjects to record examples of their own and of their spouses' contributions to each activity. A mean difference score was obtained over the 20 activities (averaging over husband and wife), with the number of examples of spouses' contributions subtracted from the number of examples of own contributions. A test of the grand mean was highly significant, $F(1, 35) = 36.0$, $p < .001$; as expected, subjects provided more examples of their own ($M = 10.9$) than of their spouses' ($M = 8.1$) inputs. The correlation between this self–other difference score and the initial measure of perceived responsibility was determined. As hypothesized, the greater the tendency to recall self-relevant behaviors, the greater was the overestimation in perceived responsibility, $r(35) = .50$, $p < .01$.

The number of words contained in each behavioral example reported by the subjects was also assessed to provide a measure of elaboration or richness of recall. The mean number of words per example did not differ as a function of whether the behavior was reported to be emitted by self ($M = 10.0$) or spouse ($M = 10.1$), $F < 1$. Further, this measure was uncorrelated with the measure of perceived responsibility, $r(35) = -.15$, *ns*.

In summary, both the measure of responsibility and the measure reflecting the availability of relevant behaviors showed the hypothesized egocentric biases. Moreover, there was a significant correlation between the magnitude of the bias in availability and the magnitude of the bias in responsibility. This finding is consistent with the hypothesis that egocentric biases in attributions of responsibility are mediated by biases in availability. Finally, the amount of behavior recalled seemed to be the important factor, rather than the richness of the recall.

Experiment 2

The data from Experiment 1 indicate that egocentric biases in availability and attributions of responsibility occur in ongoing relationships. The remaining experiments were designed to demonstrate the prevalence of these phenomena, and to investigate some of the factors that were expected to influence their magnitude.

The major purpose of Experiment 2 was to evaluate the self-esteem interpretation of the availability bias. If the availability bias is caused primarily by the motivation to enhance self-esteem, recall of a joint endeavor should facilitate an acceptance of personal responsibility after success and a denial of personal responsibility after failure. Consequently, the self-esteem interpretation implies that self-generated inputs should be more available after success than after failure. The evidence from past research that

people accept more responsibility for a success than for a failure is consistent with this reasoning (e.g., Luginbuhl, Crowe, & Kahan, 1975; Sicoly & Ross, 1977; Wortman, Costanzo, & Witt, 1973).

In Experiment 2, subjects learned several days after participating in a problem-solving task that their group had performed either well or poorly. It was hypothesized that subjects would recall a greater proportion of their own statements when the group product was positively evaluated. Because we moved to the laboratory for this experiment, it was possible to tape record the group's initial interaction. This recording provided a "reality base" against which to compare the subsequent recall of subjects.

Method

Subjects

The subjects were 37 males and 7 females selected from lists of students living at the university. Subjects were paid $5 each. All of the subjects participated in both sessions.

Procedure

The experiment was conducted in two sessions separated by a 3- or 4-day interval. Subjects reported for the first session in groups of two. They were told that the purpose of the study was to determine whether groups exhibit more social awareness than individuals. They were given 10 minutes to read a case study of Paula, a psychologically troubled person (selected from Goldstein & Palmer, 1975). Each subject in the dyad was provided with different portions of the case study. The subjects were next asked questions designed to assess their psychological understanding of Paula's difficulties. They were told to discuss each question and arrive at a joint response, taking into account the different information that each group member brought with him or her to the case. Subjects were told that their discussions were being tape recorded. The experimenter informed them that she would listen to the tapes following the session to evaluate the group's answers.

Subjects returned individually for the second session. In a random one half of the dyads, subjects were led to believe that their group had performed poorly relative to other groups in the experiment (third from the worst). In the remaining dyads, subjects were informed that their group had performed relatively well (third from the best). Subjects were then told, "Write down as much as you can recall of your group's discussion of Paula. You will only have a short time to do this, so it is unlikely that you will be able to report all or even most of what was said. It is, therefore, important that you put things down in the order that they come to you. . . . If you remember the idea, but not the exact comment, rephrase it in your own words." Subjects were told not to record who said each statement. They were simply to write what was said.

Subjects were asked to stop writing at the end of 8 minutes and to go back over their responses to indicate who said each statement during the discussion. Finally, they were asked whether, in their opinion, each statement improved or lowered their group's score, or whether they were uncertain. Subjects were debriefed at the end of the second session.

An observer who was blind to the subjects' treatment conditions contrasted subjects' recall with their original comments on the tape to assess accuracy. A statement was judged to be accurate if it represented an idea that the subject expressed during the interaction, even though the actual words used during the discussion might have differed from the words recalled by the subject. A second rater scored a random one third of the tapes, and agreement was 93%.

Results and Discussion

Availability

The proportion of statements that subjects attributed to themselves was calculated for each member of the dyad. The average proportion for each dyad served as the unit of analysis. In 21 of the 22 dyads, the subjects attributed the majority of the statements that they recalled to themselves ($p < .001$, sign test). The average proportion of subjects' own statements was .70 in the success condition and .60 in the failure condition. Each of these proportions was significantly greater than a .50 or chance expectancy, $t(10) = 9.09$, $p < .001$, and $t(10) = 3.22$, $p < .01$, respectively. Thus, in both the success and failure conditions, subjects attributed significantly more of the recalled statements to themselves than would be expected by chance. Nevertheless, as hypothesized, subjects attributed a greater proportion of the recalled statements to themselves after a success than after a failure, $F(1, 20) = 7.10$, $p < .025$. The total number of statements recalled (adding over statements attributed to self and the other person) did not differ

significantly as a function of the group's performance ($F = 1.57$).

Accuracy

Subjects' recall was compared with the taped record in a 2 × 2 between–within analysis of variance (Success vs. Failure × Self vs. Partner), with the dyad as the unit of analysis. Subjects recalled a higher percentage of their own actual statements ($M = 5.6\%$) than of their partner's actual statements ($M = 2.6\%$), $F(1, 19) = 18.37$, $p < .001$.[2] Although the means seem low, note that subjects were given only an 8-minute recall period. The group's performance level did not affect the percentage of actual statements recalled (main effect and interaction Fs < 1).

We also compared the accuracy of the statements subjects attributed to themselves with the accuracy of the statements they attributed to their partners. Sixty-nine percent of the statements that subjects attributed to self were accurate reflections of self-generated comments; 56% of the statements that subjects attributed to their partners were accurate. The difference between these two percentages was significant, $F(1, 19) = 7.06$, $p < .025$. The group's performance level did not significantly affect the accuracy of the attributed statements (success–fail main effect $F = 1.14$, interaction $F < 1$).

Most of the errors that subjects made were of two types: They recalled material from the case history that had not been mentioned in the discussion; they reported inferences and conclusions that were not contained in the case history or in the discussion. In only a few instances (approximately 2% of the errors) did subjects take credit for statements made by their partners.

Evaluations

Finally, subjects' evaluations of the statements were transcribed onto a 3-point scale: +1 (improved the group's score), 0 (uncertain), and −1 (lowered the group's score). Two scores were obtained for each subject: the average rating of comments attributed to self and the average rating of statements attributed to the other person. An analysis of variance, with the dyad as the unit of analysis, revealed a main effect for success–fail, $F(1, 20) = 14.56$, $p < .005$, and a Success–Fail × Self–Other interaction, $F(1, 20) = 5.19$, $p < .05$.

The success–fail main effect indicated that statements were evaluated more positively following success ($M = .75$) than following failure ($M = .41$). The interaction revealed that whereas subjects' evaluations of their own comments were marginally lower in the failure condition than in the success condition (M difference $= .18$, $t = 1.85$, $p < .10$), their evaluations of the other person's comments were significantly lower in the failure condition than in the success condition (M difference $= .50$, $t = 5.14$, $p < .01$).

In summary, the present study provided some evidence for the self-esteem maintenance hypothesis. Subjects attributed a higher proportion of the recalled comments to themselves after success than after failure; subjects' evaluations of the recalled statements suggested an attempt to shift the blame for failure onto their partners. On the other hand, contrary to the self-esteem interpretation, recall was egocentric even in the failure condition.

Note that the strong egocentricity obtained on the recall measure and the increased accuracy of self-generated statements may reflect, in part, the fact that subjects initially read different aspects of the case history. Since they subsequently presented this material to the other person in responding to the questions, subjects' own contributions may have received more "study time." Nevertheless, this differential is ecologically valid. A person's inputs are often derived from his or her previous history and experiences.

Experiment 3

In Experiment 3 we examined the effects of success and failure in a more natural setting. We had the players on 12 intercollegiate basketball teams individually complete a

[2] The tapes from one of the failure groups were lost; this group is omitted from the analysis.

questionnaire in which they were asked to recall an important turning point in their last game and to assess why their team had won or lost.

It is a leap to go from the self–other comparisons that we have considered in the previous studies to own team – other team comparisons. There are, however, a number of reasons to expect that the actions of one's own team should be more available to the attributor than the actions of the other team: I know the names of my teammates, and therefore, I have a ready means of organizing the storage and retrieval of data relevant to them; our success in future games against other opponents depends more on our own offensive and defensive abilities than on the abilities of the opposing team. Consequently, I may attend more closely to the actions of my teammates, which would enhance encoding and storage. Also, there are informational disparities: The strategies of my own team are more salient than are the strategies of the opposing team (Tversky & Kahneman, 1973).

If the initiatives of one's own team are differentially available, players should recall a turning point in terms of the actions of their team and attribute responsibility for the game outcome to their team. On the basis of the data from Experiment 2, it may be expected that these tendencies will be stronger after a win than after a loss.

Method

Subjects

Seventy-four female and 84 male intercollegiate basketball players participated in the study. The team managers were contacted by telephone; all agreed, following discussions with their players, to have their teams participate in the study.

Procedure

The questionnaires were administered after six games in which the teams participating in the study played each other. Thus, for the three male games chosen, three of the six male teams in the study were competing against the other three male teams. Similarly, the three female games selected included all six of the female teams. The questionnaires were administered at the first team practice following the target game (1 or 2 days after the game), except in one case where, because of the teams' schedules of play, it was necessary to collect data immediately after the game (two female teams). The questionnaires were completed individually, and the respondents' anonymity was assured. The relevant questions, from the current perspective, were the following:

1. Please describe briefly one important turning point in the last game and indicate in which period it occurred.

2. Our team won/lost our last game because. . . .

The responses to the first question were examined to determine if the turning point was described as precipitated by one's own team, both teams, or the other team. Responses to the second question were examined to assess the number of reasons for the win or loss that related to the actions of either one's own or the opposing team. The data were coded by a person who was unaware of the experimental hypotheses. A second observer independently coded the responses from 50% of the subjects. There was 100% agreement for both questions.

Results

There were no significant sex differences on the two dependent measures; the results are, therefore, reported collapsed across gender. Since team members' responses cannot be viewed as independent, responses were averaged, and the team served as the unit of analysis.

A preliminary examination of the "turning point" data revealed that even within a team, the players were recalling quite different events. Nevertheless, 119 players recalled a turning point that they described as precipitated by the actions of their own team; 13 players recalled a turning point that they viewed as caused by both teams; 16 players recalled a turning point seen to be initiated by the actions of the opposing team (the remaining 10 players did not answer the question). Subjects described such events as a strong defense during the last 2 minutes of the game, a defensive steal, a shift in offensive strategies, and so on.

The percentage of players who recalled a turning point caused by their teammates was derived for each team. These 12 scores were submitted to an analysis that compared them to a chance expectancy of 50%. The obtained distribution was significantly different from chance, $F(1, 11) = 30.25$, $p < .001$, with a mean of 80.25%. As hypothesized,

most reports emphasized the actions of the players' own team.

The percentage of players who recalled a turning point caused by their teammates was examined in relation to the team's performance. The average percentage was higher on the losing team than on the winning team in five of the six games ($p < .11$, sign test). The mean difference between the percentages on losing ($M = 88.5$) and winning ($M = 72.$) teams was nonsignificant ($F < 1$).

The players' explanations for their team's win or loss were also examined. Of the 158 participants, only 14 provided any reasons that involved the actions of the opposing team. On the average, subjects reported 1.79 reasons for the win or loss that involved their own team and .09 reasons that involved the opposing team, $F(1, 11) = 272.91$, $p < .001$. Finally, the tendency to ascribe more reasons to one's own team was nonsignificantly greater after a loss ($M = 1.73$) than after a win ($M = 1.65$), $F < 1$.

Discussion

The responses to the turning point question indicate that the performances of subjects' teammates were more available than those of opposing team members. Further, subjects ascribed responsibility for the game outcome to the actions or inactions of their teammates rather than to those of members of the opposing team. Thus, biases in availability and judgments of responsibility can occur at the group level. Rather and Heskowitz (1977) provide another example of group egocentrism: "CBS (news) became a solid Number One after the Apollo moonshot in 1968. If you are a CBS person, you tend to say our coverage of the lunar landing tipped us over. If you are a NBC person, you tend to cite the break-up of the Huntley–Brinkley team as the key factor" (p. 307).

Contrary to the data from Experiment 2, the availability bias in Experiment 3 was as strong after failure as after success. There are differences between the studies that may contribute to this discrepancy. The "egocentric" availability and attributions in the basketball experiment were team rather than self-oriented; as a result, responsibility for failure was more diffused, and subjects' self-esteem was threatened less directly. Also, unlike the group in the laboratory study, the basketball team had a future: The players could enhance their control over subsequent game outcomes by locating causality within their own team. Finally, and perhaps most important, unlike the laboratory group, the team also had a past. Team members recalled aspects of their behavior that changed and attributed the game outcomes to these variations (e.g., we win because of discipline and hustle; we lose because of a lack of discipline and hustle). What players seemed to ignore, however, was that the opposing teams might contribute to these fluctuations.

It seems likely that a tendency to perceive both teams as responsible for the game outcome might increase with the magnitude of the win or loss (assuming that large wins or losses are atypical). As Kelley (1973) noted, multiple causes are necessary to explain extreme outcomes. Although no such tendency was observed in the current study, there were too few data points (games) to provide an accurate determination.

Experiment 4

In the final two experiments, we examined the hypothesized relation between the bias in availability and the bias in attributions of responsibility more directly by introducing manipulations that should affect availability and measuring changes in attribution. In Experiment 4, subjects were required to record either their own comments (self-focus condition) or those of the other person (partner-focus condition) during a problem-solving session. At a second session, subjects were shown their notes and asked to assess the extent to which either they or their partners had been responsible for various aspects of the decision-making process. It was assumed that the partner-focus condition would enhance encoding and retrieval of the partner's contributions. Thus, the partner's inputs would be more available when assessments of responsibility were made, and subjects should assign their partner more responsi-

bility for group decision-making in the partner-focus condition than in the self-focus condition.

Method

Subjects

The subjects were 40 males recruited from the introductory psychology subject pool.

Procedure

Subjects were scheduled in pairs for the first session. In a few introductory comments, the experimenter described the difficulty of preventing people from smoking. Subjects were then told they were participating in a pilot project to assess the efficacy of "brainstorming techniques" as a means of providing possible solutions to this problem. They were further informed that solutions generated during their discussion would be sent to the Committee for the Prevention of Cigarette Smoking (a government committee). Subjects were told to follow a four-step sequence: define the problem, generate as many solutions as possible, discuss the pros and cons of each proposed solution, and finally, select a preferred solution and explain the reasons for this choice.

Subjects in the self-focus condition were asked to keep a record of their own contributions to the discussion. Subjects in the partner-focus condition were asked to keep a record of only the other person's inputs: "This will leave you free to think and develop your ideas because your partner will be doing the writing." Subjects were given about 45 minutes for discussion.

Subjects returned individually for a second session 2–3 days later. Each subject was asked to look over the notes he had taken during the previous session, "in order to refresh your recollection of the discussion."

Subjects completed the dependent measures after they had reviewed their notes of Session 1. The principal dependent variable required subjects to indicate who tended to control the course and content of the discussion during the first session. They were asked to assess this overall, and also with respect to the various stages of the discussion, on 150-mm scales with endpoints labeled "the other person" and "me."

Results

A 2 × 2 analysis of variance was performed on the data. Self- versus partner focus was a between-subjects factor. Since the dyad was treated as an experimental unit, the response made by each member of the pair was considered to be a repeated measure.

The focus manipulation had no reliable impact on attributions of responsibility (all Fs < 1). Once again, however, there was strong evidence of an egocentric bias in allocations of responsibility. Subjects reported that they had exerted more control over the course and content of each segment of the discussion than their partners ascribed to them—solutions stage: 85 versus 68, $F(1, 18) = 7.32$, $p < .025$; evaluation stage: 79 versus 70, $F(1, 18) = 4.13$, $p < .07$; final proposal stage: 86 versus 67, $F(1, 18) = 11.31$, $p < .01$; overall discussion: 89 versus 72, $F(1, 18) = 9.21$, $p < .01$. Note that for each item, A's self-attributions were, on the average, beyond the midpoint of the 150-mm scale, indicating a perceived contribution of greater than 50%; on the other hand, the partner viewed A's contributions as being less than 50% in each instance.

Discussion

The results revealed strong egocentric biases in individuals' attributions of responsibility for segments of the problem-solving task. The focus manipulation had surprisingly little effect, however. What we had viewed to be a sledgehammer manipulation turned out to be ineffective.

Why did the attention manipulation have so little impact? One possibility is that subjects may have found the written records to be relatively uninformative and relied more on their memories than on the notes in responding to the questionnaire (hence, the strong egocentric biases). The notes may have appeared inadequate, in part because they were very brief, usually about one page, relative to the length of the interaction (45 minutes). Moreover, much of what subjects wrote may have seemed irrelevant to the final decisions made by the group. In short, we may not have succeeded in focusing subjects' attention on what, from their perspective, were the important aspects of the interaction. To obtain this information they were, perhaps, only too willing to rely on their memories.

Finally, note that recent research on the relation between attention and recall in interpersonal perception settings has yielded

inconsistent results (Taylor & Fiske, 1978). This situation stands in marked contrast to the strong relation evident in the cognitive literature (Cofer, 1977; Loftus & Loftus, 1976). Unlike the constrained experimental settings utilized in cognitive research, however, the present study and those reviewed by Taylor and Fiske incorporate rich social environments. Consequently, the manipulation of attention is relatively gross; it is less certain that the individual is attending to those aspects of the situation that are relevant to the dependent measures.

Experiment 5

In Experiment 5, we again attempted to vary the individual's focus of attention so as to affect availability. In this experiment, however, we employed a manipulation designed to promote selective retrieval of information directly relevant to attributions of responsibility.

In our initial analysis, we suggested that egocentric attributions of responsibility could be produced by the selective retrieval of information from memory and that retrieval might be guided by the kinds of questions that individuals ask themselves. Experiment 5 was conducted to test this hypothesis. Subjects were induced to engage in differing retrieval by variations in the form in which questions were posed. Graduate students were stimulated to think about either their own contributions to their BA theses or the contributions of their supervisors. The amount of responsibility for the thesis that subjects allocated to either self or supervisor was then assessed. It was hypothesized that subjects would accept less responsibility for the research effort in the supervisor-focus than in the self-focus condition.

Method

Subjects

The subjects were 17 female and 12 male psychology graduate students. Most had completed either 1 or 2 years of graduate school. All of these students had conducted experiments that served as their BA theses in their final undergraduate year.

Procedure

The subjects were approached individually in their offices and asked to complete a brief questionnaire on supervisor–student relations. None refused to participate. The two forms of the questionnaire were randomly distributed to the subjects; they were assured that their responses would be anonymous and confidential.

One form of the questionnaire asked the subjects to indicate their own contribution to each of a number of activities related to their BA theses. The questions were as follows: (a) "I suggested _____ percent of the methodology that was finally employed in the study." (b) "I provided _____ percent of the interpretation of results." (c) "I initiated _____ percent of the thesis-relevant discussions with my supervisor." (d) "During thesis-related discussions I tended to control the course and content of the discussion _____ percent of the time. (e) "All things considered, I was responsible for _____ percent of the entire research effort." (f) "How would you evaluate your thesis relative to others done in the department?"

The second form of the questionnaire was identical to the above, except that the word *I* (self-focus condition) was replaced with *my supervisor* (supervisor-focus condition) on Questions 1–5. Subjects were asked to fill in the blanks in response to the first five questions and to put a slash through a 150-mm line, with endpoints labeled "inferior" and "superior," in response to Question 6.

Results and Discussion

For purposes of the analyses, it was assumed that the supervisor's and the student's contribution to each item would add up to 100%. Though the experiment was introduced as a study of supervisor–student relations, it is possible that the students may have considered in their estimates the inputs of other individuals (e.g., fellow students). Nevertheless, the current procedure provides a conservative test of the experimental hypothesis. For example, if a subject responded 20% to an item in the "I" version of the questionnaire, it was assumed that his or her supervisor contributed 80%. Yet the supervisor may have contributed only 60%, with an unspecified person providing the remainder. By possibly overestimating the supervisor's contribution, however, we are biasing the data against the experimental hypothesis: The "I" version was expected to reduce the percentage of responsibility allocated to the supervisor.

Subjects' responses to the first five questions on the "I" form of the questionnaire were subtracted from 100, so that higher numbers would reflect greater contributions by the supervisor in both conditions. Question 5 dealt with overall responsibility for the research effort. As anticipated, subjects allocated more responsibility to the supervisor in the supervisor-focus ($M = 33.3\%$) than in the self-focus ($M = 16.5\%$) condition, $F(1, 27) = 9.05$, $p < .01$. The first four questions were concerned with different aspects of the thesis, and the average response revealed a similar result: supervisor-focus $M = 33.34$; self-focus $M = 21.82$; $F(1, 27) = 5.34$, $p < .05$. Finally, subjects tended to evaluate their thesis more positively in the self-focus condition than in the supervisor-focus condition: 112.6 versus 94.6, $F(1, 27) = 3.59$, $p < .10$.

The contrasting wording of the questions had the anticipated impact on subjects' allocations of responsibility. The supervisor version of the questionnaire presumably caused subjects to recall a greater proportion of their supervisors' contributions than did the "I" form of the questionnaire. This differential availability was then reflected in the allocations of responsibility. Note, however, that the questions were not entirely successful in controlling subjects' retrieval. The supervisor was allocated only one third of the responsibility for the thesis in the supervisor-focus condition.

In light of the present data, the basketball players' attributions of responsibility for the game outcome in Experiment 3 need to be reexamined. Recall that the players were asked to complete the sentence, "Our team won/lost our last game because. . . ." This question yielded a highly significant egocentric bias. With hindsight, it is evident that the form of the question—"*Our* team . . . *our* last game"—may have prompted subjects to focus on the actions of their own teams, even though the wording does not preclude references to the opposing team. The "turning point" question in Experiment 3 was more neutrally worded and is not susceptible to this alternative interpretation.

The leading questions in these studies emanate from an external source; many of our retrieval queries are self-initiated, however, and our recall may well be biased by the form in which we pose retrieval questions to ourselves. For example, basketball players are probably more likely to think in terms of "Why did *we* win or lose?" than in terms of a neutrally phrased "Which team was responsible for the game outcome?"

General Discussion

The five studies employed different subject populations, tasks, and dependent measures. As hypothesized, the egocentric biases in availability and attribution appear to be robust and pervasive.

Determinants of the Availability Bias

Several processes were hypothesized to contribute to the increased availability of self-generated inputs. It is possible to consider how well each accounts for the existing data. *Selective encoding* and *storage* cannot have contributed to the effects of success versus failure on availability in Experiment 2 or of supervisor- versus self-focus in Experiment 5 (since these manipulations occurred long after encoding and storage took place). *Informational disparities* should not have contributed to the pretest results (subjects' time estimates were based solely on the preceding discussion), to the tendency to attribute a higher proportion of the recalled statements to oneself in the success as compared to the failure condition in Experiment 2, or to the effects of supervisor- versus self-focus in Experiment 5 (since neither performance level, as operationalized here, nor focus could affect the information initially available to the subjects). Two *motivational processes* were posited. Self-esteem maintenance does not seem pertinent to the results obtained from the two-person groups in the pretest. Nor does it account for (a) the overrecall of self-generated inputs in the failure condition of Experiment 2 and (b) the finding that players on losing basketball teams recalled the turning point of the game in terms of the actions of their teammates. The control motivation hypothesis fares somewhat better. Although focusing on one's own inputs in failure situations may lower self-esteem, it

does permit one to perceive personal control over the activity. Hence, efficacy motivation could account for these results. Nevertheless, a desire for personal efficacy does not appear to explain all of the data. The two-person groups in the pretest seem to reveal a relatively "pure" information-processing effect: It is unlikely that people would feel a need to report that they dominated casual conversations. Also, the effect of supervisor- versus self-focus in Experiment 5 appears to be mediated by differential retrieval. Efficacy considerations may have induced the subjects to report that they were major contributors to their theses; nonetheless, motivational concerns do not dictate that focusing on the supervisor's contributions will reduce one's need to assume responsibility.

In summary, selective encoding and storage, informational disparities, and motivational influences do not appear to be necessary determinants of the egocentric bias in availability. The one remaining process that was posited, *selective retrieval*, is not precluded by any of the current data; further, it receives direct support from the findings in Experiments 2 and 5.

Nevertheless, it seems premature to eliminate any of the hypothesized processes as sufficient causes of the availability bias. The tendency of spouses to recall their own contributions in Experiment 1 may reflect informational disparities; the desire to maintain self-esteem may have contributed to the effect of performance level in Experiment 2; basketball players' responses to the turning point question in Experiment 3 may well have been influenced by selective encoding and by control motivation.

We suspect that, like many cognitive phenomena (cf. Erdelyi, 1974; Erdelyi & Goldberg, in press; McGuire, 1973), biases in availability are multidetermined in real life. Multidetermination may seem an unsatisfying resolution; however, it is one that social psychologists shall probably confront increasingly as they begin to study cognitive phenomena in situ. Researchers in other sciences face parallel complexities. For example, similar cancers appear to have different etiologies, depending, among other factors, on the environment in which the patient lives (Goodfield, 1976).

The Link Between Availability and Attributions of Responsibility

The focus of the present research has been on demonstrating that the hypothesized biases in availability and attribution exist and are relatively ubiquitous. It was also hypothesized, however, that the egocentric bias in attributions of responsibility would be mediated by the bias in availability. Although the data are suggestive, we have no definitive evidence that the bias in availability *causes* the bias in responsibility. The strongest affirmative evidence is that the two biases were significantly correlated in the marriage study and that a manipulation designed to induce selective retrieval influenced attributions of responsibility (Experiment 5). In opposition, it might be contended that the covariation between the two biases is susceptible to a number of alternative causal interpretations and that there is no direct evidence that the retrieval manipulation in Experiment 5 affected availability. Conceivably, the attributions of responsibility in Experiment 5 were mediated by some other factor not yet identified. Further evidence will be required to establish whether the bias in responsibility is caused by the bias in availability. The present results suggest several additional considerations, however, concerning the determinants and pervasiveness of the biases.

Pervasiveness of the Egocentric Biases

The egocentric biases obtained in the current studies may seem inconsistent with Jones and Nisbett's (1971) proposal that actors locate causality for their actions primarily within their environment. There are a number of differences between the two paradigms that might account for the discrepancy. Most important, Jones and Nisbett were concerned with interpretation, whereas we focused on recall and judgments of responsibility. Actors could presumably overestimate their contributions to a joint product and, at the same time, locate the cause of their behavior with-

in the environment. For example, suppose that a wife who reports that she does 80% of the cleaning is asked *why* she cleans (the central question for Jones and Nisbett). She may respond that the house is dirty, an environmental attribution. Conversely, her husband, who perhaps accepts 30% of the responsibility for cleaning, may answer the same question by pointing out that his wife has a fetish for cleanliness, a trait attribution.

Thus, the current data do not speak directly to the Jones and Nisbett hypothesis. Nevertheless, our data do seem to contradict related evidence that the responses of actors are more salient and available to observers than to actors themselves (Storms, 1973; Taylor & Fiske, 1975). The critical variable may be the extent to which the observer departs from a passive role and interacts with the actor. When, as in the present research, individuals undertake complex social interactions, they alternate between the roles of speaker (actor) and listener (observer), yet much of their attention may be directed at planning and executing their own responses. Although they do not attend to themselves perceptually, they may be cognitively self-focused; therefore, self-generated inputs are likely to be more available in recall. On the other hand, passive observers may concentrate on other persons in their environment. Also, observers may be less self-absorbed when their own responses require little attention, as, for example, when they enact well-practiced behaviors (Langer, 1978, has speculated that a wide range of social behaviors require minimal thought).

These instances notwithstanding, the present research demonstrates the prevalence of self-centered biases in availability and judgments of responsibility. In everyday life, these egocentric tendencies may be overlooked when joint endeavors do not require explicit allocations of responsibility. If allocations are stated distinctly, however, there is a potential for dissension, and individuals are unlikely to realize that their differences in judgment could arise from honest evaluations of information that is differentially available.

Reference Note

1. Greenwald, A. G. *The tolitarian ego: Fabrication and revision of personal history.* Unpublished manuscript, 1978.

References

Bartlett, F. C. *Remembering.* Cambridge, England: Cambridge University Press, 1932.

Bruner, J. S. The act of discovery. *Harvard Educational Review,* 1961, *31,* 21–32.

Carver, R. P. A critical review of mathagenic behaviors and the effect of questions upon the retention of prose materials. *Journal of Reading Behavior,* 1972, *4,* 93–119.

Cofer, C. N. On the constructive theory of memory. In I. A. Uzgiris & F. Weizmann (Eds.), *The structuring of experience.* New York: Plenum Press, 1977.

deCharms, R. C. *Personal causation: The internal affective determinants of behavior.* New York: Academic Press, 1968.

Erdelyi, M. H. A new look at the new look: Perceptual defense and vigilance. *Psychological Review,* 1974, *81,* 1–25.

Erdelyi, M. H., & Goldberg, B. Let's not sweep repression under the rug: Towards a cognitive psychology of repression. In J. F. Kihlstrom & F. J. Evans (Eds.), *Functional disorders of memory.* Hillsdale, N.J.: Erlbaum, in press.

Goldstein, M. J., & Palmer, J. O. *The experience of anxiety.* New York: Oxford University Press, 1975.

Goodfield, J. *The siege of cancer.* New York: Dell, 1976.

Greenwald, A. G., & Albert, R. D. Acceptance and recall of improvised arguments. *Journal of Personality and Social Psychology,* 1968, *8,* 31–34.

Greenwald, A. G., & Sakumura, J. S. Attitude and selective learning: Where are the phenomena of yesteryear? *Journal of Personality and Social Psychology,* 1967, *7,* 387–397.

Jones, E. E., & Nisbett, R. E. *The actor and the observer: Divergent perceptions of the causes of behavior.* Morristown, N.J.: General Learning Press, 1971.

Kelley, H. H. The process of causal attribution. *American Psychologist,* 1973, *28,* 107–128.

Langer, E. J. Rethinking the role of thought in social interaction. In J. H. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research* (Vol. 2). Potomac, Md.: Erlbaum, 1978.

Loftus, G. R., & Loftus, E. F. *Human memory: The processing of information.* Hillsdale, N.J.: Erlbaum, 1976.

Luginbuhl, J. E. R., Crowe, D. H., & Kahan, J. P. Causal attribution for success and failure. *Journal of Personality and Social Psychology,* 1975, *31,* 86–93.

Malpass, R. S. Effects of attitude on learning and memory: The influence of instruction-induced sets. *Journal of Experimental Social Psychology,* 1969, *5,* 441–453.

McGuire, W. J. The yin and yang of progress in social psychology: Seven Koan. *Journal of Personality and Social Psychology*, 1973, *26*, 446–456.

Rather, D., & Heskowitz, M. *The camera never blinks*. New York: Ballantine Books, 1977.

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology*, 1977, *35*, 677–688.

Sicoly, F., & Ross, M. The facilitation of ego-biased attributions by means of self-serving observer feedback. *Journal of Personality and Social Psychology*, 1977, *35*, 734–741.

Storms, M. D. Videotape and the attribution process: Reversing actors' and observers' points of view. *Journal of Personality and Social Psychology*, 1973, *27*, 165–175.

Taylor, S. E., & Fiske, S. T. Point of view and perceptions of causality. *Journal of Personality and Social Psychology*, 1975, *32*, 439–445.

Taylor, S. E., & Fiske, S. T. Salience, attention, and attribution: Top of the head phenomena. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 11). New York: Academic Press, 1978.

Tversky, A., & Kahneman, D. Availability: A heurisitc for judging frequency and probability. *Cognitive Psychology*, 1973, *5*, 207–232.

White, R. W. Motivation reconsidered: The concept of competence. *Psychological Review*, 1959, *66*, 297–333.

Wortman, C. B., Costanzo, P. R., & Witt, T. R. Effect of anticipated performance on the attribution of causality to self and others. *Journal of Personality and Social Psychology*, 1973, *27*, 372-381.

Received February 20, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 3, 337-344

Ambient Temperature and the Occurrence of Collective Violence: A New Analysis

J. Merrill Carlsmith and Craig A. Anderson
Stanford University

Prevalent folklore suggests that riots tend to occur during periods of very hot weather. Baron and Ransberger examined 102 major riots in the United States between 1967 and 1971 and concluded that the frequency of collective violence and ambient temperature are curvilinearly related. The present article points out that the Baron and Ransberger analysis did not take account of the different number of days in different temperature ranges. The artifact is eliminated, and the probability of a riot, conditional upon temperature, is estimated. When this is done, the evidence strongly suggests that the conditional probability of a riot increases monotonically with temperature. Some general implications of such data analyses are discussed.

In a recent article in this journal, Baron and Ransberger (1978) presented an analysis of the relationship between the frequency of major riots and the ambient temperature occurring during the riots. To do this, they studied 102 major riots in the United States between 1967 and 1971. The hypothesis they wished to test, and for which they claimed confirmatory evidence, is the existence of a curvilinear relationship between the likelihood of a riot and the maximum ambient temperature at the time of the riot. This hypothesis contrasts with the prevalent folklore that riots tend to occur during periods of very hot weather. Specifically, Baron and Ransberger concluded that the likelihood of a riot increases with temperature up to the range of 81°–85° F and then decreases sharply with further increases in temperature. The evidence that they presented to support this relationship is a frequency distribution of the number of riots plotted against temperature. This frequency polygon does indeed peak in the interval 81°–85° F, falling off sharply on either side.

We contend that this relationship is an artifact of the particular way the data were examined and that an appropriate reanalysis suggests a monotonically increasing function relating the probability of riots and temperature. Basically, we argue that the Baron and Ransberger results stem from their having not taken account of base-rate differences in temperature. For example, if days in the 81°–85° F range are more common than days in the 91°–95° F range, there may well be more riots in the former range. There are, after all, many more opportunities for riots. But an appropriate analysis may well show that riots are *relatively* more common in the higher temperature range. To be sure, Baron and Ransberger did consider this possibility, but they rejected it. In our view, their rejection was premature; we consider their arguments and the weaknesses therein at greater length below.

The activities reported herein were supported in part by funds from Boys Town. However, the opinions expressed or the policies advocated do not necessarily reflect those of Boys Town. The research was carried out while the first author was a member of the Boys Town Center for the Study of Youth Development at Stanford University and while the second author was supported by a National Science Foundation Fellowship. We are indebted to Bradley Efron, Lee Ross, and Amos Tversky for their comments and suggestions on this work, and to Joy Davis for her assistance in the seemingly unending task of transcribing 58,000 temperatures.

Requests for reprints should be sent to J. Merrill Carlsmith, Department of Psychology, Stanford University, Stanford, California 94305.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3703-0337$00.75

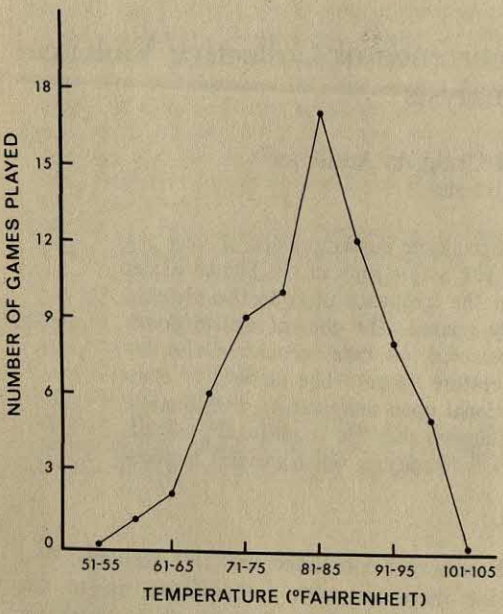


Figure 1. Frequency of baseball games played (New York Mets, 1977) as a function of ambient temperature.

Baseball, Temperature, and Base Rates

To see most clearly how this artifact may work, it is instructive to apply the same analysis used by Baron and Ransberger to a set of events that we know are not influenced by temperature. Figure 1 shows the same analysis used by Baron and Ransberger applied to the frequency of New York Mets baseball games played at home in 1977 (*The Sporting News*, April–October 1977). That is, we plot the frequency of Mets home games against the maximum ambient temperature in New York on the day of the game. (To be sure, baseball games occur primarily during the summer months, but then again, so do riots.) A brief study of Figure 1 shows a remarkable similarity to Figure 1 in the Baron and Ransberger (1978) article, and were we to follow their logic, we would have to conclude that "inspection of this figure lends support to the suggestion of a curvilinear relationship between ambient temperature and the incidence of [baseball games]" (p. 354). In view of the fact that baseball games are scheduled some months in advance, such a conclusion hardly seems warranted. Another explanation seems far more plausible. Both frequency polygons lead to erroneous conclusions, and for the same reason—the base-rate of different temperatures has not been taken into account. In our baseball example, it is not difficult to see that the frequency distribution of Mets home games is heavily influenced by the base-rate of daily maximum temperatures in New York during this period. Fewer games were played at temperatures of 91°–95° F, not because such temperatures lowered the probability of players choosing to play baseball, but because there were fewer such days. This simple example captures the essence of our critique of the Baron and Ransberger analysis, although the problem becomes a good deal more complex when we try to deal with the data they present.

The failure to consider base rates when assessing the probability of some event is hardly unique to this example. It is a problem of general concern in the analysis of data when we wish to calculate the probability of an event, conditional on the occurrence of some other event. It is also a problem that has begun to intrigue cognitive psychologists interested in subjective assessments of probability rather than formal statistical estimation. For example, in the context of a discussion of judgmental heuristics, Tversky and Kahneman (1974) comment:

> The reliance on heuristics and the prevalence of biases are not restricted to laymen. Experienced researchers are also prone to the same biases—when they think intuitively. For example, the tendency to predict the outcome that best represents the data, with insufficient regard for prior probability, has been observed in the intuitive judgments of individuals who have had extensive training in statistics. Although the statistically sophisticated avoid elementary errors, such as the gambler's fallacy, their intuitive judgments are liable to similar fallacies in more intricate and less transparent problems. (p. 1130)

Temperatures Before and After Riots (and Baseball Games)

The second major group of data presented by Baron and Ransberger is a plot of the maximum daily temperatures in the riot city during the 7 days prior to the occurrence of each riot and the 3 days following the riot. This plot shows gradually increasing tempera-

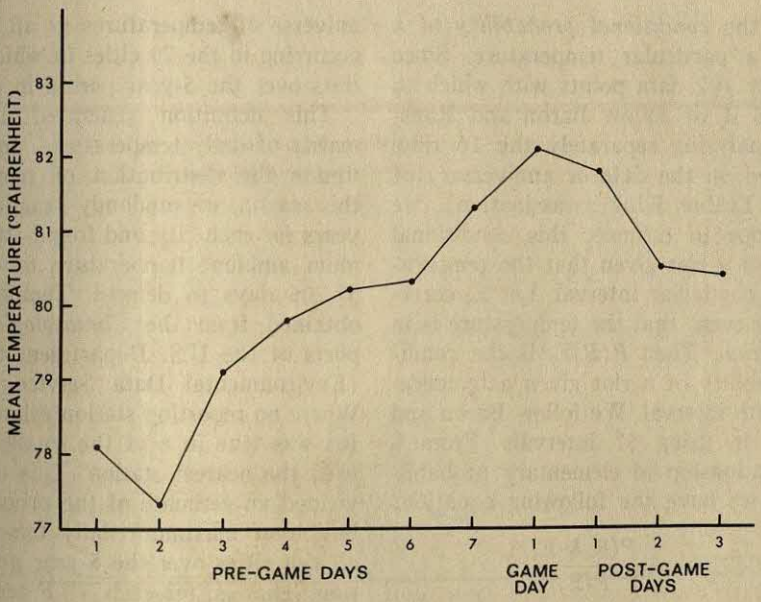


Figure 2. Mean maximum ambient temperature on the 7 days preceding the baseball game, on the game day, and on the 3 days following the game day.

tures up to the day of the riot, followed by decreasing temperatures during the 3 days following the riot. It should be noted that although this picture is consistent with their curvilinear hypothesis, it is also consistent with a wide variety of other possible functional relationships between temperature and the probability of riots. In particular, it is perfectly consistent with our hypothesis that there is a monotonically increasing relationship between temperature and the probability of a riot. On the other hand, the plot is equally consistent with a hypothesis that states that on occasional rainy (and cooler) days, riots (or baseball games) are unlikely to occur. Thus, the same picture could occur regardless of the true relationship between likelihood of a riot and temperature at the outbreak of a riot. Figure 2 shows a plot, again parallel to Baron and Ransberger's Figure 2, but done for New York Mets home games in 1977. Again we see a remarkable similarity between Baron and Ransberger's Figure 2 for riots and our Figure 2 for baseball games. It is our tentative hypothesis that our Figure 2 is mediated by a few rainy days, on which temperatures tend to be cooler and baseball games tend not to be played, but we are less certain of this artifact than we are of the artifact underlying Figure 1. In view of the fact that Baron and Ransberger's Figure 2 is consistent with a wide variety of possible functional relationships between temperature and the probability of riots, we will not pursue this issue further but return to the more fundamental question. If the postulated curvilinear relationship between temperature and the probability of riots is artifactual, what is the nature of the true relationship?

Conditional and Unconditional Probabilities

It is easier to point to the dangerous artifact underlying Figure 1 than to see a perfect solution to it. We present two alternative analyses below. Neither is immune to criticism, although both remove the obvious artifact. The consistency that emerges from these two very different analyses leads us to some confidence in the conclusions they imply, although we would emphasize at the outset our qualms about drawing any firm conclusions from this type of correlational analysis.

To see how to remove the effect of base rates, it is instructive to formalize our discussion slightly. The quantity we wish to

estimate is the *conditional probability* of a riot, given a particular temperature. Since we have only 102 data points with which to work (or 86 if we follow Baron and Ransberger in analyzing separately the 16 riots that occurred on the date or anniversary of the Martin Luther King assassination), we can only hope to estimate this conditional probability of a riot given that the temperature is in a particular interval. Let E_i correspond to the event that the temperature is in the ith interval. Then $P\{R|E_i\}$is the conditional probability of a riot given a temperature in the ith interval. We follow Baron and Ransberger in using 5° intervals. From a familiar relationship in elementary probability theory, we have the following equation:

$$P(R|E_i) = \frac{P(R \ \& \ E_i)}{P(E_i)}.$$

That is, the conditional probability of a riot, given that the temperature is in a particular interval, is given by the joint probability of a riot and a temperature in a particular interval, divided by the probability that the temperature is in that interval.[1] Examination of the equation makes it clear that Baron and Ransberger essentially estimated the joint probability without correcting for the marginal temperature distribution.

Our problem, then, is to estimate that marginal distribution. It is not such an easy problem as it might appear, since the universe from which the particular riot temperatures are to be viewed as a sample is not well defined. We might attempt to conceptualize it as the distribution of all temperatures in the United States in the 5-year period in question. But a moment's thought shows the vagueness of that conceptualization. Should that distribution be weighted by the population density in each geographical location? Does it include the temperatures at Death Valley, where there are too few people to stage a convincing riot? Alternatively, we might try the distribution of temperatures in cities larger than, say, 100,000 people, again over the 5-year period in question. But some riots occurred in much smaller cities. Furthermore, no riots occurred in Alaska. Should we then include Alaska in our universe? Our solution to this problem was to define the universe of temperatures as all temperatures occurring in the 79 cities in which there were riots over the 5-year period in question.[2]

This definition generated a 79 × 1,826 matrix of daily temperatures. In order to estimate the distribution of temperatures in this matrix, we randomly sampled 2 of the 5 years for each city and found the daily maximum ambient temperature for each of the 57,705 days so defined. Temperatures were obtained from the *Climatological Data* reports of the U.S. Department of Commerce (Environmental Data Service, 1967–1971). Where no reporting station existed in the city (as was true in 6 of the smaller towns), we used the nearest station. This method, then, yielded an estimate of the probability distribution of maximum daily temperature over all riot cities over the 5-year period in question. That is, for each 5° F temperature interval, we had a count of the number of days in which the maximum temperature was in that interval. Dividing that count by 57,705 (the total number of days sampled) yielded an estimate of the probability of a day in

---

[1] An urn model may serve to clarify this point. Suppose we imagine an urn with a large number of marbles, each marked with a temperature. Most of the marbles are white, but a few are blue. The blue marble corresponds to a riot. To estimate the marginal distribution of temperature, we draw a sample of marbles and plot the temperatures. To estimate the probability of a riot, we draw a sample of marbles and count the number of blue marbles. To estimate the joint probability of riots and temperature, we draw a sample of marbles and count the number of blue ones in each temperature range. To estimate the conditional probability of a riot given a particular temperature, we draw a large sample, count the number of blue marbles in a particular temperature range, and divide that count by the total number of marbles in that temperature range. This final number is interpreted as follows: If we are told that we have a marble in a particular temperature range, the conditional probability tells us the number of chances that it is blue. So we wish to know for our data, Given that the temperature is between 81° and 85° F, what is the conditional probability of a riot? Having estimated this conditional probability for each interval, we then ask how these conditional probabilities vary with temperature.

[2] We are indebted to Robert A. Baron for providing us with the list of cities and dates of the 102 riots used in their study.

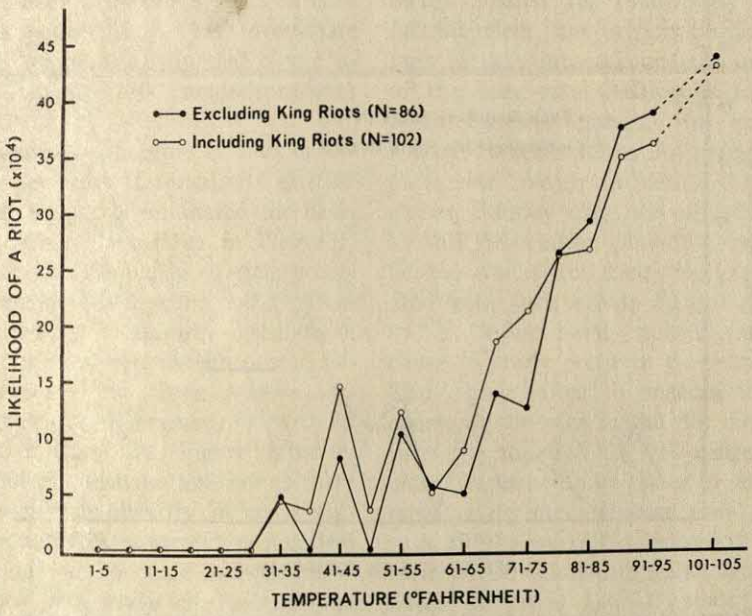


Figure 3. Conditional probability (likelihood) of a riot as a function of ambient temperature.

which the maximum temperature was in that interval. To estimate the conditional probability of a riot, given that the temperature was in a particular interval, we calculated the number of occurrences of a riot with a temperature in that range and divided by the number of days with a temperature in that range. The resulting function is shown in Figure 3.[3]

It is apparent that Figure 3 provides no evidence for a curvilinear relationship. Instead, there is a continuously increasing likelihood of riots as the temperature continues to rise, at least up through the temperature range of 91°–95° F. Contrary to the assertion of Baron and Ransberger, riots do not seem to be most likely at temperatures between 81° and 85° F; rather they become more and more likely with increasing temperature. This different function, of course, is a consequence of the fact that although there are fewer riots on days when the temperature is, say, 91°–95° F than on days when the temperature is 81°–85° F, there are many fewer days in the higher temperature range. The smaller number of riots at extremely high temperatures appears to be the result of many fewer opportunities for riots to occur; the conditional probability of a riot is larger at the higher temperature.

Three brief methodological notes are in order about Figure 3. First, there is a total of only three riots in the highest three temperature intervals, making the probability estimates extremely unstable. Consequently, we have averaged the three points, and this average is connected to the remainder of the function by a dashed line. Second, in defining the maximum daily temperature associated with each riot, we used the temperature on the day the riot began. This method contrasts with that of Baron and Ransberger, who took the average of the daily temperatures over the duration of the riot, for those riots that lasted more than 1 day. It seemed to us that if one wants to consider temperature as a causative factor in the outbreak of riots, it is more sensible to measure the temperature at the time of the outbreak rather than to include temperatures over subsequent days. Clearly, the temperature on

[3] We have followed Baron and Ransberger in estimating each function twice—once including the Martin Luther King-related riots and once excluding them.

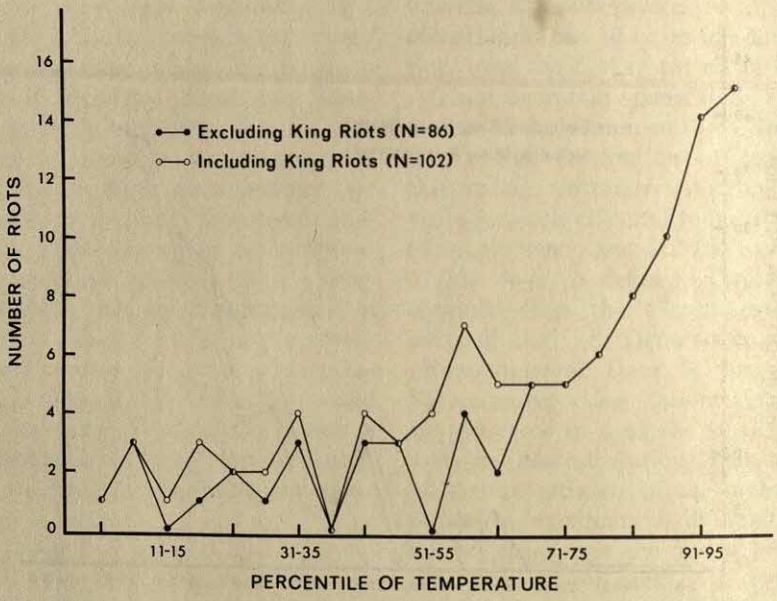


Figure 4. Frequency of collective violence (riots) as a function of the relative ambient temperature (expressed in percentiles).

July 12 cannot be more influential in determining whether a riot begins on July 10 than is the temperature on July 10. If we were to follow the Baron and Ransberger definition, there would be some minor changes in our Figure 3. Since we know that temperatures following the onset of a riot tend to be less extreme than temperatures on the first day of the riot, it is not surprising to find that extreme temperatures are slightly less common if we average over the days of the riot and that the proportion of riots at the modal temperature increases slightly. Even were we to use this averaging, the function still fails to show the precipitous drop that made Figure 1 so compelling, and the riot probability still reaches its maximum in the highest temperature range. Third, we present no inferential statistics in conjunction with this figure. It is our view that such statistics are, at best, irrelevant to these data and, at worst, seriously misleading. In their analysis of Figure 1, Baron and Ransberger present chi-square statistics. As we have already seen, such a calculation rests on an assumption that all temperature intervals are equally likely—an assumption that is demonstrably false. Furthermore, the riots show strong temporal and geographical dependencies (for example, five riots occurred in different cities in Michigan in a 3-day interval), which make assumptions of independence untenable.

An Alternative Analysis

Although Figure 3 casts severe doubts on the curvilinearity hypothesis, it is not totally convincing by itself. We have already alluded to the somewhat arbitrary definition of the universe of temperature days. Furthermore, the base-rate information used in the calculation of Figure 3 can be overly influenced by an extreme temperature distribution in one or two cities (although the large number of data points makes this somewhat implausible). Still, we are adding across rather dissimilar temperature distributions. A quite different analysis, which avoids these particular difficulties, involves looking at each riot temperature relative to all temperatures in the riot city. Thus, we estimate the cumulative distribution function (cdf) for all temperatures in City K and then look at the temperature on the day of the riot in that city relative to the cdf of all temperatures in that city. This procedure converts each riot temperature to a percentile relative to all temperatures in the riot city. Again, we use the

random sample of 2 of the 5 years in each city over the period 1967–1971 to estimate the cdf of all temperatures in that city, estimating each cdf by 730 (nonindependent) points.

Figure 4 shows the number of riots occurring at each percentile interval. By shifting to percentiles, we have eliminated the base-rate problem that plagued us in Figure 1. When we plot the frequency of riots occurring at different percentiles, the null hypothesis clearly predicts a uniform distribution. Any effect of the different likelihoods of different temperatures has been removed by expressing each riot temperature relative to all temperatures in that city. Figure 4 hardly suggests a uniform distribution, nor is there any evidence of curvilinearity. What is suggested is a monotonically increasing function. Thus, once again we conclude that the likelihood of a riot in a given city increases as the maximum ambient daily temperature in that city increases.

It should be noted that Figure 4 examines the covariation of riots and the relative temperature (relative to all temperatures in that particular city) rather than the riots' covariation with the absolute temperature. The relationship between relative and absolute temperature is strong enough (although by no means perfect) that we see no hope of using these data to answer the fascinating question of whether absolute temperature or relative temperature (relative to some adaptation level) is more important in predicting the likelihood of a riot. Rather, we see the convergence of these two functions, based on very different methods of analysis, as lending support to the general proposition that, at least for these particular riots, the probability of a riot increases monotonically with increasing maximum ambient temperatures in potential riot cities.

Attempts to Discount the Base-Rate Artifact

In view of the dramatic differences in conclusions that we draw after taking account of different temperature base rates, it is worth considering why Baron and Ransberger rejected this artifact as an explanation of their results and why we feel that the rejection was in error. They present several arguments against the plausibility of the artifactual effect, but each is particularistic and fails to take into account the overall impact of the base rate. Analogous to our use of Mets baseball games, for example, they looked at celebrations following victories in particular major athletic contests (bowl games, Stanley Cup play-offs, National Basketball Association play-offs, and the World Series) and found that temperatures associated with such events do not peak at 81°–85° F. Unfortunately, none of the events they chose to study occur in the summer months. Thus, they failed to observe that over the course of the year in the riot cities, temperatures in the 81°–85° F range are indeed more frequent than those in any other interval. The only evidence they present relevant to the overall distribution comes from their second counterargument against it. They select 11 cities and 2 months (July and August) and show that temperatures in the 81°–85° F range are not uniformly most frequent in those cities in those months. But different choices of cities, months, and temperature intervals lead to different conclusions, none of which describe the overall temperature distribution. As we have shown above, temperatures in the 81°–85° F range are in fact the most common temperatures in these cities. Baron and Ransberger's other arguments have this same particularistic quality, focusing primarily on temperatures in particular ranges in riot cities on the same dates in riot years versus nonriot years, and we do not consider them in detail here.

A Final Note

We close with a final set of cautionary remarks. We feel quite confident that these data do not provide support for the hypothesis of a curvilinear relationship between temperature and the probability of a riot. We feel reasonably confident that for these particular riots, there is good evidence for a monotonically increasing relationship between temperature and the probability of a riot.[4] However, facile generalizations from

[4] These remarks about the monotonic character of the relationship between temperature and the likeli-

these data make us very nervous. The riots are certainly not independent of one another; the dependencies cannot easily be described. The temperatures are not independent; again the dependencies are complex. Thus, from the point of view of inferential statistics, the number of true independent data points is unknown and may be small. Not only are the data fraught with all of the ambiguities of any correlational study, the data analyses are also subject to subtle, difficult, and complex effects. Even if the present data analysis seems satisfactory, there are numerous alternative explanations of the relationship. A clear understanding of the psychological effects of temperature, and particularly the effects of temperature on aggression, seems much more likely to emerge from experimental work like that of Baron (1972), Baron and Bell (1976), or Baron and Lawton (1972).

hood of riots are restricted to the normal range of temperatures. Clearly, at some point the relationship must become curvilinear. We seriously doubt that riots are likely to occur when the temperature is 120° F (although we have no data one way or the other).

References

Baron, R. A. Aggression as a function of ambient temperatures and prior anger arousal. *Journal of Personality and Social Psychology,* 1972, *21,* 183–189.

Baron, R. A., & Bell, P. A. Aggression and heat: The influence of ambient temperature, negative affect, and a cooling drink on physical aggression. *Journal of Personality and Social Psychology,* 1976, *33,* 245–255.

Baron, R. A., & Lawton, S. F. Environmental influences on aggression: The facilitation of modeling effects by high ambient temperatures. *Psychonomic Science,* 1972, *26,* 80–82.

Baron, R. A., & Ransberger, V. M. Ambient temperature and the occurrence of collective violence: The "Long, hot summer" revisited. *Journal of Personality and Social Psychology,* 1978, *36,* 351–360.

Environmental Data Service. *Climatological data.* Washington, D.C.: U.S. Department of Commerce, 1967–1971.

The Sporting News. April, 1977, through October, 1977.

Tversky, A., & Kahneman, D. Judgment under uncertainty: Heuristics and biases. *Science,* 1974, *185,* 1124–1131.

Received June 29, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 3, 345-356

Affective Space is Bipolar

James A. Russell
University of British Columbia, Vancouver, Canada

Numerous previous studies found monopolar rather than bipolar dimensions of affect (defined as emotion represented in language), but may have included methodological biases against bipolarity. The present study of self-report data ($N = 150$) on 11 affect scales showed that response format and acquiescence response style significantly shifted correlations between hypothesized opposites away from showing bipolarity. When these biases were taken into account, pleasure was found to be the bipolar opposite of displeasure and arousal of sleepiness. In turn, pleasure–displeasure and degree of arousal formed a two-dimensional bipolar space that accounted for almost all of the reliable variance in Thayer's four factors of activation plus a measure of depression. Dominance and submissiveness factors were also included in the study, but invalidity of the scales used precluded any conclusions regarding their bipolarity.

Early work on affect (emotion as represented in language) was based on the assumption that its descriptive dimensions would be bipolar: "An affective scale is a bipolar one" (Guilford, 1954, p. 264). Nowlis (e.g., Nowlis & Nowlis, 1956) similarly began his factor-analytic work on affect by hypothesizing four bipolar dimensions: pleasantness–unpleasantness, activation–deactivation, positive–negative social orientation, and control–lack of control. He gathered adjectives to measure each dimension and, in a series of studies, factor-analyzed the intercorrelations among responses to those adjectives. "How [did] these results relate to our original hypotheses? First, we had assumed that each factor would be bipolar. Actually, bipolarity was almost nonexistent in the obtained axes" (Nowlis, 1965, p. 358). His hypothesized opposites had tended to form two separate, independent monopolar factors instead of single bipolar ones.

Borgatta (1961), Clyde (1963), and McNair and Lorr (1964) subsequently reported the results of additional factor-analytic studies on affect in which monopolar, as opposed to bipolar, factors were consistently obtained. Thayer (1967) investigated what had originally been hypothesized to be a single, bipolar affective dimension, degree of activation or arousal, and obtained four independent monopolar factors, which he termed *general activation, high activation, deactivation-sleep,* and *general deactivation.* Indeed, most investigators of affect now argue against the existence of bipolar affective dimensions (e.g., Bradburn, 1969; Hall, 1977; Izard, 1972; McLachlan, 1976; Westbrook, 1976).

In other words, considerable research has tended to indicate that a person can feel, for example, both happy and unhappy or both aroused and sleepy at the same time. Based on these results, the affect scales most commonly used today were constructed to assess monopolar factors (e.g., Izard, 1972; McNair, Lorr, & Droppleman, 1971; Nowlis, 1965; Thayer, 1967), although several scales continue to be based on bipolar dimensions (e.g., Mehrabian & Russell, 1974; Spiel-

I am grateful to Lesley Walton, Klaus Schroeder, Nicole Clement, and Doreen Ridgeway for their help in carrying out this study, to James H. Steiger and A. Ralph Hakstian for their help in data analysis, and to Jerry S. Wiggins and Auke Tellegen for helpful comments on an earlier version of this article.

Send reprint requests to James A. Russell, Department of Psychology, University of British Columbia, Vancouver, British Columbia, Canada V6T 1W5.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3703-0345$00.75

berger, Gorsuch, & Lushene, 1970; Zuckerman & Lubin, 1965).

Bipolarity of Semantic Space

Interestingly, a similar controversy arose in the semantic differential literature. Osgood, Suci, and Tannenbaum (1957) found that the affective meaning of words could be adequately summarized by three bipolar dimensions: evaluation, activity, and potency. Since each of these dimensions was defined by adjective-pair scales such as good–bad, fast–slow, and strong–weak, respectively, Green and Goldfried (1965) argued that bipolarity had been assumed, indeed had been forced onto the results, by the rating procedure employed. Green and Goldfried then attempted to empirically test the assumption of bipolarity by constructing a single-adjective version of Osgood's semantic differential scales and examining the actual correlation between alleged opposites. If evaluation, say, were indeed bipolar, one would expect a high negative correlation between responses to negative evaluation words such as *bad* and those to positive evaluation words such as *good*. In fact, Green and Goldfried generally failed to find such correlations and concluded that semantic space should *not* be conceptualized as bipolar. If accepted, such a conclusion would have important implications not only for the widely used semantic differential technique but for the area of affect as well, in that the semantic differential factors have been interpreted as reflecting basic dimensions of human affect (Osgood, 1969).

Green and Goldfried's (1965) argument against bipolarity of semantic space did not go unchallenged, however. Bentler (1969) argued that high negative correlations indicative of bipolarity might have been obscured in their study by acquiescence, which is an individual-difference variable in the tendency to agree or disagree with an item regardless of its content. That is, if acquiescence were a factor in Green and Goldfried's data, it would have shifted all observed correlations between words in a positive direction and hence away from showing bipolarity ($r = -1.0$). Bentler tested his hypothesis by constructing a single-adjective version of the semantic differential and using a partial correlation technique to control acquiescence. The results provided striking evidence for his hypothesis. For example, a composite of positive evaluation words correlated only .03 with a composite of negative evaluation words—before acquiescence was taken into account. After acquiescence was partialed out, these composites correlated −.76, thus giving strong evidence for the bipolarity of semantic space.

Bipolarity of Affective Space

In the domain of affect, Meddis (1972) has recently defended the notion of bipolar dimensions. He noted two potential sources of bias in the response format used by Nowlis (1965), and subsequently by Thayer (1967), that could account for their failure to obtain bipolar dimensions. The response rating scale used by Nowlis and Thayer, shown in Table 1, assigns 1 point to "No (definitely not)," 2 points to "? (cannot decide)," 3 points to "V (feel slightly)," and 4 points to "VV (definitely feel)." First, Meddis noted, this scale is asymmetric, in that there are two categories of acceptance (V and VV) but only one of rejection (No), and, second, the option "? (cannot decide)" may preclude even ordinal data, since a response to that option could mean a variety of things other than a feeling whose intensity is somewhere between the two neighboring categories.

Meddis (1972) empirically contrasted the results obtained using Nowlis's and Thayer's response format with results obtained using a format that was more symmetric and that excluded the "cannot decide" alternative. (This last format is shown in Table 1 as the "Meddis" format.) Responses to affective adjectives were gathered with each of these formats separately and factor-analyzed. As predicted, predominantly monopolar factors tended to emerge from the data gathered with Nowlis's and Thayer's format, whereas bipolar factors tended to emerge from the data gathered with the Meddis format. When, for example, responses to 38 adjectives similar to those studied by Thayer (1967) were gathered with the Meddis format, bipolar factors were obtained: general activation ver-

sus deactivation-sleep and high activation versus general deactivation. In Meddis's (1972) words, "Whether or not we obtain bipolar factors appears to depend upon the kind of rating scale we use" (p. 183).

Additional Factors Operating Against Bipolarity

Many of the studies supporting monopolarity reviewed above did not employ Nowlis's and Thayer's response format and therefore are not subject to Meddis's specific criticisms. The conclusion of monopolarity in these additional studies may be questioned on the grounds of other methodological problems, however, such as the possibility of alternate rotations and of the influence of difficulty factors (unequal response distributions of the variables involved) in factor-analytic solutions. More importantly, these studies may also have included biases specifically against bipolar factors. First, to the extent that the sample of emotion words studied underrepresents one end of a bipolar continuum, bipolar factors are less likely to emerge. Second, if the instructions ask subjects how they felt over an extended period of time (such as over a week; e.g., Bradburn, 1969; Hall, 1977), they may describe several, perhaps opposite, emotional experiences. Finally, the correlation between two items of a scale has been shown to be spuriously inflated in proportion to the proximity of the items to each other in time or space (Guilford, 1954; Stockford & Bissell, 1949). This phenomenon, known as proximity error, would be expected to result in an overall shift toward more positive correlations among a set of items—irrespective of content—to which subjects responded within a short period of time, which of course is standard practice (and, it would seem, a necessity) when measuring affect.

In short, although numerous empirical studies of affect have concluded against bipolarity, results from these studies may have been biased by one or more of the following: an acquiescence response bias; a response format that fails to yield ordinal data; an asymmetric response format; inadequate sampling of affect terms; instructions to describe feelings during a time period long enough that several, possibly opposite, feelings might have occurred; and a proximity error, which is the tendency to respond similarly to items that are nearer in time or space.

Overview of the Present Study

In the present study, subjects described their current affective state on separate scales assessing opposite ends of hypothesized bipolar dimensions so that correlations between these scales could be empirically established. Four response formats (shown in Table 1) were employed: Meddis's (1972) format plus three others that have been extensively used in studies obtaining monopolar factors of affect, Nowlis's (1965) and Thayer's (1967), McNair and Lorr's (1964), and a true–false format.

To ensure an adequate sample of affect terms, items were chosen for the questionnaire that appeared on a priori grounds to assess primarily opposite ends of pleasure–displeasure, arousal–sleepiness, and dominance–submissiveness, since these dimensions have been suggested by a variety of sources of evidence as basic dimensions of affect (see Russell, 1978, for a brief review of this evidence). All adjectives from Thayer's (1967) four scales of activation were also included in order to explore the activation/arousal dimension more thoroughly. Thayer (1970) showed his scales to be significantly related to physiological indexes of arousal, and the relation of his four factors to the entire affective space is thus of considerable theoretical importance. Moreover, evidence provided by Russell and Mehrabian (1977) indicated that it was possible to subsume Thayer's four dimensions within a more general, three-dimensional theory of affect defined by pleasure–displeasure, degree of arousal, and dominance–submissiveness. The present study provided further evidence on that notion and tested Meddis's (1972) hypothesis that Thayer's general activation is the bipolar opposite of deactivation-sleep and that high activation is the bipolar opposite of general deactivation. Finally, a set of adjectives that

Table 1
Response Probability for Each Category of Four Response Formats

Format	Label			
Meddis (1972)	XX (definitely do not feel) .17	X (do not feel) .41	V (slightly feel) .34	VV (definitely feel) .08
Nowlis (1965) & Thayer (1967)	No (definitely do not feel) .45	? (cannot decide) .15	V (feel slightly) .32	VV (definitely feel) .08
McNair & Lorr (1974)	1 (not at all) .41	2 (a little) .36	3 (quite a lot) .18	4 (extremely) .05
True–false	N (no, I do not feel) .63	Y (yes, I do feel) .37		

Note. Probabilities were calculated across 58 items and 150 subjects for each response format.

appeared to assess depression was included, since on inspection of Thayer's items, depression seemed more likely to be the actual bipolar opposite of his general activation scale.

Method

Subjects

Subjects were 71 female and 79 male University of British Columbia undergraduates. These were students from two different psychology classes who responded to a series of four questionnaires during class time, each questionnaire on a separate day, approximately 2 weeks apart. One class responded in the fall of 1976, the other in the following spring. Ten additional subjects began, but failed to complete the study, and their data were excluded from analysis.

Materials

Eleven sets of adjectives, a total of 58 items, were used. The first 4 sets are Thayer's (1967) factors of activation, and the remaining 7 were constructed a priori to measure pleasure, displeasure, arousal, sleepiness, dominance, submissiveness, and depression. The sets were the following:

1. *General activation* (Thayer): lively, active, full of pep, energetic, peppy, vigorous, activated.
2. *High activation* (Thayer): clutched up, jittery, stirred up, fearful, intense.
3. *General deactivation* (Thayer): at rest, still, leisurely, quiescent, quiet, calm, placid.
4. *Deactivation-sleep* (Thayer): sleepy, tired, drowsy.
5. *Pleasure:* contented, happy, satisfied, pleased, joyful.
6. *Displeasure:* discontented, unhappy, dissatisfied, displeased, joyless.
7. *Arousal:* wide awake, aroused, aflame, impassioned, alert, roused.
8. *Sleepiness:* inactive, half asleep, slow, unaroused.
9. *Dominance:* dominant, controlling, influential, important, autonomous.
10. *Submissiveness:* submissive, controlled, influenced, awed, guided.
11. *Depression:* depressed, discouraged, gloomy, sad, blue, sluggish.

Four questionnaires were formed by intermixing these 58 items in two random orders and then counterbalancing each. Each questionnaire began with directions stating, "On this sheet you will find a list of words or phrases that describe different kinds of moods and feelings. Please use this list to describe your feelings *today*."

Next followed one of the four response formats (shown in Table 1). The true–false format allows two response alternatives; the remaining three formats allow four response alternatives and differ only in the verbal label attached to each alternative. For convenience, these three formats were labeled in the table by the name of investigators who introduced, or at least have used, that format in studies of affect.

Procedure

Each subject responded to the 58-item questionnaire at the beginning of class and returned it to an envelope; all envelopes were collected. Altogether, each subject responded four times, once to each of the four response formats in random order. Each subject responded only once, again in random order, to each of the four orders of presentation of the 58

items. In other words, all possible combinations of response format with order of item presentation were employed and randomly assigned to subjects. No feedback from their data was given to subjects until after the last questionnaire had been completed.

Results and Discussion

Response Format

Table 1 shows the mean probability of a response to each category of each response format. The Meddis format resulted in a roughly even, symmetric distribution of responses. The Nowlis and Thayer format, in contrast, resulted in a bimodal distribution, because subjects avoided the category labeled "? (cannot decide)"—a finding that supports the view that this category may preclude ordinal data. The McNair and Lorr format resulted in a skewed distribution, because the modal response was to the first category, "not at all." Thus, of the four formats studied, the Meddis format provided the best distribution of responses and may be preferred on these grounds alone.

It had been hypothesized that in comparison with the Meddis format, the other formats would shift correlations among items in the positive direction. A sign test was thus used in which each interitem correlation from the Meddis format was compared with its counterpart from each of the other three formats. As expected, correlations based on the other formats did tend to be greater than their counterparts based on the Meddis format: In the Nowlis and Thayer format data, 649 correlations were smaller and 1,004 were greater ($z = 8.71$, $p < .01$); in the McNair and Lorr format data, 684 were smaller and 969 greater ($z = 6.98$, $p < .01$); and in the true–false format data, 500 were smaller, and 1,153 greater ($z = 16.04$, $p < .01$). In addition, the true–false format resulted in more correlations that were greater than their counterparts not only from the Meddis format but from the others as well; Compared with the Nowlis and Thayer format data, 559 were smaller and 1,094 were greater ($z = 13.13$, $p < .01$); and compared with the McNair and Lorr format data, 641 were smaller and 1,012 were greater ($z = 9.10$, $p < .01$). There was no significant difference between correlations from the Nowlis and Thayer format and those from the McNair and Lorr format (864 smaller and 789 larger, $z = 1.82$, *ns*). In sum, when compared with the Meddis format, the three formats commonly used in studies of affect resulted in correlations shifted in the positive direction—away from evidence of bipolarity.

Content Scales

Reliability. Since 7 of the 11 scales were constructed ad hoc for this study, their internal consistency was unknown. Rather than use an extensive item-selection process, however, it was decided to simply eliminate any item from a scale if it failed to correlate positively with each of the other items on that scale. Based on data gathered with the Meddis format ($N = 150$), only one item—"controlled" on the submissiveness scale—failed this criterion, and that item was excluded from further analyses. Measures of internal consistency reliability (coefficient alpha, Cronbach, 1951) were then calculated for all 11 scales, separately for each format. (Values for 5 scales are reported later in Table 3.) Three scales showed only moderate reliability: high activation (.59 to .69), dominance (.68 to .73), and submissiveness (.56 to .65). But, all other scales showed adequate reliability (.70 to .95), and at this point, all scales were retained for further analysis.

Product-moment correlations. Table 2 gives zero-order correlations of 6 scales with hypothesized bipolar opposites, separately for each response format. Correlations between pleasure and displeasure ranged from −.43 with the true–false format to −.71 with the Meddis format. Similarly, correlations between arousal and sleepiness ranged from −.45 with the true–false format to −.62 with the Meddis format. Correlations between dominance and submissiveness were considerably lower and mostly positive, ranging from .27 with the true–false format to −.05 with the McNair and Lorr format. Correlations between Thayer's high activation and its hypothesized opposite, general

Table 2
Association Between Scales and Their Hypothesized Opposites

Scale	Hypothesized opposite	Meddis (1972) r	Meddis (1972) Partial r	Meddis (1972) Partial AQ	Nowlis (1965) & Thayer (1967) r	Nowlis (1965) & Thayer (1967) Partial r	Nowlis (1965) & Thayer (1967) Partial AQ	McNair & Lorr (1964) r	McNair & Lorr (1964) Partial r	McNair & Lorr (1964) Partial AQ	True–false r	True–false Partial r	True–false Partial AQ
Pleasure	Displeasure	−.71*	−.78*	#	−.66*	−.80*	#	−.66*	−.78*	#	−.43*	−.71*	#
Arousal	Sleepiness	−.62*	−.71*	#	−.62*	−.76*	#	−.55*	−.72*	#	−.45*	−.66*	#
Dominance	Submissiveness	.12	.05		.13	.02	#	−.05	−.19	#	.27*	.06	#
General activation	Deactivation-sleep	−.53*	−.50*		−.58*	−.66*	#	−.50*	−.55*	#	−.38*	−.52*	#
General activation	Depression	−.58*	−.60*		−.42*	−.50*	#	−.39*	−.56*	#	−.20	−.43*	#
High activation	General deactivation	−.31*	−.36*	#	−.16	−.30*	#	−.16	−.25*	#	−.14	−.34*	#

Note. r refers to zero-order correlations between the two scales listed. Partial r refers to a partial correlation with acquiescence partialed. The column AQ indicates (by the appearance of a number sign, #) that acquiescence correlated significantly ($p < .01$) with the first scale listed partialing the hypothesized opposite. $N = 150$.
* $p < .01$.

deactivation, were negative, but low and mostly nonsignificant. There were two hypothesized opposites for Thayer's general activation scale: Thayer's deactivation-sleep scale and the depression scale. Correlations between high activation and these two hypothesized opposites were negative and mostly significant, but again modest in size.

To test the hypothesis that response format influenced the correlation between hypothesized opposites, each of the six correlations from the Meddis format was compared with its counterpart from each of the other three formats.[1] Two pairs of correlations were found to be significantly different: (a) The correlation between pleasure and displeasure (−.71) from the Meddis format differed from that (−.43) from the true–false format ($z = 3.85$), and (b) the correlation between general activation and depression (−.58) from the Meddis format differed from that (−.20) from the true–false format ($z = 4.08$). Thus, the true–false response format resulted in two significant differences, both in the expected direction of a more positive correlation than that produced by the Meddis format. In the other four comparisons between these two formats, the differences were in the same direction, although not significant. The Meddis format did not, however, produce any significantly more negative correlations between opposites in comparison with either the Nowlis and Thayer or the McNair and Lorr formats. Indeed, as shown in Table 2, although 9 of these 12 comparisons were in the expected direction, there were no large differences among corresponding correlations from the four-place formats. Contrary to Meddis's (1972) hypothesis, these results suggest that the Nowlis and Thayer format contributes at most a modest bias against obtaining bipolar dimensions.

The Role of Acquiescence

Partial correlations. A clearer picture of the relationship between hypothesized opposites emerged when acquiescence was taken into account through partial correlation. For each pair of scales within a given format, an acquiescence score was calculated for each subject by summing his or her responses to all items (many of them antonyms), except those items constituting the pair of scales to be correlated. For example, the acquiescence score partialed from the correlation between

[1] A separate test was carried out for each of the six pairs of scales listed in Table 2 for a total of 18 comparisons. Each test was conducted at the .001 level (one-tailed) so that the probability of a single spuriously significant result would be kept to approximately .018 (18 × .001). The test used here for the difference between two correlations based on the same sample is discussed in Olkin (1967), but the formula is misprinted there. The correct equation was supplied by James H. Steiger through personal communication.

pleasure and displeasure consisted of responses to 48 items, which were all the items except those for pleasure or displeasure. A partial correlation was then calculated between each pair of hypothesized opposites, partialing the appropriate acquiescence score. These results are also shown in Table 2.

In all but one case (general activation/deactivation, Meddis format), partialing acquiescence shifted the correlation in the expected, negative direction, that is, in the direction supporting bipolarity. In order to assess whether acquiescence contributed significantly to the variance in the content scales, another partial correlation was calculated: the correlation of acquiescence with the first scale listed in the table, partialing its hypothesized opposite. A number sign (#) is shown in Table 2 in the column labeled AQ (acquiescence) whenever this partial correlation was significant at the .01 level, which occurred in 21 of the 24 cases. Interestingly, all three exceptions occurred with the Meddis-format data, suggesting that this format is less subject to acquiescence bias than the others. In contrast, the results for the true–false format were not only significant but sizable, suggesting that this format is particularly subject to acquiescence bias.

Pleasure and arousal. Returning to the partial correlations between content scales, we can see from Table 2 that for pleasure and displeasure, these partial correlations were quite substantial—negative and large in magnitude, ranging from −.71 to −.80. This was the case even for the data from the true–false format, in which was seen the most dramatic change due to partialing acquiescence. Similar results were obtained for arousal and sleepiness, for which the partial correlations ranged from −.66 to −.76. Thus, for these two dimensions, strong evidence of bipolarity was obtained by partialing acquiescence—regardless of which response format was involved.[2]

Dominance. Partialing acquiescence from the correlations between dominance and submissiveness shifted these correlations in the expected negative direction. Nevertheless, the final results, ranging from .06 to −.19, did not differ significantly from zero and provided no evidence of bipolarity. In light of the results for pleasure and displeasure and for arousal and sleepiness, this finding is puzzling. Its explanation seems to be that there was little valid variance in either of the two scales, but especially submissiveness. First, reliabilities for these scales were low. Second, a substantial proportion of the reliable variance appears to have been actually acquiescence response style: Correlations of dominance with an acquiescence score (based on all items except dominance or submissiveness ones) ranged from .20 (Meddis format) to .42 (true–false format); correlations of submissiveness with the same acquiescence score ranged from .32 (McNair and Lorr format) to .52 (true–false format). In short, a considerable proportion of the variance in each scale was either unreliable or invalid, leaving so little valid variance that meaningful conclusions are precluded.

Thayer's scales and depression. Thayer's (1967) four scales yielded intermediate results. The partial correlations for general activation and its two hypothesized opposites, depression and deactivation-sleep, ranged from −.43 to −.66, indicating a consistently negative relationship, although not a substantial one. The partial correlations for high

[2] Since my concern was with the bipolarity of theoretical dimensions rather than with practical prediction problems, a correction for attenuation (unreliability) provided a useful estimate of what these partial correlations would be if the two scales correlated were perfectly reliable. The pleasure–displeasure partial correlation from the Meddis–format data (−.78) was thus corrected for unreliability (as estimated by alpha coefficients) in both the pleasure and displeasure scales; the result was −.88. A similar procedure for the arousal–sleepiness partial correlation (−.71) yielded an estimate of −1.02. Such corrected correlations must be interpreted with caution, because the use of coefficient alpha as an estimate of reliability can overcorrect the correlation (as obviously happened in the case of arousal–sleepiness). Also, it is not a common practice to disattenuate *partial* correlations, although doing so would appear to be biased on the conservative side, since after partialing reliable variance, error variance would constitute a relatively greater proportion of the remainder. Nevertheless, these results at least reinforce the view that as various sources of bias and error are removed, the correlation between these hypothesized opposites approaches −1.00.

Table 3
Multiple Regression Results and Reliability Coefficients for Thayer's Four Scales and Depression

Criterion	Response format	Standardized regression weight			R	Reliability for criterion
		Pleasure	Arousal	Acquiescence		
General activation	Meddis[a]	.32	.67	.08	.84	.91
	Nowlis[b] & Thayer[c]	.19	.75	.16	.84	.92
	McNair & Lorr[d]	.30	.72	.16	.87	.95
	True–false	.15	.67	.24	.80	.93
High activation	Meddis	−.64	.38	.32	.68	.65
	Nowlis & Thayer	−.39	.40	.37	.57	.59
	McNair & Lorr	−.52	.37	.35	.68	.69
	True–false	−.42	.35	.55	.72	.66
General deactivation	Meddis	.43	−.69	.14	.57	.78
	Nowlis & Thayer	.35	−.58	.33	.49	.76
	McNair & Lorr	.27	−.60	.34	.48	.71
	True–false	.28	−.59	.34	.47	.81
Deactivation-sleep	Meddis		−.66	.23	.66	.88
	Nowlis & Thayer		−.83	.29	.77	.90
	McNair & Lorr		−.80	.29	.71	.92
	True–false		−.81	.42	.74	.88
Depression	Meddis	−.82	−.16	.22	.90	.88
	Nowlis & Thayer	−.79	−.22	.33	.87	.84
	McNair & Lorr	−.74	−.28	.40	.84	.90
	True–false	−.69	−.23	.51	.81	.84

Note. Each regression equation was based on 150 cases. All values of regression weights and multiple correlations that are entered were significant at the .01 level. Reliability of each criterion was estimated by Cronbach's (1951) coefficient alpha measure of internal consistency.
[a] 1972.
[b] 1965.
[c] 1967.
[d] 1964.

activation and its hypothesized opposite, general deactivation, ranged from −.25 to −.36, indicating a more modest but still negative relationship. Thus, although inversely related, Thayer's scales did not form exact bipolar opposites as Meddis (1972) had predicted.

A Bipolar Affective Space

Regression analyses. The results so far were far clearer for pleasure and displeasure and for arousal and sleepiness than for Thayer's four scales and depression. A series of multiple regression analyses helped to clarify the relationships among the latter five scales and their relationship to pleasure–displeasure and arousal–sleepiness as well. Multiple regression analysis was used specifically to test the hypothesis that pleasure–displeasure and arousal–sleepiness can adequately account for Thayer's four scales and depression.

To define variables for the equations, scores on all scales were first standardized ($M = 0$; $SD = 1$). Displeasure scores were then subtracted from pleasure scores, and sleepiness from arousal, to form pleasure–displeasure and degree-of-arousal predictor variables. For the third predictor variable, a separate acquiescence score was computed for each regression equation by summing scores on all items other than the items on those scales already part of that equation

eparate multiple regression equations were hen computed from data for each response ormat for five criterion variables: each of hayer's four scales plus depression. The esults are shown in Table 3.

Multiple correlation coefficients were shrunken" to provide estimates of the cor- elations in the population between the scores n the criterion predicted from the equation nd those actually obtained. These multiple orrelations were substantial and significant ($p < .01$) in every case. Since unreliability n either the criterion or the predictors limits he possible magnitude of the multiple cor- elation coefficient, the last column of Table 3 provides reliability estimates for each cri- terion variable. As can be seen, the multiple correlations for all scales except Thayer's general deactivation generally approached the limit set by the reliabilities.

A reinterpretation of Thayer's activation dimensions. The standardized regression weights shown in Table 3 provided definitions for the five criterion variables that consider- ably clarified their meaning. Approximately the same information is also shown graphi- cally in Figure 1, which presents loadings of the 11 affective scales on their first two prin- cipal components, since these components

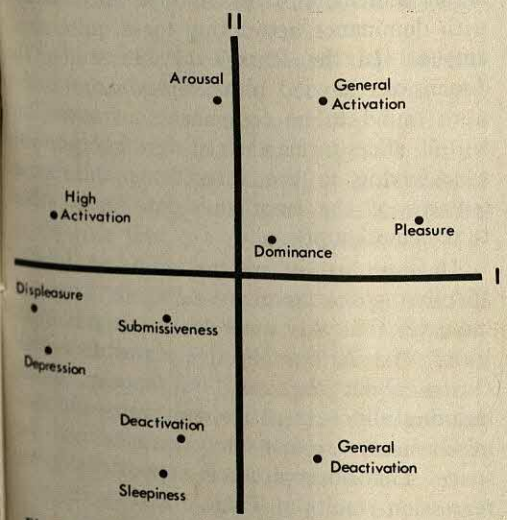


Figure 1. First two principal components of the 11 affective scales, based on data gathered with the Meddis (1972) response format ($N = 150$), after an orthogonal varimax rotation.

were readily interpretable as pleasure–dis- pleasure and degree of arousal. Both analy- ses showed Thayer's high activation and gen- eral activation to contain the expected high degree of arousal. The difference between these two scales, however, was as much the amount of pleasure as it was the degree of arousal, with general activation involving pleasure and high activation involving dis- pleasure. Similarly, Thayer's deactivation- sleep and general deactivation scales con- tained the expected low arousal, but again the two scales differed on pleasure, with gen- eral deactivation involving pleasure but de- activation-sleep involving only arousal. In short, Thayer's (1967, 1970) factor-analytic results were multidimensional not because the arousal continuum is multidimensional but because (among other things) his items varied on the pleasure dimension as well as on arousal.

In Figure 1, general activation is shown to be the *approximate* bipolar opposite of both deactivation-sleep and depression (as was suggested by the partial correlations of Table 2). General activation fell within the high pleasure/high arousal quadrant. Its hy- pothesized opposites, deactivation-sleep and depression, both fell within the opposite quadrant (displeasure/low arousal), but neither fell exactly opposite general activa- tion. The regression equations of Table 3 similarly showed deactivation-sleep to involve too little, and depression too much, displea- sure to be the precise bipolar opposite of general activation, thus explaining the mod- erate magnitude of their negative correlations shown in Table 2.

Similarly, Thayer's high activation fell within the displeasure/high arousal quadrant, and its hypothesized opposite, general deac- tivation, fell within the opposite quadrant, high pleasure/low arousal. These two scales were only slightly negatively correlated rather than exact bipolar opposites, how- ever, because each is maximally saturated with a different factor: High activation is mainly displeasure with a moderate compo- nent of arousal; general deactivation is mainly low arousal with a moderate com- ponent of pleasure.

Even though Thayer's four scales did not form exact bipolar opposites, the results summarized in Figure 1 and Table 3 show that they are part of a bipolar space. Presumably, their exact bipolar opposites could be found if adequate terms were included in the sample of terms studied. From this point of view, Thayer's scales are considerably mislabeled. Since both general activation and general deactivation involve pleasure, "general" could be more precisely replaced with "pleasant." More specifically, general activation assesses high pleasure combined with high arousal and, with items such as *lively* and *vigorous*, could be relabeled "vigor" or "excitement." General deactivation assesses high pleasure combined with low arousal and, with items such as *calm* and *placid*, could be relabeled "calmness" or "relaxation." Deactivation-sleep (with items *sleepy, tired, drowsy*) is appropriately labeled, but high activation assesses displeasure combined with high arousal and, with items such as *jittery* and *clutched up*, could be relabeled "distress."

Implications

As the obscuring influence of methodological biases was reduced in this study, it could be seen that a *bipolar affective space*, defined by pleasure–displeasure and degree of arousal, was able to represent the relationships among scales of pleasure, displeasure, arousal, sleepiness, depression, and Thayer's (1967) four factors of activation. Evidence from a previous study (Russell & Mehrabian, 1977) showed that an even greater range of affective states could be represented in the same way. In a series of regression analyses similar to those shown in Table 3, pleasure–displeasure and degree of arousal were found to account for almost all of the reliable variance in a sample of 42 commonly used scales of affect (e.g., scales of happiness, elation, anger, fear, anxiety, and depression). In another study (Russell, 1978), it was shown that the same dimensions of pleasure and arousal have emerged as the two major dimensions of affect from studies of verbal self-report, semantic differential ratings, successive-intervals scaling, and multidimensional scaling of affect terms. In other words, diverse sources of evidence are converging on a representation of affect similar to that depicted in Figure 1.

This proposed bipolar affective space may appear incomplete, however, because it is characterized by only two dimensions. Indeed, evidence has consistently confirmed the existence of dimensions of affect independent of pleasure and arousal. For example, in the Russell and Mehrabian (1977) study, dominance–submissiveness accounted for a significant proportion of the variance in scales of affect beyond that accounted for by pleasure and arousal. In the Russell (1978) study, multidimensional scaling of affect terms was used to examine more closely the dimensions of affect beyond pleasure and arousal. Results provided evidence for at least three such dimensions, interpreted as dominance/potency/control, depth or importance, and locus of causation.

Nevertheless, I would argue against the use of more than two orthogonal dimensions in a conceptualization of affect. The dimensions beyond pleasure and arousal have consistently been found to differ from pleasure and arousal in several important ways. In the Russell and Mehrabian (1977) study, pleasure and arousal accounted for by far the major proportion of variance in affect scales, with dominance accounting for a quite small amount. In the Russell (1978) study, the dimensions beyond pleasure and arousal (a) were found to be components of some, but not all, affect terms and (b) were interpretable as referring to the antecedents and consequences of the emotional state rather than to the emotion per se.

An examination of the proposed bipolar affective space as represented in Figure 1 indicates that this space lacks "simple structure." The various affective states do not all cluster about the axes, but appear to fall meaningfully (except for dominance and submissiveness) around the perimeter of the space. This observation is supported by the regression results of Table 3 and by those of Russell and Mehrabian (1977), in which affective states were typically found to be defined as *combinations* of pleasure–displeasure and degree of arousal. It may be arbitrary

erefore, where the axes are located within is space. For example, a 45° rotation of e axes shown in Figure 1 would yield two polar axes that could be labeled "excite-ent–depression" and "distress–relaxation." lore generally, the placement of the various fective states in relation to each other within e space appears to be more meaningful than e placement of the axes within the space.

The proposed bipolar affective space raises nother issue concerning the conceptualiza-ion of affect. The proposed space is based n a definition of the domain of affect that ncludes low arousal states, such as sleepi-ess and relaxation, which are not commonly ncluded on psychologists' lists of the emo-ions. For example, Tellegen [3] defines affect s *aroused feelings*. His proposed conceptual-zation of affect, which involves two dimen-sions, positive affect and negative affect, would therefore be expected to fall solely within the high arousal quadrants of Figure 1. And, indeed, his two dimensions appear to roughly correspond to the rotated dimensions of excitement and distress, respectively, de-fined above. The bipolar opposites of Tel-legen's dimensions, however, fall outside the domain of aroused feelings. Tellegen's dimen-sions therefore ought *not* to be bipolar within that domain, and this appears to be his con-clusion.

The present evidence showed that an em-pirical relationship exists between high and low arousal states and, therefore, that it is meaningful to group the two together within the same domain. The Russell and Mehrabian (1977) study suggested the same conclusion by showing that low arousal states were de-finable by the same pleasure and arousal di-mensions required to define the high arousal states. From my point of view, a fully ade-quate description of a person's affective state at any moment in time therefore must in-clude the possibility of these more mundane, but frequently encountered, low arousal states as well as the more dramatic emotions.

[3] Tellegen's conceptualization is described in Hall (1977) and was further elaborated through per-sonal communication. Needless to say, only the roughest approximation to his conceptualization is presented here.

References

Bentler, P. M. Semantic space is (approximately) bipolar. *Journal of Psychology*, 1969, *71*, 33–40.

Borgatta, E. I. Mood, personality, and interaction. *Journal of General Psychology*, 1961, *64*, 105–137.

Bradburn, N. M. *The structure of psychological well-being*. Chicago: Aldine, 1969.

Clyde, D. J. *Manual for the Clyde Mood Scale*. Coral Gables, Fla.: Biometric Laboratory, University of Miami, 1963.

Cronbach, L. J. Coefficient alpha and the internal structure of tests. *Psychometrika*, 1951, *16*, 297–334.

Green, R. F., & Goldfried, M. R. On the bipolarity of semantic space. *Psychological Monographs*, 1965, *79*(6, Whole No. 599).

Guilford, J. P. *Psychometric methods* (2nd ed.). New York: McGraw-Hill, 1954.

Hall, C. A. *Differential relationships of pleasure and distress with depression and anxiety over a past, present, and future time framework*. Unpublished doctoral dissertation, University of Minnesota, 1977.

Izard, C. E. *Patterns of emotions*. New York: Academic Press, 1972.

McLachlan, J. F. C. A short adjective check list for the evaluation of anxiety and depression. *Journal of Clinical Psychology*, 1976, *32*, 195–197.

McNair, D. M., & Lorr, M. An analysis of mood in neurotics. *Journal of Abnormal and Social Psychology*, 1964, *69*, 620–627.

McNair, D. M., Lorr, M., & Droppleman, L. F. *Manual: Profile of Mood States*. San Diego, Calif.: Educational and Industrial Testing Service, 1971.

Meddis, R. Bipolar factors in mood adjective checklists. *British Journal of Social and Clinical Psychology*, 1972, *11*, 178–184.

Mehrabian, A., & Russell, J. A. *An approach to environmental psychology*. Cambridge, Mass.: MIT Press, 1974.

Nowlis, V. Research with the Mood Adjective Check List. In S. S. Tomkins & C. E. Izard (Eds.), *Affect, cognition, and personality*. New York: Springer, 1965.

Nowlis, V., & Nowlis, H. H. The description and analysis of mood. *Annals of the New York Academy of Sciences*, 1956, *65*, 345–355.

Olkin, I. Correlations revisited. In J. Stanley (Ed.), *Improving experimental design and statistical analysis*. Chicago: Rand McNally, 1967.

Osgood, C. E. On the whys and wherefores of E, P, and A. *Journal of Personality and Social Psychology*, 1969, *12*, 194–199.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. *The measurement of meaning*. Urbana: University of Illinois Press, 1957.

Russell, J. A. Evidence of convergent validity on the dimensions of affect. *Journal of Personality and Social Psychology*, 1978, *36*, 1152–1168.

Russell, J. A., & Mehrabian, A. Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, 1977, *11*, 273–294.

Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. *Manual for the State-Trait Anxiety Inventory*. Palo Alto, Calif.: Consulting Psychologists Press, 1970.

Stockford, L., & Bissell, H. W. Factors involved in establishing a merit-rating scale. *Personnel*, 1949, *26*, 94–118.

Thayer, R. E. Measurement of activation through self-report. *Psychological Reports*, 1967, *20*, 663–678.

Thayer, R. E. Activation states as assessed by verbal report and four psychophysiological variables. *Psychophysiology*, 1970, *7*, 86–94.

Westbrook, M. T. Positive affect: A method of content analysis for verbal samples. *Journal of Consulting and Clinical Psychology*, 1976, *44*, 715–719.

Zuckerman, M., & Lubin, B. *Manual for the Multiple Affect Adjective Check List*. San Diego, Calif.: Educational and Industrial Testing Service, 1965.

Received March 9, 1978 ▪

Journal of Personality and Social Psychology
1979, Vol. 37, No. 3, 357-368

The Generalizability of Salience Effects

Shelley E. Taylor, Jennifer Crocker, Susan T. Fiske,
Merle Sprinzen, and Joachim D. Winkler
Harvard University

Stimuli that engulf attention often have a disproportionately large impact on the judgment process, even when logically irrelevant. The boundary conditions of such salience effects were examined in a scenario where subjects observed a dyadic conversation in which the visual prominence of one or another participant was manipulated. Two hypotheses were examined: (a) Salience effects are dependent on quantity of information encoded and disappear at low levels of attention, and (b) salience effects will disappear if higher involvement in the situation serves to heighten and focus attention on more relevant cues. Neither hypothesis was supported. Rather, salience effects continue to be found (a) when the perceiver is distracted, (b) whether the perceiver's impressions are assessed immediately or after a delay, (c) when the conversation has high interest value, (d) regardless of the perceiver's cognitive tuning set, and (e) when the perceiver is involved in the discussion. We conclude that salience effects are highly generalizable and that they have a significant impact on both trivial and important social judgments. We suggest that salience effects are not customarily under the control of the social perceiver, but may rather be automatic.

Considerable research leads to the conclusion that logically uninformative, irrelevant, but salient social stimuli have a disproportionately large impact on the judgment process (see Taylor & Fiske, 1978, for a review). In general, the form of this "salience" effect is that people attribute causality to stimuli that engulf their attention. Taylor and Fiske (1975), for example, found in two separate studies that when one's attention is engulfed by a particular person in the environment, one sees that person as the causal agent in the conversation. Storms (1973) found that when an individual's own behavior or that of another person is made salient via videotaped playback, the salient information is a strong influence on judgments of causality. Duval and Wicklund (1972) found that the self is perceived as more of a causal agent when attention is self-focused than when it is not, and subsequent research has upheld this finding (see Wicklund, 1975; see also McArthur & Post, 1977).

Since the early studies that focused on judgments of causality, it has emerged that salient stimuli may have an impact on perceptions other than causal ones. Specifically, it appears that the evaluative qualities of salient stimuli are exaggerated, more is learned about them, they are perceived as more representative of their social group, and they are more "available," that is, more easily brought to mind (Tversky & Kahneman, 1974).[1] (See

The research described in this article was supported by National Institute of Mental Health Grant MH 26460 and National Science Foundation Grant BNS 77-009922 to the first author. Order of listing of the other authors is purely alphabetical. The authors wish to thank numerous confederates for their help; special thanks go to Nancy L. Etcoff, Joseph D'Agostino, and Judith Livingston for their help in collecting data. The third and fourth authors were supported by National Science Foundation fellowships during the period of this research, and the second and fifth authors were supported by the first author.

Requests for reprints should be sent to Shelley E. Taylor, Department of Psychology and Social Relations, Harvard University, Cambridge, Massachusetts 02138.

[1] Some studies (e.g., McArthur & Post, 1977; Storms, 1973) suggest that salient people's behavior is more likely to be perceived as dispositionally rather than situationally based. However, the evidence on this point is equivocal (see Taylor and Fiske, 1978).

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3703-0357$00.75

Taylor & Fiske, 1978, for a review of the literature on these points.) Taylor, Fiske, Close, Anderson, and Ruderman (Note 1), for example, found that a black (in an otherwise white group), male (in an otherwise female group), or female (in an otherwise male group) drew off a disproportionate amount of attention, was perceived as being more prominent and influential, was evaluated more extremely, and was somewhat more stereotyped (perceived as representative) than was a comparable individual in a racially or sexually mixed group. Several studies have found that more information is retained about a salient other (e.g., McArthur & Post, 1977; Taylor & Fiske, 1975). Pryor and Kriss (1977) found that salient items in sentences (subjects as opposed to objects) were both more available (i.e., recalled faster, though not better) and perceived as more causal.

Salience effects are of more than theoretical interest. They have been posited as the mediators of a number of important social processes. Taylor et al. (Note 1), for example, have suggested that the salience of a novel stimulus helps explain reactions to a solo black, a solo woman, or a solo man in a token integration situation. Wicklund (1975) has suggested that self-esteem is influenced by whether the self is salient. The salience of particular cognitions affects people's social judgments and self-perceptions (e.g., Kiesler, Nisbett, & Zanna, 1969; Ross, Lepper, & Hubbard, 1975). Salience explanations have been proferred for emotional disorders such as test anxiety (Wine, 1971) and stuttering (Storms & McCaul, 1976). However, the boundary conditions around salience effects have yet to be adequately explored.

One hypothesis (Taylor & Fiske, 1975, 1978) is that salience effects depend on some minimal level of attention during exposure to the initial stimulus conditions. According to this hypothesis (Hypothesis 1), salience effects would not be expected to emerge if attention (as indexed by recall) were attenuated.

A second hypothesis (Hypothesis 2), which is in some respects opposite to the first, is that salience effects are characteristic of low levels of intensity of attention and that if a variable were introduced that increased the individual's intensity of attention, salience effects would not emerge. The assumption underlying such a hypothesis would be that people pay attention to trivial but salient stimuli when there is nothing else worth watching more closely, but that when more interesting or informative stimuli are present degree of attention is heightened and focused such that otherwise salient but trivial stimuli are ignored. Such variables might include a high level of interest value in the content of material exchanged or the subjects' involvement in the information exchanged. Since much of the research demonstrating salience effects has been conducted in unarousing, uninvolving, and low information contexts, the question of whether salience effects will generalize to more engrossing, arousing, and involving settings is a very important one (see McArthur & Solomon, 1978).

Hypotheses 1 and 2 are not necessarily incompatible. It may be that some minimum level of attention is required to obtain salience effects, but that at the highly intense levels of attention produced by high degrees of subject involvement, salience effects will disappear.

Three studies were conducted to examine the hypothesized bases of salience effects. In the first, subjects participated in a point-of-view experiment in which their information intake was either severely reduced or not by a distraction task. In the second and third studies, a variety of manipulations was introduced to see if increasing the subject's degree of attention through involvement in the nature of the material exchanged would alter salience effects. Failure to obtain salience effects under conditions of high distraction would constitute evidence for Hypothesis 1. Failure to obtain salience effects under conditions of increased attention would constitute evidence for Hypothesis 2.

Experiment 1

Experiment 1 examined whether salience effects depend on differential encoding of information during exposure to the initial stimulus conditions. In this study, subjects were distracted during either the encoding stage (i.e., when observing the conversation in

which salient information was provided) or the retrieval stage (i.e., when rating the two participants). Crossing these two variables enabled us to see if salience effects represent an encoding or a retrieval bias or a combination of the two.

Method

Subjects

Subjects were 100 Harvard–Radcliffe undergraduates who signed up for a "getting–acquainted" study. They were run in groups of 8 to 9 subjects, with two experimental confederates posing as additional subjects.

When subjects arrived for the experimental session, the experimenter explained that an informal conversation would involve two of them and that the rest of them would be observers. The experimenter "picked" the two (male) confederates to hold the conversation and seated them facing each other, with the other subjects arranged as observers. Four subjects sat behind each confederate, and if a ninth subject was present, that person sat in the control condition. That is, for four subjects, one of the confederates (A) engulfed their visual attention, and for the other four subjects, the other confederate (B) engulfed their visual attention. Control subjects had equal visual access to both confederates. This procedure has been successfully used as a manipulation of salience in previous studies (see, for example, Taylor & Fiske, 1975).

Procedure

The observers were instructed that the experiment was concerned with several specific aspects of the conversation, so that separate instruction sheets would be passed out to each person. Subjects were told that their task might seem odd, but that both psychological and linguistic aspects of interaction were being studied in the experiment. All subjects were told to listen to the conversation without participating in it and without taking notes. For the undistracted condition, the instructions read:

> As you are watching and listening, please observe the conversation in general. Form an impression of the whole discussion. Imagine you are overhearing these people in a public place. Afterwards, we will be asking you about your observations of these people and the entire conversation.

The other half of the subjects received instructions that constituted the encoding distraction manipulation:

> As you are watching and listening, please observe the conversational usage of pronouns. Count the total times I, me, and you are used. Keep a combined score for the conversation as a whole. Afterwards, we will be asking you about your observations of these people and the entire conversation.

The confederates then carried on a "getting–acquainted conversation," as if they had just met. They exchanged information about summer jobs, college major, career plans, hometown, and the like. The experimenter stopped the conversation after they had exchanged the standard information. The confederates were asked to leave and wait down the hall.

At this point, the decoding distraction manipulation was introduced. Half the subjects were seated at chairs with writing surfaces, and they were told to fill out the impression questionnaire. The other half of the subjects were taken into the next room and seated at a large table. The experimenter mentioned that someone else was working in the room, but had said that the subjects could use it. When the subjects entered the room, they found a second experimenter setting up audiovisual materials ostensibly for an unrelated experiment. As these subjects attempted to fill out the dependent measures, the second experimenter continually readjusted the volume of a tape recording, changed and focused colorful slides, and generally bombarded the subjects with distracting audiovisual stimuli.

Dependent Measures

Subjects rated both confederates, and the order of rating (salient first/nonsalient first) was counterbalanced. For each confederate, subjects were asked first to recall what he had said. They then rated how friendly, talkative, and nervous he was, to what extent each of these qualities was dispositionally caused, and to what extent it was situationally caused. The last five questions assessed causal influence in the conversation: To what extent did this person set the tone of the conversation between the two people? To what extent did this person determine what kind of information would be exchanged during this conversation? To what extent did this person cause his partner to behave as he did? To what extent did you form an impression of this person? And, how much did this person influence the conversation as a whole? All responses were on 9–point scales with labeled endpoints.

After subjects had completed recall, dispositional–situational, and conversational causality measures for each confederate, they were asked to report the results of their observation task (if any), for example, the number of pronouns they had counted. They were also asked how successful they felt at the observation task, how much the situation interfered with the observation instructions, and how much the situation interfered with filling out the questionnaire. When they had finished, subjects were paid $2 each and fully debriefed. None expressed suspicion about the study's purpose.

Results

The dependent measures were analyzed by a four–way analysis of variance. The between–subjects variables were encoding distraction (distracted/not), retrieval distraction (distracted/not), and confederate faced (A/B). Confederate rating (A/B) was a within–subjects variable for all measures except the manipulation checks.

Manipulation Checks

On the question assessing the encoding distraction manipulation, subjects who counted pronouns felt that the observation situation (i.e., counting pronouns) interfered significantly more than did undistracted subjects, $F(1, 91) = 4.84$, $p < .03$; Ms = 4.80 vs. 3.72. There were no other significant effects in that analysis.

On the question assessing ease of filling out the questionnaire (the retrieval distraction manipulation check), retrieval distraction subjects reported no more difficulty than no-retrieval distraction subjects; however, retrieval distraction subjects took much longer to fill out the questionnaires than did the no-retrieval distraction subjects. This informal observation indicated that despite the lack of difference on the manipulation check, the manipulation of retrieval distraction may have been successful. (On the other hand, subjects may have circumvented the retrieval distraction manipulation by taking longer.) Interestingly, it was the encoding distraction subjects who showed a difference on this distraction measure, $F(1, 90) = 6.64$, $p < .02$. Apparently, the encoding distraction subjects' difficulty in observing the conversation made them find it hard to answer the questions. This was particularly true for the distracted subjects facing Confederate B, $F(1, 90) = 4.80$, $p < .03$. There were no other significant effects. Finally, the question assessing how successful subjects felt at the observation task, a potential covariate for later analyses, showed no effects at all.

Causality Measures

The five conversational causality measures (set tone, determine information exchanged, cause partner's behavior, influence conversation, form impression) were summed to form a general index of causality, and the analysis showed that the salience effect emerged clearly, $F(1, 91) = 13.29$, $p < .001$. Means are presented in Table 1. Importantly, this effect did not interact with either distraction manipulation, either separately or combined (all Fs < 1). Thus, the data show a simple salience effect, uninfluenced by both distraction manipulations.

Table 1
Causality and Evaluation Ratings: Experiment 1

Condition	Rating			
	Causality index	Friendly	Talkative	Nervous
Facing A				
Rate A	22.58	6.56	6.40	4.73
Rate B	21.80	6.30	4.91	4.90
Facing B				
Rate A	18.85	5.72	5.42	5.10
Rate B	26.11	7.49	5.79	4.06

Note. Causality index values range from 5 to 45; other scales were 9-point rating scales. Higher values indicate greater rated perception. F values for salience interaction effect range from 5.25 to 22.53, with 1 and 75 degrees of freedom; ps range from <.03 to <.001.

In addition to the salience effect, the causality ratings showed a main effect for confederate, $F(1, 91) = 8.65$, $p < .005$, such that one confederate (B) was consistently seen as more causal than the other. This confederate effect accounts for the reversal of the predicted mean pattern in the no distraction–facing A cell, but does not affect the salience results. The question How much did this person talk?—also a measure of prominence—showed salience effects such that the salient actor's estimated talking time was rated higher, $F(1, 92) = 16.90$, $p < .001$. There was also a confederate effect on this measure, $F(1, 92) = 6.16$, $p < .02$.

The dispositional and situational causality measures were summed to form separate scales of each. No salience effects emerged, nor were there any other significant effects.

Recall

There was no effect of salience on recall (F close to zero). However, the encoding distraction (counting pronouns) manipulation resulted in those subjects remembering virtually nothing of the conversation, $F(1, 92) = 20.45$, $p < .001$, compared with subjects who simply observed without any distraction. Recall also showed a main effect for confederate, $F(1, 92) = 45.65$, $p < .001$.

Evaluations

There were two evaluative measures: friendly and nervous. Salience effects emerged on the friendliness rating, $F(1, 92) = 41.76$, $p < .001$, such that the salient actor was seen as especially friendly. Similarly, salience effects emerged on the nervousness rating, $F(1, 91) = 6.24$, $p < .02$, such that the confederate a subject faced was seen as relatively less nervous than the nonsalient confederate. On both personality evaluations, the salient actor's rating differed substantially from the nonsalient actor's rating. Since both means were on the same side of the midpoint, and the salient actor's ratings were always more extreme, this pattern represents an exaggerated evaluation of personality attributes due to salience, a pattern similar to that found in previous studies (Taylor & Fiske, 1978).

Both variables also showed confederate effects: friendly, $F(1, 92) = 23.09$, $p < .001$; and nervous, $F(1, 91) = 3.18$, $p < .07$. Finally, the encoding distraction influenced subjects to rate both confederates as less friendly, $F(1, 92) = 9.51$, $p < .003$.

To summarize, then, salience effects on causality ratings and evaluations were found in this study. The hypothesis that such effects depend on differential encoding of information during exposure to initial stimulus material was not supported, however.

Experiment 2

Experiment 2 was designed as a test of Hypothesis 2: that salience effects are eliminated when degree of attention is heightened and focused by a more engaging or informative setting. In order to examine conditions with high external generalizability, we manipulated intensity of attention by having subjects observe a debate on a current issue with which they were either involved or uninvolved. In addition to examining the usual dependent variable of causality (the main test of the hypothesis), we included secondary measures of persuasiveness, attitudes toward the issue, and attitudes toward the debate to see if they would be affected by salience. Specifically, if salience exaggerates affect, one's beliefs about a debater's persuasiveness, efficacy, and the like should increase when one agrees with the debater's position and he or she is salient, and they should decrease when one disagrees with the debater and he or she is salient.

Method

Subjects

Ninety-one Harvard–Radcliffe undergraduates (46 males, 45 females) participated as paid volunteers and were selected on the basis of their position on the issue debated and the importance of that issue to them. Involvement and position on the issues was assessed by drawing subjects from a pool of participants in an "attitude survey" ostensibly unconnected with the present study and run by a different experimenter. The attitude survey asked participants to indicate their position on each of 34 different issues and to indicate the importance of each issue to them. Two issues were used in the present study: banning cigarettes and alimony for unmarried couples. Subjects selected for the involved conditions rated the issue significantly more important to them ($M = 5.67$ on a 9-point scale) than uninvolved subjects ($M = 2.50$, $t = 13.25$, $p < .001$). There was a small but nonsignificant difference between the involved and uninvolved subjects on their position on the issue, with involved subjects slightly more opposed ($M = 1.38$) than uninvolved subjects ($M = 1.72$, $t = 1.40$, *ns*).

All the survey participants were informed that several people in the psychology department needed subjects for experiments and that they should leave their name and number if they were interested in participating as paid volunteers in studies unconnected to the survey.

Procedure

Survey participants who were opposed on one of the two issues were contacted by telephone and recruited for an experiment on persuasion. The same procedure as in Experiment 1 was used, with the following changes. One involved subject and one uninvolved subject were assigned to each of the three viewpoint conditions. To support the impression that the confederates were simply subjects

assigned to the debater role, the confederates were given 2 minutes to think about and jot down arguments supporting the side of the issue they had been assigned to argue. Two different pairs of confederates assisted with the study, and each confederate argued the "pro" side of the issue in some groups and the "anti" side in other groups.

Dependent Measures

Subjects rated both confederates, and the order of rating (salient first/nonsalient first) was counterbalanced. Subjects were asked to recall all the personal information and arguments that each of the debaters gave.

Subjects then completed 9-point measures assessing which confederate was more causal (To what extent did this person set the tone of the discussion? How much did this person talk in the discussion?); which confederate was more persuasive (How much did this person persuade the other?); replicability of the outcome (If a rematch of this discussion were arranged, how different would the outcome be?); and mutability of the outcome (How likely is it that this person could have said anything different to be more persuasive?). Subjects also judged which confederate had won the argument, which confederate was more convincing, and which one was more thorough.

Finally, the subjects were asked to indicate their position on the issue, and they filled out several manipulation checks (How interesting did you think the discussion was? How much did you care who won the argument? How much do you care about this issue? How personally involved did you feel in the discussion?).

When all the subjects in the group had completed the dependent measures, they were carefully debriefed, asked for their impressions of the study, informed of its purpose, and paid $1.50 each. During the debriefing, two subjects indicated a suspicion that the debaters were actually confederates of the experimenter. When questioned, they indicated no understanding of the hypothesis of the study, so the data from these subjects were included in the analysis. Subjects were asked not to discuss the study with other potential subjects.

Results

Manipulation Checks

Four manipulation checks of involvement were included, and analysis indicated that the involvement manipulation was successful. In response to the question, How much do you care about this issue? involved subjects indicated that they cared significantly more ($M = 4.67$) than uninvolved subjects ($M = 3.01$), $F(1, 78) = 22.94$, $p < .001$. Involved subjects also reported feeling significantly more personally involved in the discussion ($M = 3.53$) than uninvolved subjects ($M = 2.34$), $F(1, 78) = 7.46$, $p < .008$. Differences between the involved and uninvolved subjects were not significant on their ratings of how interesting they found the discussion (involved $M = 3.74$; uninvolved $M = 3.46$), $F(1, 78) = .50$, $p < .49$, or how much they cared who won (involved $M = 2.58$; uninvolved $M = 2.12$), $F(1, 78) = 1.33$, $p < .26$.

Causality Measures

Ratings on two causality measures used in earlier studies (To what extent did this person set the tone of the discussion? and How much did this person talk in the discussion?) were correlated .54 ($p < .001$) and did not correlate higher than .25 with any other dependent variables. Accordingly, a causality measure was created for each confederate by summing the talkativeness and the setting-the-tone ratings. A repeated measures analysis of variance with viewpoint (facing pro, facing con, or control), involvement (2 levels), and issue (2 levels) as the between-subjects factors and person rated (pro or con) as the within-subjects factor was performed on the causality index. The point-of-view effect (Viewpoint × Person Rated) was weak and nonsignificant, $F(2, 79) = 1.022$, $p < .40$. However, the interaction of the point-of-view effect with involvement (i.e., Viewpoint × Person Rated × Involvement) was significant, $F(2, 79) = 4.95$, $p < .01$, and revealed that the point-of-view effect was strong for highly involved subjects but irregular for low-involvement subjects, as the means in Table 2 indicate. There was also a significant Viewpoint × Person Rated × Involvement × Issue interaction on the causality item, which showed that the failure to replicate the point-of-view effect for uninvolved subjects was mainly due to a reversal of the predicted mean pattern for uninvolved subjects in the smoking issue condition only, $F(2, 79) = 3.65$, $p < .04$. Finally, there was a significant main effect for person rated such that subjects rated the person they agreed with (con) as more causal ($M = 7.45$) than the person they disagreed with ($M = 6.51$), $F(1, 79) = 4.90$, $p < .03$.

Table 2
Mean Causality Ratings as a Function of Involvement: Experiment 2

	Involvement			
	High		Low	
	Person rated		Person rated	
Viewpoint	A	B	A	B
Facing A	7.32	6.53	5.88	7.80
Control	7.20	7.40	5.86	6.86
Facing B	5.13	8.69	7.37	7.39

Note. Subjects agreed with Confederate B and disagreed with Confederate A; higher numbers indicate greater causality. Viewpoint × Person Rated × Involvement interaction was significant at $F(2, 79) = 4.95$, $p < .01$.

Recall

Subjects were asked to recall as many of the arguments and personal details given by each confederate as they could. There were no significant salience effects on either recall measure. There was a significant Person Rated × Involvement interaction on recall of arguments such that highly involved subjects recalled more of the arguments they agreed with (the con arguments) ($M = 44.52\%$; $M = 41.91\%$ for pro arguments), whereas uninvolved subjects recalled more of the arguments they disagreed with ($M = 49.87\%$; for pro arguments, $M = 43.50\%$), $F(1,79) = 4.79$, $p < .04$. There were also uninterpretable main effects for issue on several measures that will not be reported here.

Evaluations

No significant salience effects (Viewpoint × Person Rated) or Salience × Involvement interactions emerged on the secondary judgments (persuasiveness, repeatability, mutability, convincingness, thoroughness, winning the argument, and attitude change).

Discussion

Why did salience effects replicate under conditions of high involvement but not under conditions of low involvement? If other point-of-view studies are uninvolving, as we have suggested, this failure to replicate the effect under the low-involvement condition is quite puzzling. One possibility is that the low-involvement condition is less involving than the typical point-of-view study, whereas the high-involvement condition is more involving. Although it is impossible to directly compare the degree of subject involvement across studies, this explanation seems possible. Subjects chosen because they indicated that an issue was not important to them may well be less involved than subjects viewing an exchange of personal information, who may in turn be less involved than subjects viewing a discussion on a topic that they rated as very important to them. Anecdotal support for this explanation is provided by the experimenter's report that involved subjects paid close attention to the arguments and suggested during the debriefing points that the debaters had missed; however, uninvolved subjects spent time looking around the room, playing with their pencils, and generally looking very bored. This explanation, then, suggests that some minimal degree of attention is required to achieve salience effects; that is, the subject must be paying enough attention for the salient information to actually be salient.

While this explanation would seem to contradict the results of Experiment 1, in fact, it does not. The salience manipulation is a visual one (i.e., who engulfs one's visual field). In Experiment 1, though subjects' verbal recall was eliminated, their attention to the salience manipulation did not wane, whereas in Experiment 2, it was the visual attention of low-involvement subjects that seemed to wander. Further research will determine whether this explanation is viable or whether, since the nonreplication occurred in only one of the low-involvement conditions, it is simply a fluke of the stimulus materials.

Experiment 3

Experiment 3 was designed to conceptually replicate and extend the results of Experiment 2 and, by implication, the test of Hypothesis 2. We pursued the issue of whether salience effects (overrated causal role of sali-

ent but trivial stimuli) are eliminated when other factors conspire to heighten and focus attention on more relevant aspects of the situation. In this study, half the subjects were given a cognitive tuning set (Zajonc, 1960) to receive information, and half were given a transmission set to communicate information to others. We reasoned that a transmission set prompts subjects to heighten and focus their attention to information necessary for transmitting a cogent impression (e.g., an individual's behavior) as opposed to salient but trivial information (i.e., his or her visual prominence), whereas a receiving set more closely mirrors the set of subjects in previous studies.

A second issue we explored in Experiment 3 was the effect of timing on salience effects. It may be that when information has been highly salient, its impact on the judgment process is strong only immediately after the experience, when the salient information and the judgment process are in close contiguity, and not if judgments are delayed to a later time. That is, it is possible that following a delay, salient information assumes its rightful, limited role in judgments.

Method

Subjects

Eighty-eight Harvard-Radcliffe students (56 males, 32 females) were scheduled in groups of 6. Subjects were told that the study's purpose was to find out what happens during self-disclosure conversations (that is, how observers form impressions of people who are disclosing characteristics about themselves) and what observers do with information about others at different points in time.

We used a self-disclosure conversation because this kind of conversation is less ritualistic, stylized, and formal than getting-acquainted conversations. In addition, the confederates would be conversing in impression-formation terms; that is, the confederates would be describing themselves in ways that would tend naturally to allow subjects to form impressions. For these reasons, this kind of conversation is more interesting to the subjects than the getting-acquainted conversations used in previous research.

Cognitive Tuning Manipulation

As observers, the subjects were asked to watch a videotape of a self-disclosure conversation while performing a particular task. Sheets of paper were passed out describing the task each was to perform. Half of the subjects in each group were given the "receiver" manipulation:

> Your task as you watch the videotape is to begin to form an initial impression of both of the participants. You should use the information each person provides about himself to begin to think about the kind of people they are. After you have finished watching the videotape, during the second part of the study, you will be receiving additional information about the participants that should help you form better impressions of them both. In other words, your task while watching the videotape is to begin to think about the sorts of people the participants are, impressions that should be clearer after you have received additional information about them during the second part of the study.

The others received the "transmitter" manipulation:

> Your task as you watch the videotape is to form a clear impression of both of the participants. You should use the information each person provides about himself to figure out what kind of people they really are. After you have finished watching the videotape, during the second part of the study, you will be asked to communicate all you can about both people to others. In other words, your task is to determine, from watching their conversation, the sorts of people the participants are in order to communicate your impressions to others during the second part of the study.

When the videotape ended, subjects were asked to write down what their task was, as a manipulation check.

Point-of-View Manipulation

The experimenter told the subjects that the conversation they would be watching had been videotaped during another part of the experiment that had involved subjects having a self-disclosure conversation with other subjects. Subjects were told that cameras had been set up to give a split-screen image, but in the videotaping of the particular conversation they would be watching, one of the cameras had not worked properly; instead of a split-screen picture, they would see one in which only one of the two conversationalists was clearly visible. The subjects then watched a videotape of two male confederates, A and B, ostensibly having a self-disclosure conversation. The videotape for all groups began with a brief introduction in which both confederates could be seen, and then, following a blackout in the film, half of the groups of subjects could only see A over the shoulder of B, whereas the other half of the groups could only see B over the shoulder of A.

The confederates discussed characteristics that they liked and did not like about themselves and characteristics that other people liked and did not like about them. The confederates talked about the same length of time and exchanged similar kinds of

moderately personal information. The conversation lasted about 15 minutes.

Delay Manipulation

After watching the videotape, subjects in the delayed conditions were dismissed; two subjects per group (one receiver and one transmitter) returned after an hour, and two subjects (one receiver and one transmitter) returned after a day. The remaining two subjects per group (one receiver and one transmitter) completed the dependent measures immediately.

The tuning, point-of-view, and delay manipulations produced a fully crossed $2 \times 2 \times 3$ between–subjects design.

Dependent Measures

The experimenter asked the subjects to fill out a questionnaire about the self-disclosure conversation they had observed. Each subject rated both confederates; the order of rating the confederates was counterbalanced. The first part of the questionnaire, for each confederate, was a recall task; subjects were asked to remember what the confederate said that he and others liked and did not like about himself. In the second part, subjects rated how friendly, talkative, and nervous the confederate was, on 9-point scales with labeled endpoints. For each of the three behaviors, the subject also rated the extent to which the behaviors were caused by dispositional qualities of the person (9-point scale) and by situational factors (9-point scale). In the third part, on 9-point scales, subjects rated how much the confederate set the tone of the conversation, was able to draw out more from the other than he had to give up about himself, caused his partner to behave as he did, truly disclosed about himself, and influenced the conversation as a whole. Subjects also rated the degree to which they were able to form an opinion of the confederate, received a favorable impression of the confederate, and liked him. Last, they were asked to rate how easy it was for them to answer the questions in this part of the questionnaire. Subjects then filled out an identical set of questions about the other confederate.

Subjects took approximately 20 minutes to complete the ratings. Subjects were then fully debriefed. The debriefing revealed that subjects were not suspicious about the point–of–view manipulation and did not guess the true nature of the experiment. Subjects took the cognitive tuning manipulation seriously; transmitters used the debriefing to relate their impressions of the confederates, whereas receivers expected to receive additional information. The hypotheses of the experiment were then discussed, and the subjects were paid; subjects in the immediate condition received $2 each, and subjects in the delayed conditions received $3 each.

Table 3
Mean Causality and Evaluation Ratings as a Function of Salience: Experiment 3

Measure	Rating of confederate faced	Rating of confederate not faced
Causality index	14.32	11.76
Talking	5.98	5.32
Recall of positive characteristics	2.95	2.41
Favorable impression	4.50	3.35
Ease of ratings	6.36	5.88
Liking	4.09	3.18

Note. Causality index values range from 3 to 27; other scales are 9-point rating scales. Higher values indicate greater rated perception. F values for salience interaction effect range from 5.25 to 22.53, with 1 and 75 degrees of freedom; ps range from $<.03$ to $<.001$.

Results

Dependent measures were analyzed by analyses of variance using cognitive tuning, confederate faced, and delay as between–subjects factors. Because each subject rated both confederates, confederate ratings were a repeated measure.

Manipulation Checks

The manipulation check verified that the cognitive tuning manipulation had worked for most of the subjects (73 of the 88 subjects remembered their task after they had viewed the videotape).

Causality Effects

On the basis of high average intercorrelations, subjects' ratings of confederates' influence on the tone of the conversation, influence on the conversation as a whole, and influence on partner were combined into a causality index for purposes of analysis.

As Table 3 demonstrates, the salience effect emerged. Subjects perceived that the confederate they observed was more causal in the conversation than the confederate they did not observe, $F(1, 76) = 7.45$, $p < .01$. There were no other interactions or main effects on this causality index. Even when sub-

jects who had not remembered the tuning manipulation were removed, there were no significant cognitive tuning main effects or interactions with the causality index. The question How much did this person talk? is also a measure of prominence and showed similar salience effects, such that subjects perceived the salient confederate as talking more than the nonsalient confederate, $F(1, 75) = 5.25, p < .03$.

Given that the observer has decided who was influencing the conversation, will the observer go on to make situational and personal attributions about why the confederate behaved as he did? As mentioned, for the three behaviors measured—friendliness, talkativeness, and nervousness—subjects indicated how dispositionally caused and how situationally caused each was for each confederate. The analysis revealed that subjects perceived that the nervousness of the confederate they observed was caused less by the personal characteristics of the confederate and more by the characteristics of the situation, whereas they perceived that the nervousness of the confederate they did not observe was caused more by the personal characteristics of the confederate and less by the characteristics of the situation (personal, $F(1, 72) = 8.73, p < .01$; situational, $F(1, 72) = 4.76, p < .03$. This finding constitutes a reversal of the mean pattern predicted by the actor–observer hypothesis (see Storms, 1973). However, because of the nature of the attribute "nervous," it may be that the data pattern is consistent with a view of the salient person as "in charge." [2]

Evaluations

Significant salience effects also emerged on recall of positive characteristics mentioned by the confederate, $F(1, 76) = 22.53, p < .001$; the degree to which the confederate made a favorable impression, $F(1, 76) = 9.30, p < .01$; the ease with which the subject could rate the confederate, $F(1, 76) = 15.39, p < .001$; and how much the subject liked the confederate, $F(1, 76) = 8.18, p < .01$. The means of these measures (see Table 3) revealed that subjects could more easily form impressions, particularly favorable impressions, about the confederate they faced than about the confederate they did not. As in Experiment 1, the salient person's attributes were exaggerated in a positive direction. Confederate effects also emerged on these impression measures, but they are uninterpretable and will not be reported here.

Summary

Experiment 3 found strong salience effects on a wide variety of measures in a self-disclosure situation. However, the impact of cognitive set and whether impressions were assessed immediately or later on the magnitude of those salience effects was negligible.

General Discussion

Three studies examined two hypotheses regarding the effects of salient information on social judgments. The first was that salience effects are dependent on attention and that differential attributions are made to stimuli for which large amounts of information have been gathered. Results indicated that as long as subjects actually attend to the salient information, salience effects occur and are not increased or decreased by more or less retention of information. Indeed, the fact that salience effects occurred even when subjects had taken in virtually none of the verbal information exchanged in the scenario indicates that a very low degree of verbal attention can still produce biased attributions of causality to a visually salient person.

The second hypothesis tested was that salience effects are characteristic of low levels of intensity of attention and that if some varia-

---

[2] There were some uninterpretable interaction effects with some of the measures: (a) a Delay × Cognitive Tuning interaction with talking, because of personal characteristics, $F(2, 75) = 6.32, p < .01$; (b) a Delay × Cognitive Tuning × Person Faced interaction with nervous, because of situational characteristics, $F(2, 72) = 3.14, p < .05$; (c) a Delay × Cognitive Tuning × Salience interaction with how easily the subjects could rate the confederate, $F(1, 75) = 4.62, p < .01$; and (d) a Cognitive Tuning × Person Faced interaction on perceptions of how much the confederate talked, $F(1, 75) = 7.94, p < .01$.

ble that increases intensity of attention (such as subject involvement in the scenario) were introduced, salience effects would not emerge. The rationale for this prediction was that heightened involvement in the scenario would increase attention to relevant cues, thereby reducing the impact of otherwise salient but trivial stimuli. Two studies tested this hypothesis and found that under conditions of involvement with the issue, more engrossing conversations, and a cognitive tuning set to transmit rather than receive, salience effects continued to emerge.

The results of these studies suggest that salience effects have at least some degree of external generalizability. In addition to the conditions of generalizability tested in the present studies (involvement, tuning, distraction, and delay), the results of two pilot studies also extend the domain of applicability of salience effects. In a pilot study by Gartrell (Note 2), salience effects were obtained even when one actor clearly dominated the other in a conversation. In a second pilot study (Taylor & Winkler, Note 3) assessing the impact of arousal on salience effects, power analyses indicated that salience effects would emerge under conditions of arousal but that their magnitude would not be substantially altered by level of arousal. Taking all this information together, we conclude that salience effects do generalize to more engrossing and involving situations than those examined in earlier studies and that salience effects are not dependent on either subject set (transmitting versus receiving) or on immediate assessment of subjects' impressions. The fact that salience effects continue to emerge in interesting, involving, and arousing conditions also makes more plausible the argument that salience may mediate consequential social perceptions.

The results of the present study are also highly suggestive regarding how salience effects themselves are mediated. Elimination of verbal recall leaves the effect untouched (Experiment 1), whereas wandering visual attention may be responsible for the failure to replicate in Experiment 2. The obvious conclusion to draw is that salience effects may well be visually mediated. However, results of an effort to model the mediation process (Fiske, Kenny, & Taylor, Note 4) do not unequivocally support this contention, though the indicators of both verbal and visual recall in those two studies are by no means infallible. In any case, given inconclusive data, we prefer to hedge on the issue of mediation for the present.

The present studies indicate that salience effects are not pushed around very much by the social variables that would be expected to interact with them. Nor are salience effects a function of the subject's set or of the juxtaposition of the salience scenario and the rating task. Indeed, what is striking about the point–of–view effect is that it is quite reliable and practically unmovable. What is suggested to us by this fact is that although the effect is manifested on variables of social interest such as causality or prominence ratings, the effect itself may be an automatic perceptual bias (not unlike optical illusions, orienting to novel stimuli, and the principles of figural emphasis uncovered by Gestalt psychologists). Thus, we would argue that salience effects are automatic responses to stimulus qualities, which the perceiver has never learned and which occur without intention (cf. Schneider & Shiffrin, 1977).

Reference Notes

1. Taylor, S. E., Fiske, S. T., Close, M., Anderson, C., & Ruderman, A. *Solo status as a psychological variable: The power of being distinctive.* Manuscript submitted for review, 1977.
2. Gartrell, D. *Dominance–submissiveness of actor, point of view of observer, and the attribution of causality in interactions.* Paper presented to the annual meeting of the Canadian Sociology and Anthropology Association, Edmonton, Canada, May 1975.
3. Taylor, S. T., & Winkler, J. *Caffeine arousal and the generalizability of the point-of-view effect.* Unpublished manuscript, Harvard University, 1979.
4. Fiske, S. T., Kenny, D. A., & Taylor, S. E. *Path analysis of attentional mediation in attribution.* Manuscript in preparation, 1977.

References

Duval, S., & Wicklund, R. A. *A theory of objective self-awareness.* New York: Academic Press, 1972.
Kiesler, C. A., Nisbett, R. E., & Zanna, M. P. On inferring one's beliefs from one's behavior. *Jour-*

nal of Personality and Social Psychology, 1969, *22,* 321–327.

McArthur, L., & Post, D. Figural emphasis and person perception. *Journal of Experimental Social Psychology,* 1977, *13,* 520–535.

McArthur, L. Z., & Solomon, L. K. Perceptions of an aggressive encounter as a function of the victim's salience and the perceiver's arousal. *Journal of Personality and Social Psychology,* 1978, *36,* 1278–1290.

Pryor, J. B., & Kriss, M. The cognitive dynamics of salience in the attribution process. *Journal of Personality and Social Psychology,* 1977, *35,* 49–55.

Ross, L., Lepper, M., & Hubbard, M. Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm. *Journal of Personality and Social Psychology,* 1975, *32,* 380–392.

Schneider, W., & Shiffrin, R. M. Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review,* 1977, *84,* 1–66.

Storms, M. D. Videotape and the attribution process: Reversing actors' and observers' point of view. *Journal of Personality and Social Psychology,* 1973, *27,* 165–167.

Storms, M. D., & McCaul, K. D. Attribution processes and emotional exacerbation of dysfunctional behavior. In J. H. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research* (Vol. 1). Hillsdale, N.J.: Erlbaum, 1976.

Taylor, S. E., & Fiske, S. T. Point of view and perceptions of causality. *Journal of Personality and Social Psychology,* 1975, *32,* 439–445.

Taylor, S. E., & Fiske, S. T. Salience, attention, and attribution: Top of the head phenomena. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 11). New York: Academic Press, 1978.

Tversky, A., & Kahneman, D. Judgment under uncertainty: Heuristics and biases. *Science,* 1974, *185,* 1124–1131.

Wicklund, R. A. Objective self-awareness. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 8). New York: Academic Press, 1975.

Wine, J. Test anxiety and direction of attention. *Psychological Bulletin,* 1971, *76,* 92–104.

Zajonc, R. B. The process of cognitive tuning in communication. *Journal of Abnormal and Social Psychology,* 1960, *61,* 159–167.

Received March 10, 1978 ▪

Journal of Personality and Social Psychology
1979, Vol. 37, No. 3, 369–379

Interpersonal Consequences of Person Perception Processes in Two Social Contexts

Lawrence A. Messé and Gary E. Stollak
Michigan State University

Rita W. Larson
Menlo Park, California

Gerald Y. Michaels
University of Michigan (Ann Arbor)

This research examined the proposition that perceiver-based perceptual processes affect social behavior. Approximately 1,100 undergraduates were exposed to a videotape that portrayed a male or female child interacting with an adult in a playroom. In Study 1, subjects who "saw" the child emit (a) primarily positive behaviors (i.e., subjects who were positively biased), (b) about equal numbers of positive and negative behaviors (i.e., subjects who were accurate), or (c) primarily negative behaviors (i.e., subjects who were negatively biased) then engaged in cooperative task activities with a 7-year-old child. In Study 2, a subset of these subjects engaged in a discussion with another undergraduate about three issues on which they apparently disagreed. Systematic analyses of these interactions suggested that perceptual processes affected social behavior; for example, negatively biased subjects tended to act in a more authoritarian manner in their encounters with the child, whereas positively biased subjects were the least effective in the discussion task. These and the other findings are discussed in terms of their implications for understanding the relationship between person perception and overt social behavior.

There is a set of theories in psychology (e.g., Combs & Snygg, 1959; Kelly, 1955; Rogers, 1959) and sociology (e.g., Cottrell, 1966; Mead, 1934) that explains human behavior in part in terms of person perception processes. For example, Kelly (1969) postulates that how a person "anticipates" events —including interpersonal encounters— through cognitive and perceptual processes affects the course of those events. Similarly, perhaps even more strongly, Combs and Snygg (1959) argue that all human behavior is determined by perceptual mechanisms and processes. This position is also similar to the proposition of role theory (Cottrell, 1966) and symbolic interactionism (Mead, 1934) that people's "definitions of the situation" have a large influence on the course of their interpersonal encounters.

Given that person perception processes are seen by such theorists as important determinants of social behavior, it is somewhat surprising that few studies have attempted to establish empirically the link between perception and behavior (cf. Hastorf, Schneider, & Polefka, 1970; Shrauger & Altrocchi, 1964; Taguiri, 1969). Notable exceptions to this

This article is a revision of two papers presented as part of a symposium held at the 1977 annual meeting of the American Psychological Association. The research reported here as Study 2 was part of the third author's doctoral dissertation, conducted under the supervision of the first two authors, which she submitted to the Department of Psychology, Michigan State University. This research was supported by National Institute of Mental Health Grant MH24250, "Adult Perceptions and Child Behavior Dysfunctions." We wish to thank Linda Giacomo for her help and support, especially during the early stages of this research, and Roger Buldain for assisting with the statistical analyses. We also wish to thank the numerous undergraduates who provided invaluable assistance in collecting and scoring the data. The cooperation of the staff and administration of the East Lansing, Michigan, school district, especially Robert Docking, Malcolm Katz, and Dallas Wegener, is also gratefully acknowledged. Finally, we wish to express our gratitude to the 170 or so children who served as the actors and participants for this research.

Requests for reprints should be sent to Lawrence A. Messé, Department of Psychology, Michigan State University, East Lansing, Michigan 48824.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3703-0369$00.75

statement are the seminal work of Kelley (1950) on the consequences of a set to see another as warm or cold, the research on people's reactions to attractive or "stigmatized" others (e.g., Kleck, Ono, & Hastorf, 1966; Snyder, Tanke, & Berscheid, 1977), and—although not usually thought of in this perspective—studies of race effects on interpersonal encounters (e.g., Harrison, Messé, & Stollak, 1971).

While studies such as Kelley's demonstrate that in general, person perception mechanisms play a role in people's overt social behavior, they do not establish the link between such behavior and an important class of person perception variables: those factors in the perceiver (e.g., personality, cognitive structure, etc.) that affect his or her impressions and judgments (Kaplan, 1976; Shrauger & Altrocchi, 1964). Perceptual set, as it was induced by Kelley, is a situational variable, whereas physical appearance variables are primarily factors in the target person. Thus, the relationship between perceiver-based person perception variables and social behavior has not yet been clearly established.

The present research was an attempt to establish the link between social behavior and one perceiver-based person perception variable: the differential attention that persons pay to the behavior of others, especially children. Obviously, children emit a large number of cues or behaviors. Adults who observe them often evaluate these behaviors as "good" (acceptable, desirable, valuable, etc.) or "bad" (unacceptable, undesirable, immoral, etc.). We speculated that some people might be more attuned than others to the "bad" behaviors that children emit, and/or they might be more likely to interpret children's ambiguous behavior as bad; such people could be called negatively biased. Others might be more sensitive to children's good behaviors and/or be more likely to interpret behavior as good; these people would be positively biased. Still others might be more "balanced," that is, not be differentially sensitive to good or bad behavior and/or primed to interpret ambiguous behavior one way or the other. This speculation about "interpersonal perceptual styles" (IPS) is similar to the position of Kaplan (1976) that persons can have general response dispositions that influence their judgments of others.

In the present research, we explored the hypothesized link between IPS and interpersonal behavior in two social contexts: In Study 1, we examined the relationship between IPS and the behaviors that occurred during a cooperative encounter between an adult and a child; in Study 2, we explored the relationship between IPS and behavior within the context of a less cooperative interaction, a discussion between two adults of issues about which they disagreed.

Study 1

We measured IPS by examining people's perceptions of the same stimulus event—videotape excerpts from what was supposed to be a relatively long-term encounter between an adult and a child. We then observed encounters between children and people who, through their responses to the common stimulus event, were identified as negatively biased, positively biased, or perceptually accurate. Given the rather short-term and cooperative nature of this encounter, we expected (a) that IPS would relate more strongly to the behavior of the perceiver (i.e., the adult) than to the behavior of the child with whom she or he interacted and (b) that the behavior of negatively biased persons would differ (i.e., be more controlling and dominant) from that of positively biased and accurate persons, who, in turn, would not differ much from each other.

Method

Subject Recruitment and Selection

An advertisement was placed in the student newspaper of Michigan State University that solicited for persons who were interested in participating in "behavioral research" for pay. Interested persons were instructed to come to a 45-minute interview (dates, times, and place were listed in the advertisement) for which they would be paid $1. The advertisement also stated that those persons who were chosen to participate further in the project could earn $5 per hour for 1 to 3 hours of their time.

Table 1
Mean IPS Scores and Cell Frequencies

Sex of subject	Sex of child in videotape	Subject's IPS category: Positive Score	Positive *n*	Accurate Score	Accurate *n*	Negative Score	Negative *n*
F	F	6.80	14	−.24	13	−16.20	11
	M	7.02	12	−.07	10	−16.38	13
M	F	9.52	12	.07	12	−18.19	12
	M	9.07	14	.81	13	−16.30	11

Note. IPS = interpersonal perceptual style. An IPS score is the difference between the numbers (out of 25 each) of positive and negative behaviors that a subject reported perceiving the target child emit.

About 1,100 persons responded to the advertisement and came for the interview. These sessions were arranged so that potential subjects were examined in groups of 18–25 persons. Each person first completed a brief information sheet that asked his or her name, address, and telephone number, as well as the days and hours during the week that the respondent typically would be available to participate in the study. Then the group was shown the videotaped excerpts of the adult–child playroom encounters, under the guise that the project was concerned mainly with how various groups of people—professional psychologists, older adults, undergraduates, and so on—evaluated the activity of a "graduate student in training." After viewing the tape, the members of the group individually filled out a behavior checklist about the children in the videotape, supposedly so that the interpersonal context of the graduate student's evaluation could be assessed; then they completed instruments that were designed to tap their judgments about the graduate student.

The child behavior checklist was scored by subtracting the number of "negative" child behaviors that a respondent saw from the number of "positive" behaviors that she or he saw; such scores could range from +25 to −25, since the checklist contained 25 behaviors of each type (see below). There were 1,068 scorable checklists. From these, potential subjects were selected for the adult–child encounters such that there were 30 persons of each sex (plus any additional persons, if any, whose IPS score equaled that of the 30th potential subject selected) who best matched the definitions of positive bias, balance, and negative bias. These potential subjects met the following criteria: (a) They checked at least 20 behaviors in all; (b) they were identified as negatively biased if they checked at least 14 more negative behaviors than positive behaviors; (c) they were identified as accurate if the number of negative and positive behaviors that they checked did not differ by more than one; (d) they were identified as positively biased if they checked at least 5 more positive behaviors than negative behaviors.[1] Table 1 presents the mean IPS score and the number of each type of respondents who were selected to participate further.

Children Participants

Through the cooperation of the administration of the East Lansing, Michigan, school district, the parents of all children in second-grade classes were contacted via letter and asked if they would permit their child to volunteer (for $5) to play with an undergraduate in a playroom that was located at the Department of Psychology, Michigan State University. Parents who were interested returned a postcard to that effect, and a sample of these parents was contacted via telephone. In this conversation, the nature of the child's participation was explained in detail, and if the parent agreed, an appointment was made to have the child come to the playroom for the encounter with the undergraduate. In this way, the cooperation and participation of 168 7-year-old children were obtained.

Instruments and Materials

Standard perceptual stimulus (SPS). A script was constructed in which a "graduate student in training"—the word *clinician* was not used to avoid the possibility that the term would induce a negative set in subjects—acted rather passively in a series of encounters with a child. The child was identified as one of several volunteers who were paid to participate with the student. During these encounters, the child, while engaging in a variety of

[1] The criteria used to identify negatively and positively biased respondents differed substantially. This difference occurred because relatively few respondents (less than 2%) checked at least 14 more positive behaviors than negative behaviors, so that, of necessity, the identification criterion for positive bias had to be lowered.

activities, emitted a number of behaviors; some of these behaviors could be construed as negative (e.g., cheating at a game), some could be "seen" as positive (e.g., sharing food with the adult), and some could be perceived as neutral or ambiguous (e.g., painting with watercolors). An attempt was made to portray approximately equal numbers of positive and negative behaviors.

Two female and two male child actors (who ranged in age from 9 to 11 years old) learned the script and were videotaped while they played this role with a female graduate student. The SPSs of the female and male actors who were judged to do the best, most natural jobs were used to determine the perceptual styles of the undergraduates, as described above. About half the respondents saw the tape of the female target child, and half saw the tape of the male. Table 1 presents the number of male and female subjects of each perceptual style who were exposed to the male or female target child.

Child Behavior Checklist (CBC). The CBC, which was used to determine the interpersonal perceptual style scores of the respondents, was a modification of an instrument developed by Ferguson (e.g., Ferguson, Partyka, & Lester, 1974). In its present form it presented respondents with 64 "behaviors" (e.g., "Played with toys in a rough way"; "Did what the adult asked her [him] to do") and asked them to check those behaviors that applied to the child in the videotape that they had just viewed. Fourteen of the items were nonscored "fillers"; the remainder consisted of 25 "positive" behaviors and 25 "negative" behaviors that had differentiated parental perceptions of clinic-referred children from parental perceptions of non-clinic-referred children (Ferguson et al., 1974) and that a sample of about 100 undergraduates judged to be clearly positive or negative. As noted above, IPS was determined by subtracting the number of negative behaviors that a respondent checked from the number of positive behaviors that he or she checked.

Evaluation of graduate student in training. Two forms were used to permit respondents to evaluate the graduate student—the reason they were given for asking them to view the SPS. One was a modification of rating scales that were developed by Bessell and Palomares (1970); the other was a checklist composed of 36 trait words taken from Anderson (1968).[2]

Tasks

The adult subjects interacted with the child volunteers in playroom facilities that were located at the Department of Psychology, Michigan State University. Every adult–child pair engaged in three cooperative tasks, each of which lasted for 10 minutes. In the first ("free play"), both participants were told that they could do what they wanted (i.e., play together or separately, read, just sit there, etc.). In the second task, the pair was asked to reproduce to the best of their abilities a rather complicated figure on an "Etch-a-Sketch." In this task the adult had control of only one knob (and, therefore, one direction), and the child controlled the other knob. Their task was to coordinate their activities so that they produced a reasonable facsimile of the sketch. In the third task, the adult was given a list of proverbs (e.g., "Haste makes waste") and told to try to teach the meaning of them (as many as he or she wished) to the child. The interactions were videotaped through one-way mirrors.[3]

Design

This study used a factorial design whose dimensions were 2 (sex of undergraduate) × 2 (sex of target child; here target child refers to both the sex of the child in the SPS and that of the child with whom the subject interacted, since care was taken to insure that these were always the same) × 2 (IPS of the subject: positive, accurate, or negative, as defined above). Fourteen subjects, chosen at random from the 15 or more potential subjects in each condition, were examined in each cell of the design, but because of equipment failure and the other mechanical problems, final cell sizes ranged from 10 to 14 (see Table 1).

Procedure

All phases of the research were conducted by undergraduate assistants, although one of the first two authors was always available in case some difficulty arose. One assistant escorted the child into the playroom, and a few moments later, a second assistant brought the adult into the room, introduced him or her to the child, and then left. The first assistant then instructed the pair concerning the first task, after which she or he left the room and went to the observation room where a third assistant was videotaping the playroom activity. After 10 minutes, the assistant reentered the playroom to instruct the participants concerning the second task, again leaving the room once the instructions had been read.

---

[2] Since these instruments were not used further in this study, they are not described in any detail here. Note, though, that there was a relationship between these measures and the subjects' perceptions of the child in the videotape (mean $r = .21$). These results support Kaplan's (1976) hypothesis that dispositions in person perception tend to be applied in general within a social context. Thus, in the present context (i.e., the videotaped excerpts), the disposition to see the child as negative or positive appeared to generalize somewhat to perceptions of the adult as well.

[3] Of course, the undergraduates, the children, and their parents all were informed of this procedure beforehand, and all gave their informed consent about this (and all other) aspects of their participation in the study.

This same operating procedure was also employed for the third task. After the session was completed, both the undergraduate and the child were paid the specified amounts of money and thanked for their participation.

Coding of the Adult–Child Interactions

Eight undergraduates (who were kept unaware of the nature of the study until they had finished this task) were trained to use a modified circumplex system of classifying social behavior (Freedman, Leary, Ossorio, & Coffey, 1951). While the 16 specific categories of the system were retained, the scoring procedure was changed from rating scales of intensity to frequency counts of the number of social acts (Bales, 1950) emitted that fell within a given category.[4] Each videotape was scored independently by two coders, and the number of tapes that the various coder pairs scored ranged from 3 to 12.

The average interjudge reliability (correlation coefficient across coder pairs) for each circumplex category ranged from .29 (for suspect) to .93 (for hate), with an overall mean of .48. Because of the somewhat erratic interjudge reliabilities and because of large disparities across coders with respect to the total number of acts that were coded per tape, it was thought best to transform the raw data by converting them to z scores—within a coder within a category—and calculating the mean z score per category across each coder pair. In this way, coders who scored more acts would not influence the results unequally, and, as much as possible, random error would be reduced. Moreover, the frequencies with which interactants displayed categories of behaviors differed markedly. Thus, only half of the categories—structure, cooperate, passively question, dominate, helplessness, help, submission, and reassure—accounted for about 95% of the behaviors that were exhibited. Given the infrequency of behavior within the remaining categories, it was thought best not to examine them further.

Results

The dependent measures—the mean z scores within a coder pair for each of the eight most frequently displayed circumplex categories—were subjected to a 2 (sex of subject) × 2 (sex of target child) × 3 (IPS of subject) × 2 (interactant's score, adult's or child's, a within-dyad measure) multivariate analysis of variance (MANOVA).[5]

Effects of IPS

Given the theoretical framework of the present research, we expected that effects of IPS would appear as a function of who (adult or child) was emitting the behavior; that is, we expected IPS to affect the adult's behavior in this relatively brief encounter to a greater extent (and in different ways, if any) than it would the child's behavior. Thus, we expected the analysis of variance to yield a significant multivariate IPS × Interactant interaction. The MANOVA, in fact, indicated that this was the case, $F(16, 248) = 1.53$, $p < .08$.

Subsequent examination of the univariate IPS × Interactant effects revealed that five of the eight dependent measures yielded a significant interaction: dominate, $F(2, 131) = 5.47$, $p < .005$; structure, $F(2, 131) = 2.84$, $p < .075$; help, $F(2, 131) = 2.77$, $p < .075$; cooperate, $F(2, 131) = 2.88$, $p < .075$; and submit, $F(2, 131) = 4.55$, $p < .02$. These effects were analyzed further via tests of simple effects, and where appropriate, individual comparisons (Winer, 1971, p. 384) were performed between the negative and accurate conditions and between the accurate and positive conditions. Table 2 summarizes the results of these analyses by presenting those cell means that generated significant simple main effects for IPS.

The MANOVA also indicated that the Sex of Subject × IPS × Interactant multivariate in-

---

[4] We felt that the circumplex system was a very judgmental task for undergraduates, even in this simpler form. The somewhat erratic interrater reliabilities that were obtained confirmed that there was a need to be concerned about this issue.

[5] There is some controversy concerning the appropriate procedure to use to select α for a multivariate effect, especially when, as in the case of the present research, a number of dependent measures are used to test rather specific hypotheses (see, for example, Bock, 1975, pp. 422–425). The present research used a slight modification of a procedure suggested by Bock (1975, p. 423) in which the overall, multivariate α is a function of the product of increasingly more stringent univariate βs. For both studies, the β for the "first" univariate was set at .9, and each successive β was raised by a multiple of .025 or .015. Thus, given that Study 1 examined eight dependent measures:

$$\alpha_o = 1 - [(.9)\ (.925)\ (.95)\ (.975)\ (.99)\ (.9925)\ (.995)\ (.9975)] = .248.$$

Likewise, for the 16 dependent measures in Study 2, $\alpha_o = .250$.

Table 2
Mean Social Behavior Scores Reflecting Significant Effects Involving IPS

Circumplex category	IPS category of adult: Negative bias	Accurate	Positive bias	F[a]
	Adult's behaviors			
Dominate	.128[b]	−.060	−.088	3.37**
Structure	.168[b]	−.193	.012	3.71**
Help by males	−.111[b]	.068	.227[b]	2.37*
Submit	−.146[b]	.110	.053	2.69*
	Child's behaviors			
Dominate	−.195[b]	.078	.138	3.94**
Help with male adults	−.006[b]	.235	−.125[b]	2.69*

Note. IPS = interpersonal perceptual style. Values are mean z scores (standardized within each coder and averaged over each coder pair) for each circumplex category that was analyzed. Thus, each social behavior score had an overall mean of 0 and a σ equal to 1.
[a] $df = 2, 262$.
[b] Value differs significantly from accurate IPS mean.
* $p < .10$.
** $p < .05$.

teraction was significant, $F(16, 248) = 1.43$, $p < .13$. This effect reflected one significant univariate interaction, for help, $F(2, 131) = 3.17$, $p < .075$, which was examined further by simple effects tests and individual comparisons. These findings are also summarized in Table 2.

As Table 2 indicates, a number of reasonable results emerged from these comparisons. Negatively biased adults tended to display more dominance and less submission than did accurate perceivers; they also engaged in more acts of structuring. Moreover, negatively biased male subjects were less helpful than were their accurate counterparts. When compared to accurate perceivers, positively biased subjects also engaged in more structuring activity, whereas positively biased male adults tended to be more helpful.

Table 3 presents the mean scores for cooperative behaviors displayed as a function of IPS and interactant. While the simple effects analysis of this dependent variable did not reveal any significant effects for IPS within interactants, it did yield an interesting pattern of results when this variable was further examined from the perspective of differences between the interactants within each IPS category. Such a perspective is especially appropriate for cooperative behavior, because this activity can only be effective when it is expressed by both parties. The analysis indicated that children displayed a greater amount of this behavior than did the negatively biased adults with whom they were paired, whereas the reverse occurred when the adults were positively biased. However, the difference between children and adults was not significant when the adults had an accurate IPS. Thus, the dyads composed of a child and a perceptually balanced adult yielded the most even distribution of cooperative activity.

Three other findings relevant to IPS are worthy of note, since they also involve the behavior of children (see Table 2). The analyses revealed that children displayed less dominance toward negatively biased adults than towards accurate adults, and children were more helpful with accurate males than with either negatively biased or positively biased males.

Thus, taken together, these findings support the position that perceiver-based person perception processes are related to interpersonal behavior.

Sex Differences

The analyses also yielded two significant multivariate effects involving the sex of the interactants: the Sex of Subject × Interactant interaction, $F(8, 124) = 3.24$, $p < .005$, and the Sex of Child × Interactant interaction, $F(8, 124) = 2.02$, $p < .05$. While not a focus of this research, the results of the systematic exploration of the underlying significant ($p < .05$) univariate differences are summarized briefly here, in part because they reflect on the validity of the observational data: (a) Male adults emitted more acts of dominance than did females; (b) children displayed more dominance toward female adults; (c) females were more cooperative than males; (d) children were more cooperative with male adults; (e) male subjects were less submissive; (f) children were more submissive to males; and (g) male children displayed more helpful acts than did their female counterparts.

Discussion

The results supported the basic premise that served as the impetus for this research: Social behavior is related to perceiver-based person perception processes. Moreover, the specific findings that did emerge, on the whole, were reasonable.

Negatively biased adults tended to act to constrain the behavior of the child. This result is especially striking, given the social psychology of the observational setting—that is, the interactants were strangers to each other, their tasks were highly cooperative in nature, they knew they were being observed, and they interacted for a relatively brief period of time (30 minutes)—in which both adult and child most likely were "putting their best foot forward." The pattern of social behavior that negatively biased persons exhibited would be expected of those people who have a propensity to focus on children's negative behaviors, since it is likely that persons of this type would feel that they could not trust the child to act "appropriately" without "adequate" supervision.

Table 3
Mean Cooperative Behavior Scores as a Function of IPS and Interactant

Interactant	IPS of adult		
	Negative	Accurate	Positive
Adult	−.047	−.041	.102
Child	.122	−.089	−.048

Note. IPS = interpersonal perceptual style. As in Table 2, values are mean standardized scores—in this case, of the frequencies of cooperative behaviors that interactants displayed. Again, the overall mean was 0, and σ was 1.

On the other hand, positively biased persons, especially males, tended to be more helpful, if also somewhat more structuring. Again, such activity is reasonable in someone who is relatively inattentive to what most people would consider to be negative behavior in children. In this cooperative social context, though, a positive bias probably was congruent with the child's actual behavior in the playroom.

In addition, the behavior of the children to some extent was related to the adult's IPS, but not as much as we had expected. Perhaps the magnitude of the effect of adult IPS on the child's behavior is reasonable, given that this initial encounter between adult and child was brief and known to be transitory. It is interesting to speculate, however, that if IPS continues to affect adult behavior past the initial stage of an interpersonal relationship with a child, it could come to have a substantial impact on the child's behavior as well.

One conclusion that could be derived from the present findings is that a positive perceptual bias promotes positive interpersonal functioning. No doubt this was somewhat the case for the social context that we examined in Study 1—a cooperative, brief encounter with a child. However, it appeared reasonable to explore the possibility that positive bias would be related to less effective behavior in other social contexts—for example, when persons must deal with each other in an encounter that is not entirely cooperative. This possibility was the focus of Study 2.

Study 2

Specifically, in Study 2 the relationship between IPS and interpersonal behavior in a

mildly confrontational dyadic interaction was examined via three separate sets of dependent measures: a poststudy questionnaire, which measured the participants' reactions to their partner; observers' ratings of the subject and of the total interaction; and structural measures of the interaction, which consisted of the times to completion and the types of outcomes that occurred in a revealed differences task. We expected that, as in Study 1, IPS would be related to the social behaviors that occurred in an initial encounter between the perceiver and another person; and, since this encounter required that persons actively defend their opinions to someone whose position was opposed to theirs, we further expected that positively biased interactants would behave the least effectively.

Method

Subjects and Research Assistants

Subjects were 48 male and 48 female undergraduates who previously had participated in Study 1. Eight subjects were randomly chosen from the 10–14 subjects in each condition of that study, and each was paid $5 to participate in Study 2.[6]

Four confederates (two male and two female), four coders (two male and two female), and a male experimenter were selected from undergraduate psychology majors. These assistants were kept unaware of the purpose of the research and the reasons why subjects were selected to participate until all data were collected. They received course credits for participation in the study.

Design

This study used a $2 \times 2 \times 3$ factorial design, with sex of the subject, sex of the confederate, and IPS, respectively, as independent variables. Each cell contained eight dyads (subject and confederate). Subjects were paired with a confederate of the same sex as that of the target child whom they had viewed in the standard perceptual stimulus.

Procedure and Instrumentation

The subjects were observed interacting with a peer (confederate) in a structured situation that consisted of a revealed differences task. After the subject and confederate had completed a 17-item attitude questionnaire, the experimenter selected 3 items (those the subject had responded to with a strong opinion) and instructed the dyad members to *try* and reach an agreement about the items during the next 30 minutes.[7] All confederates were trained such that they could always advocate, in a rational and nonthreatening manner, a position opposite to whatever the subject defended. The procedure used in this study was very similar to the revealed differences technique developed by Strodtbeck (1951).

After completion of the interaction segment of the study, the subject and confederate were asked to fill out a poststudy questionnaire. In the first part, the respondents indicated along a 7-point scale how well each of 36 personality traits described their partner. The endpoints for the 7-point scales were labeled "very" and "not at all." The 18 positive and 18 negative trait scores were averaged to yield a total positive score and a total negative score. The second part of the questionnaire contained five questions. Four items consisted of evaluating the respondent's partner on 7-point dimensions of persuasiveness, likability, ability to win affection and liking from others, and ability to fit in with the respondent's circle of close friends. The remaining item assessed how comfortable the individual felt during the study.

Coding of Peer Interaction

Two coders independently rated the interaction occurring in a session. Each coder rated an equal number of subjects from each condition. Coders viewed the session through a one-way mirror and heard the interaction through headphones. The coders rated the dyad and subject using five 7-point semantic differential scales. The scales consisted of rating the total interaction (positive–negative) and the subject's

---

[6] Note that 2 to 3 weeks elapsed between subjects' participation in the two studies. Note also that subjects never were informed that the two studies were connected in any way, nor were they told why they were selected to participate. They were led to believe, when they originally responded to the advertisement, that a group of researchers who could pay people to participate in their separate studies were pooling their recruitment efforts. Of course, the personnel that subjects saw in each study were entirely different.

[7] Actually, to make training of the confederates a a more reasonable task, only 8 of the 17 issues were ever chosen for discussion: (a) legalization of marijuana use, (b) women's roles, (c) amnesty for war resisters, (d) the future of the family, (e) Ford's pardon of Nixon, (f) the Protestant Ethic, (g) ecology versus progress, and (h) the role of the university. Because of their salience to subjects at the time that the study was conducted, Topics a and e were almost always used, with the remainder used less frequently and more or less equally. Note, however, that there was no relationship between topics chosen for discussion and subjects' sex and IPS.

Table 4
Means of Measures Yielding Significant IPS Effects

Variable	IPS category: Negative bias	Accurate	Positive bias	*F*
Coders' overall rating of dyads containing male subjects[a]	4.93	5.21	3.46[b]	4.87***
Proportion of time spent reaching moderate disagreement with female confederate	.245	.278	.057[b]	3.24**
Coders' rating of friendliness when confederate was female[a]	5.57	5.67	4.55[b]	2.87*
Coders' overall rating of dyads containing female confederates[a]	5.14	4.92	3.55[b]	4.04**
Interactants' rating of each other's persuasibility when confederate was female[a]	4.03	3.70	2.93[b]	3.04**

Note. IPS = interpersonal perceptual style.
[a] Scores for these variables could range from 1 (very negative, unfriendly, etc.) to 7 (very positive, friendly, etc.).
[b] Value differs significantly from mean for IPS accurate category.
* $p < .075$.
** $p < .05$.
*** $p < .01$.

intelligence, comfortableness, friendliness, and nondefensiveness.[8]

Results

The dyad was used as the unit of analysis, with the exception of the four global ratings that the coders made only of the subject. The 16 dependent measures, comprised of the coders' ratings, the structural measures, and the poststudy questionnaire, were subjected to a 2 (sex of subject) × 2 (sex of confederate) × 3 (IPS) multivariate analysis of variance.[9]

The MANOVA revealed two significant multivariate effects involving IPS: the IPS × Sex of Subject interaction, $F(32, 100) = 1.34$, $p < .14$, and the IPS × Sex of Confederate interaction, $F(32, 100) = 1.25$, $p < .21$. These results reflected six significant univariate effects: The IPS × Sex of Subject interaction was significant for the coders' overall rating of the dyads, $F(2, 65) = 5.24$, $p < .01$; the IPS × Sex of Confederate interaction was significant for the proportion of time that dyads spent discussing issues that resulted in moderate disagreement, $F(2, 80) = 4.38$, $p < .02$; for the coders' rating of the subjects' friendliness, $F(2, 65) = 2.47$, $p < .1$; for nondefensiveness, $F(2, 65) = 2.59$, $p < .1$; for overall rating, $F(2, 65) = 3.36$, $p < .05$; and for the interactants' rating of each other's persuasibility, $F(2, 80) = 3.18$, $p < .05$.

Table 4 summarizes the results of univariate simple effects analysis by presenting the rele-

[8] Coders also scored the interactions in a more molecular fashion, using Borgatta's (1963) Behavior Scores System. While these data yielded a number of reasonable differences as a function of IPS, these findings tended to be redundant with, and somewhat less sensitive than, the data derived from the coders' semantic differential ratings. Thus, for reasons of efficiency, the results of the molecular coding are not presented in this report.

[9] Because of mechanical and procedural difficulties, we were unable to obtain coders' ratings for 15 dyads; moreover, assorted data were missing for 4 other dyads as well. Thus, the sample for the MANOVA had to be restricted to the 77 dyads for which there were complete data. However, it was thought best, where possible, to examine univariate effects—when the multivariate effect was significant—from analyses of variance that were performed on the larger sample.

vant means and significant F ratios. As predicted, the data presented in Table 4 suggest that positively biased subjects were the least effective participants in this mildly confrontational social encounter. Coders rated the interactions involving positively biased males the most negatively; likewise, coders were the least positive in their impression of interactions between positively biased subjects and female confederates. They also judged positively biased subjects to be the least friendly toward their female partners. Moreover, this combination of interactants spent the least time coming to moderate disagreement—which probably was the most appropriate outcome, since confederates were programmed to concede points very slowly—and rated each other as the least persuasive.

Sex Effects

The MANOVA also revealed that the multivariate main effect of sex of confederate was significant, $F(16, 50) = 4.24$, $p < .0001$. Examination of the underlying univariate effects revealed that while the coders judged subjects to be more intelligent and comfortable when they interacted with a female confederate, the interactants themselves were much less favorably impressed with each other (i.e., they rated each other as more negative, less likable, less persuasive, and so forth) when the confederate was female.

Discussion

The results from Study 2 replicated those of Study 1, in that, as predicted, IPS was related to overt social behavior. More specifically, in the somewhat confrontational social context of Study 2, positively biased perceivers engaged in the most dysfunctional interaction, especially when paired with a female partner. During the interaction, the female confederates were assertive, confident, and rational, which is contrary to the common stereotype of females as submissive, weak, and emotional. Perhaps the positively biased perceivers had the most difficulty with female confederates because their self-image is firmly grounded on the traditional stereotype of what a male and female should be. Therefore, when the female did not behave according to her traditional role, the positively biased perceivers might have felt the most threatened.

Taken together, the two studies suggest that persons with an accurate IPS exhibit the most adaptive social behavior across social contexts. Persons with a negative IPS were reasonably adept at interacting within a situation that contained an element of conflict (i.e., the revealed differences discussion), but were less able to behave appropriately in a more cooperative setting. Persons with a positive IPS showed the opposite pattern; they did very well when interacting with a child who was on his or her "good behavior," but appeared to be somewhat inept when confronted with a mild social conflict.

Thus, the two studies demonstrate that there is a link between at least one perceiver-based person perception process and social behavior. Clearly, though, further work is needed. The theoretical perspective that served as the basis of this work (e.g., Combs & Snygg, 1959) postulates that perceptual mechanisms like IPS cause social behavior. It also could be, however, that IPS and interpersonal behavior are related in part because they both are manifestations of some other variable (e.g., need for security). This latter interpretation is plausible, and it is likely to be valid to some extent. IPS is likely to be determined by a number of personality variables that also have a direct impact on social behavior. But it is also likely that there is a direct, causal link between person perception and social behavior as well. As with many phenomena in social psychology, it is probably the case that person perception processes like IPS are both concomitant variables with, and causes of, interpersonal behavior. Establishing empirically the precise relationship between IPS, other components of personality, and overt social behavior would be a most difficult task. However, the findings of the present research suggest that the undertaking of this task would be a worthwhile and fruitful endeavor.

References

Anderson, N. H. Likableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology,* 1968, *9,* 272–279.

Bales, R. G. *Interaction process analysis: A method for the study of small groups.* Cambridge, Mass.: Addison-Wesley, 1950.

Bessell, H., & Palomares, U. H. *Methods in human development: Rating scale booklet.* San Diego: Human Development Training Institute, 1970.

Bock, R. D. *Multivariate statistical methods in behavioral research.* New York: McGraw-Hill, 1975.

Borgatta, E. F. A new systematic interaction observation system: Behavior scores system. *Journal of Psychological Studies,* 1963, *14,* 25–44.

Combs, A. W., & Snygg, D. *Individual behavior: A perceptual approach to behavior* (Rev. ed.). New York: Harper & Row, 1959.

Cottrell, L. S., Jr. The analysis of situational fields in social psychology. In A. P. Hare, E. F. Borgatta, & R. F. Bales (Eds.), *Small groups: Studies in social interaction* (Rev. ed.). New York: Knopf, 1966.

Ferguson, L. R., Partyka, L. B., & Lester, B. M. Patterns of parent perception differentiating clinic from non-clinic children. *Journal of Abnormal Child Psychology,* 1974, *2,* 169–181.

Freedman, M. B., Leary, T. F., Ossorio, A. G., & Coffey, H. S. The interpersonal dimension of personality. *Journal of Personality,* 1951, *20,* 143–161.

Harrison, M. G., Messé, L. A., & Stollak, G. E. The effects of racial composition and group size on interaction patterns in preschool children. *Proceedings of the 79th Annual Convention of the American Psychological Association,* 1971, *6,* 325–326. (Summary)

Hastorf, A. H., Schneider, D. J., & Polefka, J. *Person perception.* Reading, Mass.: Addison-Wesley, 1970.

Kaplan, M. F. Measurement and generality of response dispositions in person perception. *Journal of Personality,* 1976, *44,* 179–194.

Kelley, H. H. The warm–cold variable in first impressions of persons. *Journal of Personality,* 1950, *18,* 431–439.

Kelly, G. A. *The psychology of personal constructs: A theory of personality* (2 vols.). New York: Norton, 1955.

Kelly, G. A. Ontological acceleration. In B. Maher (Ed.), *Clinical psychology and personality: The selected papers of George Kelly.* New York: Wiley, 1969.

Kleck, H. H., Ono, H., & Hastorf, A. H. The effects of physical deviance upon face-to-face interaction. *Human Relations,* 1966, *19,* 425–436.

Mead, G. H. *Mind, self and society.* Chicago: University of Chicago Press, 1934.

Rogers, C. R. A theory of therapy, personality and interpersonal relationships as developed in the client-centered framework. In S. Koch (Ed.), *Psychology: A study of a science* (Vol. 3). New York: McGraw-Hill, 1959.

Shrauger, S., & Altrocchi, J. The personality of the perceiver as a factor in person perception. *Psychological Bulletin,* 1964, *62,* 289–308.

Snyder, M., Tanke, E. D., & Berscheid, E. Social perception and interpersonal behavior: On the self-fulfilling nature of social stereotypes. *Journal of Personality and Social Psychology,* 1977, *35,* 656–666.

Strodtbeck, F. L. Husband–wife interaction over revealed differences. *American Sociological Review,* 1951, *16,* 468–473.

Taguiri, R. Person perception. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (Vol. 3, 2nd ed.). Reading, Mass.: Addison-Wesley, 1969.

Winer, B. J. *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill, 1971.

Received March 17, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 3, 380–390

Affective States, Expressive Behavior, and Learning in Children

John C. Masters, R. Christopher Barden, and Martin E. Ford
University of Minnesota

Two experiments with children are presented that illustrate the effects of emotional states on learning and validate experimental affect-induction procedures in which individuals dwell upon thoughts of affect-provoking experiences. Positive affective states enhanced learning, and negative states retarded it dramatically. Ratings of children's facial expressions confirmed that positive affect-induction procedures elicited happy expressions, and negative inductions elicited sad ones. Additionally, positive affect inductions enhanced children's apparent interest, involvement, and arousal, and negative inductions decreased them. These measures were related to learning but proved not to be the sole mediators of the impact of affective states on learning. The thoughts children generated for affect induction illustrated their recognition of naturalistic experiences that induce affective states. These results indicate that young children possess the potential for the cognitive self-control of their own affective states, and the effects on learning indicate that even transient mood states may produce lasting changes in behavior.

Experimentally Induced Affective States

Studies of experimentally induced affective states pose a significant contribution to the scientific study of such states and the determining role that emotions may play in shaping other behavior patterns. It has already been clearly demonstrated that induced affective states in children affect patterns of self-gratification (Moore, Underwood, & Rosenhan, 1973; Rosenhan, Underwood, & Moore, 1974), altruism (Harris & Siebel, 1975; Moore, et al., 1973; Rosenhan et al., 1974; Underwood, Froming, & Moore, 1977), and aggression (Harris & Siebel, 1975), and mood states thus seem likely to determine the performance of a broad range of social behaviors, many of which are so far unstudied. There are also compelling arguments for the possibility that affective states may influence the effectiveness of learning processes and may consequently have far-reaching effects on the development of new behavior patterns through social learning or on intellectual functioning in achievement settings.

Consequences of Affective States for Behavior and Learning

Theoretically, the effects of a given affective state may be due to a variety of mediating processes, perhaps different ones, depending upon the behavior pattern being affected. For behaviors such as self-gratification or altruism, it has been proposed that the pleasant or aversive nature of a given state serves as a motivational determinant for behaviors that will maintain the state if it is positive or reverse the state if it is negative (Moore et al., 1973; Rosenhan et al., 1974; Underwood et al., 1977). The influence of affective variables on persistence at effortful behavior may be mediated by reinforcement effects, if the variables are of positive or negative valence, such as favorable or unfavorable self-evaluation

Preliminary versions of this experiment constituted an undergraduate summa cum laude honors thesis by the second-named author. The able assistance of Craig Binger is gratefully acknowledged.

Requests for reprints should be sent to John C. Masters, Institute of Child Development, 51 East River Road, University of Minnesota, Minneapolis, Minnesota 55455.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3703-0380$00.75

(Masters & Santrock, 1976) and if they are consequent to the effortful behavior. There is some evidence, however, that the impact of self-evaluations and their emotional concomitants may affect learning through incentive or other motivational mechanisms that are not consequent to learning but actually occur in anticipation of intellectual mastery (Masters, Furman, & Barden, 1977). Thus, mood states bearing no contingent relationship to performance may affect performance, learning, or mastery not through reinforcement processes but through motivational or arousal components.

Affective states may also have an indirect effect on behavior and learning—an impact mediated by the actions of others whose behavior towards an individual is influenced by that individual's apparent affective state. There is ample evidence that people's beliefs and biases about the characteristics of others influence their behavior towards these others (Mischel, 1968; Rosenthal, 1966, 1969). Certainly, a child who appears happy or sad, interested, involved, or aroused may elicit different treatment from parents, teachers, peers, or other agents of socialization, and this treatment is then likely to influence behavior and apparent competence in social or intelectual contexts. Facial expressions of emotion are common and socially significant. They are also scientifically significant and have been used successfully to index internal affective states (Ekman, Friesen, & Ellsworth, 1971; Ekman et al., 1972). However, investigations utilizing experimental affect-induction procedures have failed to gather direct validating evidence, such as changes in affective facial expressions. The demonstration of appropriate changes in facial expressions following different affect-induction manipulations would strengthen the assumption that affective states were truly induced, as well as give some indication of changes in a child's social stimulus characteristics as a function of the affect-inducing experience.

Two experiments are presented that examine the effects of positive (happy), neutral, and negative (sad) affective states on children's mastery of a learning problem. In addition to the affective valence, the tempo of a given emotional state was independently manipulated within an induction procedure to establish active or passive forms of the positive, neutral, and negative affective states.[1] Three basic predictions were made: (a) Positive affective states will enhance learning, and negative states will interfere; (b) a tempo component that is active will promote effective learning, and one that is passive will not; and (c) induced active tempo will increase the speed of ongoing learning behavior (average time to solve individual problems). These predictions were based on the following reasoning:

1. Positive states may also enhance the effectiveness of social reinforcers (right–wrong feedback delivered by an experimenter; Byrne & Clore, 1970; Gouaux & Gouaux, 1971; Izard, 1964).

2. Negative affective states will interfere with learning through distraction, by diverting attention away from the task and onto attributes of the self (Mischel, Ebbesen, & Zeiss, 1973) and by failing to induce any incentive for mastery (Masters et al., 1977).

3. Positive affective states may also motivate greater attention to the task plus a striving for success, since success experiences will maintain the affective states (Masters et al., 1977; Mischel et al., 1973; Rosenhan et al., 1974).

4. Induced active tempo will enhance learning accuracy through increased arousal, interest, and involvement in the task, compatible with the long history of research bearing upon the facilitative effects of moderate arousal on learning (e.g., Izard, 1964; Izard, Wehmer, Livsey, & Jennings, 1965; Velten, 1968; Woodworth & Schlosberg, 1954).

Experiment 1 was conducted to evaluate the accuracy of the predictions regarding the effects of affect and tempo on learning. Ex-

[1] It should be noted at this point that the interaction of arousal and valence may produce different labels for affective states (e.g., active–positive might be termed elation; passive–positive could be serenity; active–negative might be agitated depression or anger; passive–negative might be depression, sadness, or grief). Such labels will not be used here, nor were they utilized in the experiments, since the influence of such labeling was not under investigation.

periment 2 replicated Experiment 1 and in addition provided data relevant to hypotheses concerning other consequences of affective states. In this experiment, children's facial expressions were recorded on videotape following the affect and tempo induction procedure; these expressions were rated according to the degrees of various emotions that were apparent and according to expressional indexes of factors related to learning, such as interest, involvement, and arousal (Ekman et al., 1971; Ekman et al., 1972). For purposes of convergent and discriminant validation of the affect-induction procedure, children's facial expressions were rated on convergent dimensions of happiness (positive affective state) and sadness (negative affective state) and on divergent dimensions as well (e.g., fear, disgust, surprise). To provide additional validity evidence, the thoughts that children generated for the affect induction were independently rated for affective tone by a panel of adult raters.

Method

Subjects

For each experiment, 48 4-year-old children served as subjects. Children were drawn from nursery schools in a large metropolitan area. They were equally divided by sex and randomly distributed according to socioeconomic status and racial background. The experimenter was a male adult.

Procedural Overview

For both experiments, children were brought to an experimental room by the experimenter. Then the affect induction commenced and was followed by the learning task. At the end, children were asked to recall the thought that they had generated for the affect induction, and then, if they had been in a negative affect condition, a positive affect induction was included to offset any lingering effects of the negative induction. In Experiment 2, just before and immediately following the learning task, children were asked to sit for a moment, during which time their facial expressions were silently recorded via a video camera standing in the corner of the room.

Experimental Manipulations

There were three levels of affect and two of tempo, resulting in six experimental conditions. The affect-induction procedure was similar to that employed by earlier investigators (Masters & Furman, 1976; Rosenhan et al., 1974) and involved asking the child to generate a thought of the appropriate affect and tempo and to concentrate on it for 30 sec. The child did not communicate the content of his or her thought to the experimenter until the experiment was completed, to minimize any experimenter bias. At that time, a record was made of the thought content. The affect and tempo induction was as follows:

> You know, (subject's name), sometimes things happen to us that make us feel happy, and sometimes things happen to us that make us feel sad, and sometimes things happen to us that don't make much difference to us. Can you remember something that happened to you that made you feel:
>
> (positive affect, active tempo) so happy that you just wanted to jump up and down?
> (positive affect, passive tempo) so happy that you just wanted to sit and smile?
> (neutral affect, active tempo) like jumping up and down?
> (neutral affect, passive tempo) just like sitting?
> (negative affect, active tempo) so sad that you wanted to jump up and down?
> (negative affect, passive tempo) so sad that you just wanted to sit and frown?
>
> Now (subject's name), I want you to use your imagination, all right? Our imagination really works sometimes, doesn't it? It can really make us believe things, can't it? All right now, (subject's name), just look at my light box (the learning equipmment) and really use your imagination for a few seconds and think of something that makes you feel (same as above). That's right, think of something that makes you feel (same as above).

The child was then given 30 sec to stare at a plain gray box and think.

Learning Task

Following the affect/tempo induction procedure, children were introduced to the learning task, one involving shape discrimination (Masters et al., 1977). There were 12 individual discrimination problems making up a trial block, and a "tower of lights" indicated a child's performance by illuminating a series of lights, one for each problem solved correctly by the child. Children were given up to 10 trial blocks to achieve perfect mastery of the 12 problems, and an additional trial block was run for those children whose performance was less than perfect on the 10th trial, in order to make certain that the right–wrong feedback during that trial had not allowed them finally to achieve mastery. There were three stimuli in each problem, one of which was designated correct, and their order in a horizontal array was varied for each presentation of the problem. The measures of rate and accuracy of learning utilized for purposes of analysis were the amount of time taken to reach perfect mastery, the number of trials to perfect

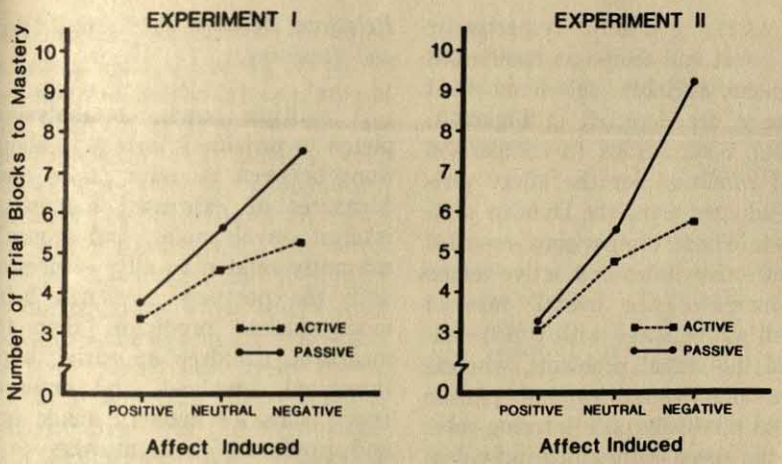


Figure 1. Mean number of trial blocks (maximum = 10) to perfect mastery for both experiments as a function of affective valence and active or passive tempo.

mastery, the total number of errors, and the mean amount of time spent considering individual problem solutions.

Independent Measures of Expressed Affect

In Experiment 2, children's facial expressions were videotaped for a 30-sec period at two times during the experimental session. The first recording occurred immediately following the affect-induction period and just prior to the learning task (pretask), providing an index of the impact of the induction procedure and allowing the assessment of affect, attention, and arousal measures that might be predictive of learning performance. The second recording occurred immediately upon the completion of the learning task (posttask) and tapped the continued impact of the induction procedures as they had persisted throughout the learning task, plus any effects generated by the learning experience itself.

Facial expression ratings were completed according to the categories defined by Ekman (Ekman et al., 1971). The following categories were utilized: happiness, sadness, pleasantness, interest, involvement, arousal, disgust, fear, pain, and surprise. Two raters independently viewed the videotaped segments without knowledge of the nature of the experiment or the hypotheses involved. They were alerted to consider the duration of a facial expression, its relative intensity, and the frequency of its occurrence during the 30-sec period, in order to judge the intensity (Ekman et al., 1972). All ratings were on a 9-point scale, with 1 equated to the absence of a particular state but not the presence of its opposite (e.g., 1–happiness did not imply any degree of sadness). Rater reliability was assessed for all subjects in terms of percentage of agreement between the raters and ranged from .80 to 1.00 across all dimensions (median agreement = 94%).

Thoughts Utilized in Affect and Tempo Inductions

The thoughts that children generated on the affect induction in Experiment 2 were submitted to five adult judges who rated them on both affective tone and tempo. After the initial ratings were completed, the raters were informed of the experimental manipulations and were allowed to read the experimental instructions for all conditions. They were then asked to review the thoughts once more and estimate (a) which affect induction the child had experienced and (b) which tempo induction the child had experienced. Average interrater agreement across all five raters ranged from .65 to .92, with a median value of .73.

Results

Effects of Induced Affective States on Speed and Accuracy of Learning

The four learning measures (total time to mastery, number of trials to mastery, total number of errors, and mean amount of time spent on each problem) were significantly intercorrelated; consequently, a single multivariate analysis of variance was performed. The results were identical for the two experiments, and since Experiment 2 also contained the measures of affective expression, only the statistical results of Experiment 2 are specifically reported.

The multivariate analysis revealed significant main effects due to both affect and tempo, $F(8, 144) = 53.59$, $p < .001$, and

$F(4, 144) = 43.11$, $p < .001$, respectively. The effects of affect and tempo as manifested on one dependent variable, the number of trials to mastery, are depicted in Figure 1. Individual group comparisons (a comparison to the neutral condition for the affect variables) were conducted using the Duncan multiple-range rest. These comparisons revealed that positive affective states and active tempo significantly increased the overall rate of learning as well as the speed with which children processed individual problems, whereas both negative affective states and passive tempo produced relatively poor learning rates and retarded the speed with which individual problems were processed.

The interaction between affect and tempo also proved significant, $F(8, 144) = 12.11$, $p < .001$. In general, the difference between active and passive tempo was enhanced for the negative affect condition, although this was less true for the mean time spent on each problem than it was for the other learning measures. This interaction, which can be seen particularly well in Figure 1 with respect to one learning variable, may be due to a partial ceiling effect for the positive affect conditions, in which children often learned in as few trial blocks as possible, many showing perfect mastery after only the second block of learning trials (accuracy on the first trial block was chance, since correct solutions were not given in advance).

It is of interest to note here the contrast between the different effects of affective states on the overall number of trials to task mastery and the speed with which children produced solutions for individual problems. Those affect and tempo conditions that promoted fewer errors and rapid overall mastery in fewer trials also increased the speed with which children solved individual problems. Under conditions of various affect and tempo inductions, the "impulsive," rapid answer was more likely to be correct, and the slower, "reflective" response incorrect. The correlation between number of trials to mastery and mean time per problem was $r(48) = .47$, $p < .01$.

Relations Between Expressed Affect and Learning

A multiple regression analysis was completed to provide information about the relations between affective states and learning. Measures of expressed happiness, sadness, interest, involvement, and arousal were significantly related to all measures of learning, with the posttask measures being slightly more efficient predictors than the pretask measures. Children appearing happier, more interested, involved, and aroused required fewer trials for mastery, made fewer errors, and produced their answers to individual problems more rapidly. Children appearing sadder and not particularly interested, involved, or aroused required more trials, made more errors, and spent longer on the solution of individual problems.

Since the measures of expressed affect and of interest, involvement, and arousal were quite highly correlated with one another, median $r(48) = .51$, further multiple regression analyses were undertaken to determine the degree to which expressed affect and interest/arousal variables contributed to the variance in learning. The multiple correlations between the prediction set and two measures of overall rate and accuracy of learning were .72 and .64 ($df = 6, 41$, $p < .01$) for trials to mastery and number of errors, respectively. Significance tests for individual beta weights revealed that expressed happiness and sadness were the most powerful predictors of the rate and accuracy of overall learning, and expressions of apparent task involvement or arousal contributed little. The multiple correlation using affect and interest/arousal variables to predict processing speed for individual problems was $R(6, 41) = .51$, $p < .05$. Only expressed sadness had a significant beta weight in the predictor equation for the amount of time spent on individual problems—a result that interestingly parallels the clinical picture of depression as one of psychomotor retardation.

To test further the implication that expressed pretask and posttask interest, involvement, and arousal were less powerfully related to learning than affect, a stepwise multiple regression analysis was performed,

this time entering the predictors in the equation in two orders. First the affect predictors were entered, followed by the interest/arousal variables, and then the reverse. For the rate and accuracy of overall task mastery, each set of variables accounted for a significant amount of variance, but the magnitude of the contributions was very different. The affect measures added between 20% and 30% more variance to the equation when added after the interest/involvement/arousal variables, whereas the latter added only from 1% to 3% more variance when the affect measures had been included first (but this slight addition was still significant). With respect to the amount of time children spent on individual problems, the stepwise regression analysis revealed that the interest/involvement/arousal variables did not contribute significantly to the predictor equation.

These results indicate that positive and negative expressed affective states are strongly associated with the overall rate and accuracy of children's learning, and negative states influence the speed of cognitive processing, or at least the rapidity with which a solution is reported. Apparent interest, involvement, and arousal are also related significantly to learning, but the relation is modest at best, and it does not appear that these variables are major mediating factors in the relation between affective states and learning.

Convergent and Discriminant Validation of Affect-Induction Procedures

Facial expression ratings. In order to assess the convergent and discriminant validity of the affect-induction procedures, analyses were conducted to determine the effects of the various experimental conditions on pretask ratings of affective states from the facial expression data. Two types of convergence and discrimination were examined: (a) the degree to which affect-induction manipulations influenced appropriate facial expression ratings and tempo-induction manipulations did not and (b) the degree to which the affect-induction procedures influenced the appropriate affect ratings (pleasantness, happiness, sadness) in an understandable fashion and did not influence ratings of divergent affective states (anger, disgust, surprise, pain).

Table 1
Mean Degree of Expressed Affect Following Positive, Neutral, and Negative Affect Inductions

Affect induction	Affective expression rating[a]	
	Happiness	Sadness
Positive	6.76	1.03
Neutral	3.91	1.44
Negative	2.91	3.04

Note. $n = 16$ for each mean.
[a] Maximum = 9.00; minimum = 1.00.

Analyses of the effects of sex, affect induction, and tempo induction on pretask ratings of happiness and sadness revealed significant effects only for affect induction (happiness: $F(2, 36) = 24.99$, $p < .01$; sadness: $F(2, 36) = 5.44$, $p < .01$. Means are presented in Table 1, and they reveal clearly that the effects were appropriate for the different induction conditions. Individual group comparisons indicated that for happiness, the positive affect condition differed significantly from both the neutral and negative conditions. For sadness, the negative affect condition differed significantly from the neutral and positive conditions, and the positive condition differed from the neutral condition as well. Figure 2 illustrates the facial expressions of four children, two following the induction of a positive state and two following the induction of a negative state.

Discriminant validity was revealed by an examination of the effects of the affect-induction procedures on the ratings of divergent emotional states. There were no significant effects whatsoever from the affect-induction procedures on ratings of anger, disgust, surprise, or pain. A significant main effect of induced affect was observed for ratings of fear, $F(2, 36) = 5.01$, $p < .05$. These ratings were all quite low, and children in the neutral condition were rated as showing somewhat more fear ($M = 2.16$) than children in either the positive ($M = 1.07$) or negative ($M = 1.25$) affect conditions. This finding

Figure 2. Facial expressions, photographed from videotape, of four children participating in Experiment 2—two from the positive affect-induction conditions and two from the negative affect-induction conditions.

may indicate simply a mild degree of apprehension in children after the difficult assignment of determining a neither happy nor sad experience to dwell upon, with no feedback given regarding their success.

Analyses of affect- and tempo-inducing thoughts. The thoughts generated by children were rated for their degree of intrinsic happiness, sadness, activity, or passivity. These ratings were analyzed to determine the degree to which the experimental conditions of affect and tempo induction inclined young children to generate thoughts that were identifiable by adults as matching the intended affect or tempo manipulation. The ratings were submitted to two-way analyses of variance with primary dimensions of affect and tempo induction. The means for all analyses are presented in Table 2.

The analysis of happiness ratings revealed a significant main effect for the affect induction, $F(2, 23) = 22.67$, $p < .01$. The thoughts produced by children to effect the positive affect induction were rated significantly more happy than those produced by children in the neutral condition, who in turn produced thoughts judged happier than those from children in the sad condition. Analysis of sad ratings also revealed a significant main effect of affect induction, $Fs(2, 23) = 21.20$, $p < .01$. The thoughts produced by children in the sad condition were rated more sad than those from children in the neutral condition, who in turn produced thoughts rated more sad than those from children in the happy condition.

Analysis of the activity and passivity ratings revealed main effects of tempo, $Fs(1, 139) = 5.82$ and 6.51, $p < .01$, for activity and passivity, respectively, and a significant interaction between activation and affect induction, $Fs(2, 23) = 8.47$ and 7.46, $p < .01$. Thoughts from children in the active tempo conditions were rated significantly more active

Table 2
Mean Ratings of the Happiness, Sadness, Activity, and Passivity of Children's Self-Generated Thoughts as a Function of Affect and Tempo Inductions

Rating dimension	Tempo induction	Affect induction: Positive (n = 16)	Neutral (n = 16)	Negative (n = 16)
Happiness	Active	7.95	5.35	2.16
	Passive	7.07	4.05	1.93
Sadness	Active	1.25	2.60	6.72
	Passive	1.23	2.60	6.43
Activity	Active	5.85	7.05	3.60
	Passive	4.50	2.70	5.10
Passivity	Active	2.70	2.20	5.12
	Passive	4.07	6.30	3.93

than those from children in the passive conditions, and the reverse was true for the passivity ratings. However, these findings were qualified by the interactions. Thoughts from children in the neutral condition were clearly recognizable as being from either the active or passive activation induction according to both rating dimensions. However, this distinction was greatly reduced for thoughts from children in the positive affect condition. Furthermore, the thoughts generated by children in the negative affect condition were commonly *mis*judged by raters. There was a tendency to judge thoughts from the passive–sad condition to be more active than those from the active–sad condition, and vice versa for the passive ratings. These results indicate that the presence of a clear affective tone, especially a negative one, tends either to confuse children in their generation of active or passive scenes or to produce thoughts that are active or passive to them but not to adult raters judging those thoughts. The latter seems to be the more likely explanation, since the experimental results revealed effects of tempo on learning, without any significant reversals as a function of the affect induced.

Discussion

Role of Affect in Social Learning

The finding that affect strongly influences learning enhances our understanding of the role that affective variables may play in development. Mood states may be transient, but if they influence adaptive learning, their consequences may be widespread and of long duration. The current affective state of an individual in any context is an important person variable that is likely to mediate any influence of cognitive or contextual learning variables and may well constrain or amplify the possibility for prior experience and learning to mold adaptive behavior. The present findings also indicate that the impact of emotional states should not be underestimated. The affective states induced in this experiment were surely less varied than the states induced by uncontrollable experiences children have in the natural environment. Most of the thoughts generated by children represented social experience with parents, teachers, or peers, and the thoughts must surely have induced weaker affective states than did the original experiences.

The implications of the present findings extend beyond the academic or intellectual context. Although the present experiment utilized an intellectual learning task, there seems little reason to believe that affective states will not also affect children's social learning in nonintellectual contexts. Thus, children's sensitivity to social reinforcement or punishment or to the behavior of peer or adult models may depend at least in part upon the children's ongoing affective state. The effectiveness of social learning variables, punishment for instance, may also be mediated

by the affective states that are induced by the social learning variable itself during the learning interaction. The role of affect as a participant or mediating variable in intellectual or social learning simply has not yet been explored.

The fact that children's self-generated thoughts can create intended affective states raises the possibility that children possess at least the potential ability to manipulate their own affective states and possibly intervene in a therapeutic fashion to terminate the persistence of a negative affective state or enhance an existing positive state. There is some evidence to indicate that children may ameliorate unpleasant experiences and affective states by selective exposure to contexts that are likely to induce opposing states or provide desired experiences (Masters, Ford, & Arend, Note 1). The present results indicate that the "power of positive thinking" or a rational–emotive analysis (Ellis, 1962, 1971) is not without application to young children, and cognitive abilities to induce affective states may very early become part of an individual's armamentarium of self-regulatory skills, providing some degree of personal control over mood states and their consequences.

Tempo and Affect

The results of the present two investigations indicate that the tempo component of an affective state may be an important contributor to the effects achieved on a broad range of behaviors. The learning data also revealed that inductions with an active tempo enhanced learning in much the same way that the induction of a positive affective state did. Nevertheless, in many instances three was a significant interaction between activation and affect, illustrating an apparent interdependence of the valence and activation components of specific affective states. Labels exist in the language that appear to be appropriate for different combinations of affect and tempo (e.g., positive/active = elation or joy; positive/passive = serenity, pleasure, or perhaps contentment), and the present findings indicate that a conceptual dissection of emotional states into hypothesized dimensions beyond merely that of affective valence is a heuristic procedure for future research.

Children's Comprehension of Affect and Affect-Inducing Experiences

Little is known about children's concepts of affect, but the present results indicate that even those of preschool children correspond quite well to those of adults. Preschool-age children's comprehension of the affect that they do feel or are expected to feel in various circumstances is well developed, and they comprehend the determinants of a given affective state well enough to propose appropriate antecedent conditions. In the present experiment, it was difficult to tell which of the thoughts that children generated were recollections of their own past experiences and which were more normative nominations of experiences anyone might have that would produce the affective state in question. Some thoughts were of everyday experiences: sitting and watching TV (neutral), helping to clean up the house (negative), jumping on one foot (neutral), or going on a vacation to Chicago (positive). Others, however, appeared to be concocted rather than recalled, such as having a turtle throw a shell at one (negative) or when Santa comes and leaves no present (negative). The response of at least one child indicated an awareness of vicarious experiences that may produce an affective state: watching a friend on a bike make a funny jump (positive). Very young children thus appear to have the ability to identify affect-inducing experiences and a comprehensive understanding of culturally defined affect labels. The dimensions of these skills and knowledge are at present only poorly understood.

The diversity of content spontaneously utilized by children in generating the affect-inducing thoughts also suggests the need for further investigation into the classes of experiences that may generate affective states. Just as the effects of success or failure on levels of aspiration are constrained by the contextual character of that experience, so that, for example, academic failure will not necessarily lower aspirations for athletic accomplishments, the contextual nature of af-

affect-inducing experience may focus the impact of the induced state upon relevant or related subsequent behaviors. For example, in the present experiments, some children generated thoughts that were distinctly social in nature (dad helping me jump up and down; when grandma visits; when my father spanked me), whereas others thought about contexts that were private (candy; sitting and watching TV; breaking toys accidentally). A private–social dimension such as this may be relatively unimportant for a private activity like learning but be influential in determining subsequent social behavior such as altruism. Similarly, some children generated thoughts about experiences with intrinsic social comparison components (everybody's eating candy but me), whereas others nominated contexts that contained peers, but without any comparative overtones (when my friend comes over and then leaves in a huff; when a friend calls me a copycat). In short, the effects of affective states on various social and intellectual behaviors may be discriminated according to aspects of the precipitating experiences other than simply the affective valence.

Factors Mediating the Behavioral Consequences of Affective States

The factors mediating the effects of affective states on behavior have rarely been explored. In the present instance, expressive indicators of both pretask and posttask interest, involvement, and arousal very clearly were influenced by induced affective states and were related to children's learning. However, careful analysis revealed that interest/arousal factors were related to learning in a manner independent of the affective factors, and they were clearly not a prime mediator for the effects of affective states on learning. The mediating role of attentional and arousal factors cannot be ruled out, because of the possibility that facial expressions are not the best indicators of ongoing attentional behavior during learning. There is growing evidence, however, that changes in attentional behavior are not a major consequence of affective states. Underwood et al. (1977) have already shown that measures of incidental attention are unaffected by induced affective states. More recently, Isen, Shalker, Clark, and Karp (1978) found that positive affective states affected the recall of positive memories, which suggests that the impact of affective states on learning may occur during retrieval or recall performance rather than during attentional or encoding phases.

Validation of Experimental Affect-Induction Procedure

The inclusion of independent assessments of affective states provided clear convergent and discriminant validity for the experimental affect-induction procedure. The availability of techniques to manipulate emotional states under controlled conditions and in normal individuals is an important methodological advance, and there should be further attempts to demonstrate the validity of this manipulation through the use of independent indexes of induced affective states. In the present case, the use of expressed affect and interest/arousal indexes also provides information about another aspect of emotional states, the degree to which they may alter the social stimulus characteristics of the person experiencing them. Among the factors mediating the effects of emotional or mood states on behavior are the social reactions of others, reactions which may in part be determined by the appearance of the person experiencing the affective state. The present results indicate that expressional changes during affect may involve aspects or judgments of appearance other than those given affective labels per se. Thus, for example, the impact of affective states on learning in the intellectual context may be shaped by a teacher's judgments of a pupil's interest or involvement rather than by judgments of an ongoing mood state. The external, social consequences of internal and presumably private states merit direct investigation.

Conclusion

In summary, the role that affective states may play in intellectual and social behavior is becoming ever more documented. The pre-

sent results indicate that the experimental procedure of inducing affective states is valid and effective and can provide an important investigative tool in exploring the role that affective states may play in shaping cognitive and social aspects of both the normal and disordered behavior of children and adults. Further descriptive investigations are certainly in order, providing both examples of the social and private experiences that are likely to induce affective states and examples of the range of behaviors and capabilities that are sensitive to affective determinants. Also of importance is the further investigation of the nature of affective states, including the elucidation of companion factors, such as tempo, that influence the impact of affective states and the cognitive, attentional, and social mechanisms that mediate observed effects on learning and behavior.

Reference Note

1. Masters, J. C., Ford, M. E., & Arend, R. A. *Strategies for the neutralization or remediation of aversive social experiences*. Paper presented at the biennial meetings of the Society for Research in Child Development, San Francisco, March 1979.

References

Byrne, D., & Clore, G. A reinforcement model of evaluative responses. *Personality*, 1970, *1*, 103–128.

Ekman, P., Friesen, W. V., & Ellsworth, P. *Emotion in the human face: Guidelines for research and integration of findings*. New York: Pergamon, 1971.

Ekman, P., Liebert, R. M., Friesen, W. V., Harrison, R., Zlatchin, C., Malmstrom, E. J., & Baron, R. A. Facial expressions of emotion while watching televised violence as predictors of subsequent aggression. In G. A. Comstock, E. A. Rubenstein, & J. P. Murray (Eds.), *Television and social behavior, reports and papers: Vol. 5. Television's effects: Further explorations*. Rockville, Md.: National Institute of Mental Health, 1972.

Ellis, A. *Reason and emotion in psychotherapy*. New York: Lyle Stuart, 1962.

Ellis, A. *Growth through reason*. Palo Alto, Calif.: Science and Behavior Books, 1971.

Gouaux, C., & Gouaux, S. M. The influence of induced affective states on the effectiveness of social and nonsocial reinforcers in an instrumental learning task. *Psychonomic Science*, 1971, *22*, 341–343.

Harris, M. B., & Siebel, C. E. Affect, aggression, and altruism. *Developmental Psychology*, 1975, *11*, 623–627.

Isen, A. M., Shalker, T. E., Clark, M., & Karp, L. Affect, accessibility of material in memory, and behavior: A cognitive loop? *Journal of Personality and Social Psychology*, 1978, *36*, 1–12.

Izard, C. E. The effects of role-played emotion on affective reaction, intellectual functioning, and evaluative ratings of the actress. *Journal of Clinical Psychology*, 1964, *20*, 444–446.

Izard, C. E., Wehmer, C. M., Livsey, W., & Jennings, J. R. Affect, awareness and performance. In S. S. Tomkins & C. E. Izard (Eds.), *Affect, cognition and personality*. New York: Springer, 1965.

Masters, J. C., & Furman, W. Effects of affective states on noncontingent outcome expectancies and beliefs in internal or external control. *Developmental Psychology*, 1976, *12*, 481–482.

Masters, J. C., Furman, W., & Barden, R. C. Effects of achievement standards, tangible rewards, and self-dispensed achievement evaluation on children's task masters. *Child Development*, 1977, *48*, 217–224.

Masters, J. C., & Santrock, J. W. Studies in the self-regulation of behavior: Effects of contingent cognitive and affective events. *Developmental Psychology*, 1976, *12*, 334–348.

Mischel, W. *Personality and assessment*. New York: Wiley, 1968.

Mischel, W., Ebbesen, E., & Zeiss, A. Selective attention to the self: Situational and dispositional determinants. *Journal of Personality and Social Psychology*, 1973, *27*, 129–142.

Moore, B. S., Underwood, B., & Rosenhan, D. L. Affect and altruism. *Developmental Psychology*, 1973, *8*, 99–104.

Rosenhan, D. L., Underwood, B., & Moore, B. S. Affect moderates self-gratification and altruism. *Journal of Personality and Social Psychology*, 1974, *30*, 546–552.

Rosenthal, R. *Experimenter effects in behavioral research*. New York: Appleton-Century-Crofts, 1966.

Rosenthal, R. Interpersonal expectations: Effects of the experimenter's hypothesis. In R. Rosenthal & R. L. Rosenow (Eds.), *Artifact in behavioral research*. New York: Academic Press, 1969.

Underwood, B., Froming, W. J., & Moore, B. S. Mood, attention, and altruism: A search for mediating variables. *Developmental Psychology*, 1977, *13*, 541–542.

Underwood, B., Moore, B. S., & Rosenhan, D. L. Affect and self-gratification. *Developmental Psychology*, 1973, *8*, 209–214.

Velten, E. A laboratory task for the induction of mood states. *Behaviour Research and Therapy*, 1968, *6*, 473–482.

Woodworth, R., & Schlosberg, H. *Experimental psychology*. New York: Holt, Rinehart & Winston, 1954.

Received April 7, 1978

Journal of Personality and Social Psychology
1979, Vol. 37, No. 3, 391-394

Stimulus Recognition May Mediate Exposure Effects

Michael H. Birnbaum and Barbara A. Mellers
University of Illinois at Urbana-Champaign

Moreland and Zajonc presented stimuli with differential numbers of exposures to subjects and obtained measures of affect (e.g., ratings of liking) and ratings of familiarity. Exposure frequency and ratings of familiarity were both significant predictors of affect in a multiple regression equation. Moreland and Zajonc concluded that there are two independent effects, and thus the exposure effect could not be explained by a stimulus recognition factor alone. However, these results can be explained by the theory that exposure frequency affects a single mediator that is imperfectly correlated with ratings of familiarity and affect. Thus, the null hypothesis that recognition mediates the exposure effect cannot be refuted by the partial correlation and regression analyses of Moreland and Zajonc.

Moreland and Zajonc (1977) varied the frequency with which stimuli were presented and obtained several dependent variables including ratings of familiarity and a measure of liking (affect). Since the rating of affect could be predicted from rated familiarity and exposure frequency and since both regression coefficients were significant, it was argued that there are two "independent" effects. However, the present article shows that the theory that one factor, subjective recognition, mediates the effect of the independent variable on both dependent variables can predict this outcome.

Single-Mediator Theory

Suppose the manipulated variable, exposure frequency, affects a mediator, subjective recognition,

$$S = g(A^*) + s, \tag{1}$$

where S is subjective recognition, A^* is actual exposure frequency, s is a random residual, and g is some function (e.g., logarithmic) of actual frequency.

Suppose that ratings of liking (L) and ratings of familiarity (R) are correlated with subjective recognition,

$$L = aS + l, \tag{2}$$

and

$$R = bS + r, \tag{3}$$

where a and b are linear constants, and l and r are residuals that are uncorrelated with S, s, or each other.

It follows that the intercorrelation between each pair of variables can be expressed as the product of the correlations between the two variables and the mediating factor, subjective recognition:

$$\rho_{ij} = f_i f_j, \quad i \neq j, \tag{4}$$

where ρ_{ij} is the correlation between variables i and j; f_i and f_j are the correlations between the variables and subjective recognition. Since each correlation is the product of the correlations with a mediating factor, S, the path model defined by Equations 1, 2, and 3 is equivalent to a one-factor model.

Interpretation of Partial Correlations in Single-Mediator Theory

The least squares regression coefficient in the (standardized) equation predicting ratings of liking from transformed exposure frequency, $A = g(A^*)$, and rated familiarity is given by

We thank Richard Moreland for providing us with the correlation matrices in Table 1.

Requests for reprints should be addressed to Michael H. Birnbaum, Department of Psychology, University of Illinois, Champaign, Illinois 61820.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3703-0391$00.75

Table 1
Correlations of Variables from Moreland and Zajonc (1977)

	1	2	3	4	5
1. Frequency		.42	.64	.53	.41
2. Affect	.66		.40	.35	.25
3. Familiarity	.58	.53		.47	.54
4. Confidence	.34	.25	.29		.48
5. Accuracy	.41	.30	.55	.42	

Note. Correlations above and below diagonal are for Experiments 1 and 2, respectively. Frequency represents log $(f + 1)$, where f is actual exposure frequency. Liking and affect are corresponding dependent variables for Experiments 1 and 2, respectively.

the expression:

$$\beta_{LR\cdot A} = \frac{\rho_{LR} - \rho_{RA}\rho_{AL}}{1 - \rho_{RA}^2}, \quad (5)$$

where $\beta_{LR\cdot A}$ is the beta coefficient for rated familiarity in the equation predicting rated liking with transformed frequency (A) partialed out. (The corresponding partial correlation has the same numerator as Equation 5.)

Substituting Equation 4 into Equation 5 yields

$$\beta_{LR\cdot A} = \frac{f_L f_R(1 - f_A^2)}{1 - f_R^2 f_A^2} = \frac{\rho_{LR}(1 - f_A^2)}{1 - \rho_{RA}^2}, \quad (6)$$

which shows that unless subjective recognition is perfectly correlated with frequency ($f_A^2 = 1$), the beta is expected to be non-zero and to have the same sign as ρ_{LR}.

Similarly, the beta for A in the equation predicting liking will be given by

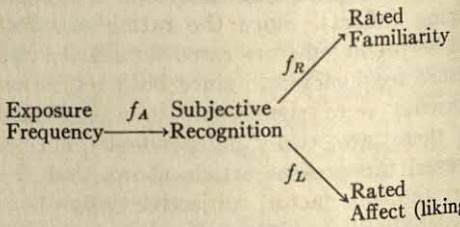
$$\beta_{LA\cdot R} = \frac{\rho_{LA}(1 - f_R^2)}{1 - \rho_{RA}^2}, \quad (7)$$

which shows that the coefficient of frequency is expected to be non-zero and to have the same sign as the correlation between exposure frequency and liking, unless rated familiarity is perfectly correlated with subjective recognition ($f_R^2 = 1$). Thus, the finding of a positive partial correlation between liking and frequency with rated familiarity partialed out is consistent with both the single-mediator (null) hypothesis and the alternative. This partial correlation is therefore not diagnostic of the presence of two independent effects on liking, as contended by Moreland and Zajonc (1977).[1]

One or Two Mediators?

The correlations between exposure frequency, rated familiarity, and affect (liking) for Experiments 1 and 2 of Moreland and Zajonc (1977) are shown in Table 1. The off-diagonal correlations among these three variables can be perfectly fit by the single-mediator model in each case.[2] For Experiment 1, the correlations with the mediator are .82, .78, and .51, respectively; for Experiment 2, the correlations with the recognition mediator are .85, .68, and .78 for frequency, rated familiarity, and affect (liking), respectively.

Consequently, the three correlations are consistent with the representation shown below, in which each correlation among observed variables is the product of their correlations with subject recognition (Equation 4):



. (8)

This model can be extended to describe the intercorrelations among the other dependent variables measuring recognition.

The alternative to Equation 4 is that there is an additional effect of frequency on liking that operates independently of the effect of

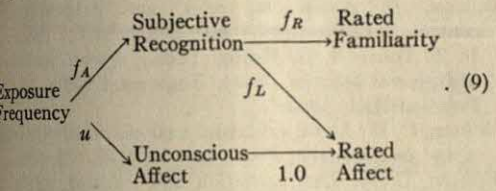
[1] Should a partial correlation have a sign opposite to that of the original correlation (i.e., $\rho_{ij\cdot k}\rho_{ij} < 0$), it would not be consistent with a one-mediator model. Though it may seem counterintuitive, a *negative* partial correlation between exposure frequency and rated familiarity with liking partialed out would have been inconsistent with the single-mediator model, but in agreement with the presence of a second effect of frequency on liking.

[2] For a three-variable, positive intercorrelation matrix, the path coefficients (factor loadings) are given by the following equation:

$$f_i = \sqrt{\frac{\rho_{ij}\rho_{ik}}{\rho_{jk}}},$$

where f_i is the correlation between Variable i and the mediating variable.

frequency on recognition. This theory can be epresented as follows:



. (9)

The single-mediator model (Expression 8, Equation 4) implies that the correlations atisfy the following:

$$\rho_{ij} \leq \frac{\rho_{ik}}{\rho_{jk}} \leq 1/\rho_{ij}, \tag{10}$$

ince the correlations with subjective recog- ition (f_i) are less than or equal to 1.0, and Equation 10 is equivalent to the following:

$$f_i f_j \leq f_i/f_j \leq 1/f_i f_j. \tag{11}$$

For example, the single-mediator model would ot be able to account for correlations for $_{LR}$, ρ_{LA}, and ρ_{RA} of .7, .7, and .35, respec- ively, since $\rho_{LA}/\rho_{RA} = 2$, which is outside he limits set by .7 and 1/.7. However, Equa- ion 10 was satisfied for exposure, affect, and ny of the three recognition measures for both xperiments.

These analyses show that the regression nd partial correlation results reported by Moreland and Zajonc (1977) can be explained y the null hypothesis of a single mediator. Good fit of the null hypothesis (Expression 8) hat recognition mediates the exposure effect oes not rule out the possible presence of nconscious affect (as in Expression 9). How- ver, the burden of proof clearly remains on hose who would conclude, with Moreland nd Zajonc (1977, p. 191), that "the relation- hip between stimulus exposure and affect oes not depend on the operation of higher rder cognitive processes."

Parallel With Subception

The present issue has interesting parallels n perception (Garner, Hake, & Eriksen, 1956) nd learning (Dulany, 1968). Eriksen (1956, 960) replicated and reinterpreted an experi- ent of Lazarus and McCleary (1951) that laimed to demonstrate subception. In Erik- en's (1956) replication, subjects were trained to give a verbal response to each of several squares, one of which was paired with shock. In the testing phase, galvanic skin response (GSR) and verbal responses were measured to each stimulus. It was found that the stimulus–GSR relationship was maintained when verbal response was partialed out. Eriksen (1960) noted that the results are consistent with the hypothesis that the stimulus leads to a perception that is imperfectly correlated with the two dependent variables with uncorrelated errors. To find an *absence* of the partial correlation between stimulus and GSR with verbal response partialed out would require that the verbal response be perfectly correlated with perception. In terms of Thurstone's law of categorical judgment, the category boundaries would have to show zero variance. Should the subject shift the limens during the experiment, the partial correlation will be expected to be non-zero.

Causation and Correlation

Brewer, Campbell, and Crano (1970) have cautioned against the use of multiple regression and partial correlation in nonexperimental settings. They reviewed 10 psychological studies in which partial correlations had been used inappropriately to argue for multiple effects without testing the rival, one-factor hypothesis. This paper shows that the criticisms of partial correlation analysis made by Brewer et al. (1970) can apply to certain experimental as well as nonexperimental studies.

Although factor and path analyses may be useful for examining the number and nature of the *effects* in the dependent variables, they do not allow one to draw inferences about causation. Unless the variables are experimentally manipulated, numerous alternative causal models remain consistent with the data. To argue for *multiple causes*, an investigator must demonstrate that *two* or more experimentally *manipulated* variables contribute independently to the measured dependent variables. One should not draw causal inferences from correlations unless the independent variables are truly independent, that is, manipulated experimental variables.

Conclusions

The present article shows that the one-mediator model may be a reasonable null hypothesis for data obtained in experimental studies in which one variable is manipulated and several are measured. Indeed, a positive partial correlation between exposure frequency and liking with rated familiarity partialed out does *not* evaluate the null hypothesis that a single variable mediates the effect of the independent variable on both dependent variables, as shown by Equation 7. Thus, the theory that stimulus recognition mediates the exposure effect on liking is not refuted by the partial correlation and regression analyses of Moreland and Zajonc (1977).

References

Brewer, M. B., Campbell, D. T., & Crano, W. D. Testing a single-factor model as an alternative to the misuse of partial correlations in hypothesis-testing research. *Sociometry*, 1970, *33*, 1–11.

Dulany, D. Awareness, rules, and propositional control: A confrontation with S-R behavior theory. In T. Dixon & D. Horton (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, N.J.: Prentice–Hall, 1968.

Eriksen, C. W. An experimental analysis of subception. *American Journal of Psychology*, 1956, *69*, 625–634.

Eriksen, C. W. Discrimination and learning without awareness: A methodological survey and evaluation. *Psychological Review*, 1960, *67*, 279–300.

Garner, W. R., Hake, H. W., & Eriksen, C. W. Operationism and the concept of perception. *Psychological Review*, 1956, *63*, 149–159.

Lazarus, R. S., & McCleary, R. A. Autonomic discrimination without awareness: A study of subception. *Psychological Review*, 1951, *88*, 113–122.

Moreland, R. L., & Zajonc, R. B. Is stimulus recognition a necessary condition for the occurrence of exposure effects? *Journal of Personality and Social Psychology*, 1977, *35*, 191–199.

Received April 13, 1978 ▪

Journal of Personality and Social Psychology
1979, Vol. 37, No. 3, 395-412

A Psychological Taxonomy of Trait-Descriptive Terms: The Interpersonal Domain

Jerry S. Wiggins
University of British Columbia, Vancouver, Canada

The eventual aim of the research reported is the development of a comprehensive taxonomy of trait-descriptive terms in the English language. Building on earlier work of Allport, Norman, and Goldberg, preliminary a priori distinctions were made among different domains of trait categories. General procedures for developing structured taxonomies within domains were illustrated with reference to the interpersonal domain. Theoretical considerations dictated the definition of the universe of content, the choice of measurement model, and the procedures for classifying terms within the domain. Eight adjectival scales were developed as markers of the principal vectors of the interpersonal domain. The substantive, structural, and psychometric characteristics of these scales were found to be highly satisfactory. Hence, they may prove useful both as assessment devices in their own right and as reference points for the classification of variables in personality and social psychology.

Personality is that branch of psychology which is concerned with providing a systematic account of the ways in which individuals differ from one another (Wiggins, Renner, Clore, & Rose, 1971). Individuals differ from one another in a variety of ways: their anatomical and physiognomic characteristics; their personal appearance, grooming, and manner of dress; their social backgrounds, roles, and other demographic characteristics; their effect on others or social stimulus value; and at any given moment in time, their temporary states, moods, attitudes, and activities. But the principal goal of personality study is to provide a systematic account of individual differences in human *tendencies* (proclivities, propensities, dispositions, inclinations) to act or not to act in certain ways on certain occasions.

Although such tendencies are commonly called "traits," the use of that term in contemporary psychological discourse carries with it implications of a particular theoretical commitment, a preferred method of scientific investigation, and a philosophical preference for certain kinds of explanation in theory construction. Hence, it is necessary to make it clear at the outset that an interest in human tendencies (traits) does not imply a theoretical precommitment to such issues as whether traits are manifestations of generative or causal mechanisms (Allport, 1937); whether trait attributions reflect specific cognitive processes of observers (Heider, 1958); whether traits are best construed idiographically or nomothetically (Allport, 1937; Bem & Allen, 1974; Kelly, 1955); or whether consistencies in human tendencies are largely due to environmental or situational constancies (Mischel, 1968).

In my view, consistent patterns of human conduct constitute the *basic data of personality study,* which require rather than provide explanation (Wiggins, Note 1). In approach-

The early stages of this research were supported by Grant MH 12972 from the National Institute of Mental Health, U.S. Public Health Service (Lewis R. Goldberg, principal investigator). More recent work was supported by Canada Council Research Grant S75-0109. An earlier version of this article was presented at the meeting of the Society of Multivariate Experimental Psychology, Gleneden Beach, Oregon, November 1975.

Requests for reprints should be sent to Jerry S. Wiggins, Department of Psychology, University of British Columbia, Vancouver, British Columbia, Canada V6T 1W5.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3703-0395$00.75

ing this task of accounting for individual differences in human tendencies, I share with others the conviction that the natural language provides a convenient starting place (Allport, 1937; Cattell, 1957; Goldberg, Note 2; Norman, Note 3). The universe of content of human tendencies is contained within the covers of an unabridged dictionary of the English language. I also share with these writers the conviction that an adequate taxonomy of trait-descriptive terms must precede meaningful empirical studies of human tendencies. There have been a number of systematic efforts to develop a personality taxonomy over the past 40 years, and my own work has attempted to capitalize on these earlier efforts.

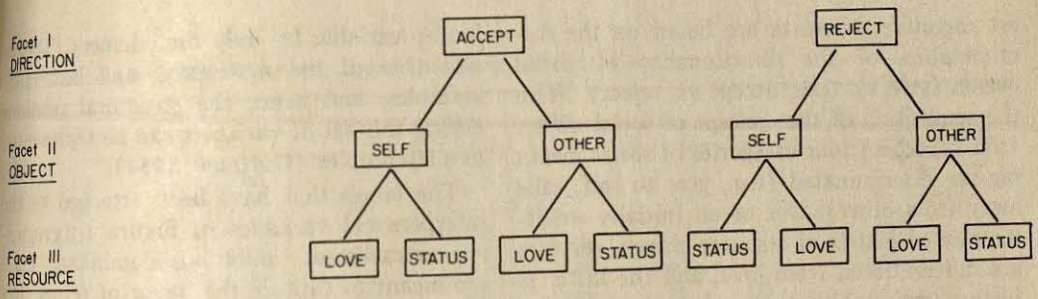
Allport and Odbert (1936) examined the approximately half million separate entries or derivatives included in *Webster's New International Dictionary* (1924) for terms that appeared "to distinguish the behavior of one human being from that of another" (p. 24) and identified a pool of 17,953 terms having this characteristic. Norman (Note 3) scanned the entire contents of *Webster's Third New International Dictionary Unabridged* (1961) for additional terms that had not been included in the Allport–Odbert list. The universe of content thus defined was approximately 27,000 terms. Subsequently, Norman and Goldberg were able to reduce this list by eliminating obscure, inappropriate, and archaic terms. The focus of all of these investigations has been on a subset of approximately 3,600 terms that Allport initially called "stable biophysical traits." Allport considered these to be "real" traits as opposed to temporary states, moods, social roles, physical characteristics, and so forth.

However one conceives of stable biophysical traits, I see them as no more nor less real than any other kinds of human characteristics. I agree with Guilford (1959) that "a trait is any distinguishable, relatively enduring way in which one individual differs from others" (p. 6). But I see the major taxonomic task as that of specifying the different *kinds* of ways in which individuals differ from each other. One *kind* of way in which individuals differ from each other is in terms of what they do to each other. Thus, I think that one of the most important categories of traits is that which may be designated "interpersonal." Within the realm of things that people do to each other, it is desirable to make a further theoretical distinction between interpersonal exchanges based on love and status and interpersonal exchanges based on goods, money, and services (Foa & Foa, 1974). The former are called interpersonal traits and the latter material traits.

Another equally "real" way in which individuals are distinguishable from one another is in terms of their styles of emotional reactivity, which we refer to as "temperament." It is also possible to distinguish individuals on the basis of the particular roles and status they hold within the framework of our social institutions. Additionally, there are "character" terms that represent appraisals of an individual based on a code of proper behavior. And finally, there are many words that refer to individual differences in qualities of mind as manifested in thought, perception, and speech.

Consider the following trait-descriptive adjectives: *aggressive, miserly, lively, ceremonious, dishonest,* and *analytical.* I would classify these terms as representative of the six trait categories just mentioned, namely, interpersonal traits, material traits, temperamental traits, social roles, character, and mental predicates. Here I differ from Allport, Norman, and Goldberg, who view all of these terms as "stable biophysical traits," without attempting to differentiate the different kinds of descriptive jobs such terms perform for us. To these earlier authors' important distinction between stable traits and temporary states, moods, social evaluations, and so forth, I would add finer distinctions *within* the category of stable traits.

The present article describes our first efforts to develop a taxonomy of personality trait-descriptive terms in the English language. The universe of content is taken to be Norman's (Note 3) total lexicon of 18,125 terms. Within this broader framework, I intend to focus on the "prime" categories of Norman, which involve 4,063 relatively familiar and nonobscure terms. Even restricting our attention to prime terms, we are still



INTERPERSONAL VARIABLES								
NO (Gregarious - Extraverted)	+	+	+	+				−
PA (Ambitious - Dominant)	+	+	+				−	−
BC (Arrogant - Calculating)	+	+				−	−	−
DE (Cold - Quarrelsome)		+			−	−	−	−
FG (Aloof - Introverted)				+	−	−	−	
HI (Lazy - Submissive)			+	+	−	−		
JK (Unassuming - Ingenuous)			+	+	−			
LM (Warm - Agreeable)	+		+					

Figure 1. Facet composition of interpersonal variables (after Foa & Foa, 1974). (+ = acceptance; − = rejection.)

confronted with a list of staggering size. Our enterprise required a clear-cut game plan or research strategy to guide us through this sea of words.

We began by defining a limited taxonomy on an a priori basis. We attempted to specify different *domains* of human characteristics based on the different descriptive jobs performed by the words within the domains. Next, we attempted preliminary a priori classifications within domains, mainly as a means of keeping track of terms and lightening our clerical burdens. The overall strategy for developing a taxonomy within a single domain may be honorifically described as "iterative," although "trial and error" may be a more apt descriptor. Our main concern was the avoidance of premature "fixing" of a taxonomy. Hence, all preliminary classifications were tentative in nature and subject to continual revision in the light of other developing categories. This rather complex interactive strategy will be illustrated with reference to the development of a taxonomy within the domain of interpersonal traits.

The Domain of Interpersonal Traits

The line of reasoning to be presented originated in the interpersonal theory of Sullivan (1953) and was later operationalized in a set of personality measurement procedures by Leary (1957) and his colleagues. Foa (1961, 1965) integrated this theory with Guttman's (1954) order analysis, and the general orientation was extended to domains other than the interpersonal by Rinn (1965). Notable interpersonal systems based on a circumplex model have also been described by Becker and Krug (1964), Benjamin (1974), Lorr and McNair (1963), Schaefer (1959), and Stern (1970). This general framework was integrated with interpersonal exchange theory (Homans, 1961) by Carson (1969) and by Foa and Foa (1974).

In their most recent theoretical statement, Foa and Foa (1974) describe the development of cognitive categories of social perception as a progressive differentiation of structure involving facets of *directionality*, *object*, and *resource*. A somewhat simplified version of Foa and Foa's representation of this structure is illustrated at the top of Figure 1. The earli-

est cognitive schemata are based on the discrimination of the directionality of social events (give vs. take, accept vs. reject). With the acquisition of the concept of social object (self vs. other) four categories of social meaning are discriminated (e.g., give to self, take away from other). Out of an initially undifferentiated matrix of resource classes, services are differentiated from love, and the latter is further differentiated into love and status. With the distinction between love and status, earlier facets become reorganized, and it is possible to distinguish eight features of social meaning (e.g., granting status to oneself, denying love to another). These eight features may be thought of as part of a semantic code strip that provides the basic discriminations for encoding and decoding interpersonal events (Osgood, 1970).

Within the above context, interpersonal events may be defined as *dyadic interactions that have relatively clear-cut social (status) and emotional (love) consequences for both participants (self and other)*. This definition provides a theoretical basis for distinguishing interpersonal traits from other categories of trait descriptors, such as temperament, moods, cognitive traits, and physical characteristics.

Under the semantic features in Figure 1 are listed eight theoretical interpersonal *variables*. The organization of these variables is thought to be determined by an interrelated set of societal rules that impart meaning to social events (Wiggins, Note 1). Thus, actions that have the same profile of semantic features are categorized as belonging to the same response class. The semantic features of each of the eight variables appear as rows in Figure 1. Note that the first variable (NO) is coded on all the positive (accept) categories for self and other with respect to both love and status. The first variable is in marked contrast to FG for which all values are negative (reject). Variables NO and FG have no features in common, but since acceptance and rejection are logically opposed concepts, it would be expected that the two variables would be strongly negatively correlated. Similar relationships can be seen to exist between PA and HI, BC and JK, and DE and LM. Note also that each variable differs from its adjacent variable by only one element. This is also true of the first (NO) and last (LM) variables, and hence the structural relations among this set of variables can be represented as a circumplex (Guttman, 1954).

The labels that have been attached to the interpersonal variables in Figure 1 (gregarious–extraverted, ambitious–dominant, etc.) are meant to capture the flavor of terms that share the same profile of semantic features and may serve more as tags than as definitions. Thus, for example, the rather inelegant label of "lazy–submissive" is attached to interpersonal transactions involving incompetence, passive resistance, submission, or obedience. These otherwise diverse attributes share in common the semantic features of denying status to self, denying love to both self and other, and granting status to other. A fuller listing of representative terms in each category may be found in Table 2, later in this article.

Were we to collect personality measurements on the eight variables listed in Figure 1, their intercorrelations would, in theory, show the pattern illustrated in Table 1. The correlation of a variable with itself is assumed to be unity, so that the principal diagonal of this matrix contains ones. The correlations along this main diagonal are large and positive, and they decrease across successive minor diagonals to the $n/2$nd variable, where they are a minimum. The correlations then increase up to a large positive value in the off-diagonal matrix. The circumplexity of this or any other set of variables can be evaluated directly from the intercorrelation matrix. A rigorous procedure for assessing the circumplexity of a set of variables involves plotting the correlations of each variable (ordinate) with the other variables in sequence (abscissa). This procedure should generate a series of overlapping sine curves that can be evaluated for goodness of fit (Stern, 1970).

An alternative procedure for evaluating circumplexity is to extract the first two principal components from the matrix of intercorrelations and to examine the plot of the variables on the two components. Figure 2 presents the plot of the eight variables on two principal

Table 1
Hypothetical Correlations Among Eight Interpersonal Variables

Variable	PA	BC	DE	FG	HI	JK	LM	NO
PA	1.00							
BC	.50	1.00						
DE	.00	.50	1.00					
FG	−.50	.00	.50	1.00				
HI	−1.00	−.50	.00	.50	1.00			
JK	−.50	−1.00	−.50	.00	.50	1.00		
LM	.00	−.50	−1.00	−.50	.00	.50	1.00	
NO	.50	.00	−.50	−1.00	−.50	.00	.50	1.00

Note. PA = ambitious-dominant; BC = arrogant-calculating; DE = cold-quarrelsome; FG = aloof-introverted; HI = lazy-submissive; JK = unassuming-ingenuous; LM = warm-agreeable; NO = gregarious-extraverted.

components extracted from the intercorrelations in Table 1. As can be seen from this figure, the intercorrelations among variables in Table 1 form a perfect, equally spaced circumplex. This pattern follows, both theoretically and empirically, from the pattern of shared semantic features among the variables illustrated at the bottom of Figure 1. Variables that share three features in common are adjacent to each other on the circle; variables that have no features in common are opposite each other.

There are two distinct advantages to the representation of interpersonal variables by a two-dimensional circumplex. The first is that it provides an explicit conceptual definition of the universe of content of interpersonal behavior. Any behavior that meets the definition of a meaningful interpersonal event given above must be capable of being represented as a vector originating from the center of the circle. Thus, the specific system proposed here is potentially falsifiable. The second advantage of the circumplex model is that it

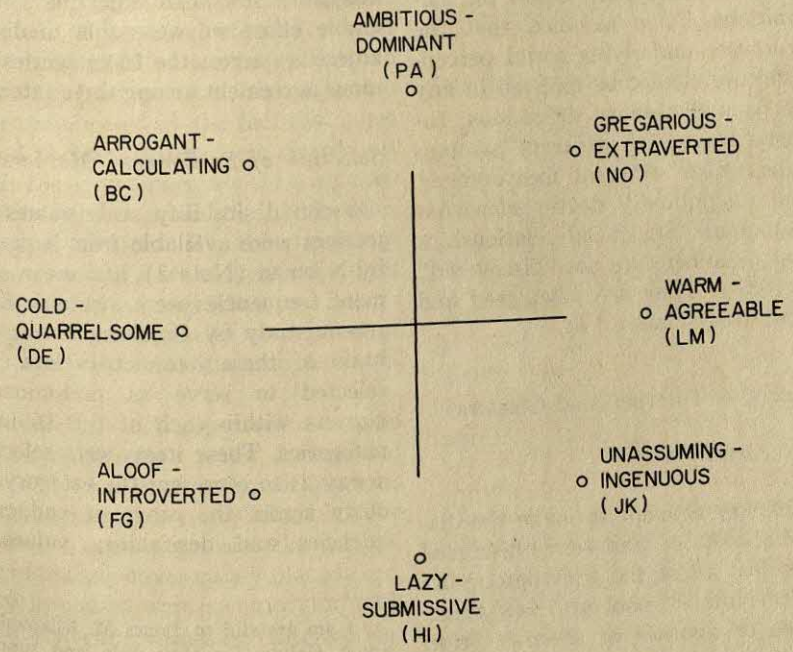


Figure 2. Perfect circumplex of interpersonal variables.

alerts the investigator to noticeable "gaps" in the interpersonal space of a given set of variables. Strictly empirical procedures of variable selection are likely to deemphasize the importance of certain variables that are implied by the logic of the circumplex system but that are underrepresented in the English language. Although the system of interpersonal variables under discussion is limited to eight variables, it could in principle be equally well represented by 16, 32, or 64 variables. The thinness with which we slice the circumplex pie is limited by the reliability with which respondents can distinguish between closely synonymous words or phrases.

It should be evident that the preceding theoretical considerations substantially constrain the final form that a taxonomy of interpersonal traits may assume. The definition of interpersonal events, the specification of their underlying facet structure, and the selection of a measurement model to represent relationships among variables all express a commitment to a particular, albeit widely shared, conceptualization of the domain under investigation. In this sense, the eventual taxonomy of trait-descriptive terms will be a "psychological" taxonomy rather than a strictly "semantic" taxonomy based on dictionary definitions. It is assumed that the semantic structures underlying social perception in this culture cannot be inferred in any obvious way from dictionary definitions. Investigators who start with different assumptions and who utilize different measurement models would undoubtedly devise somewhat different taxonomies. Structural relationships of the kind at issue here are not "discovered" (Loevinger, 1957). They are postulated and then evaluated for goodness of fit.

Development of Interpersonal Clusters

Rational Categorization

In a preliminary attempt to make the distinction among *kinds* of trait terms discussed above, I classified all of the trait descriptors in Goldberg's (Note 2) pool of 1,710 adjectives into one or another of seven a priori categories. At that stage in the development of the taxonomy, the principal distinctions were between interpersonal trait terms, temperamental trait terms, and mental predicate terms. Tentative categories of "attitudinal terms" and "constancy" were employed at this point, as well as a "miscellaneous" category that involved various dimensions that could not be clearly classified into other categories. In the initial sorting of the 1,710 adjectives, approximately 800 terms were identified as "interpersonal" under our working definition of interpersonal traits.

For purposes of preliminary categorization, we selected Leary's (1957) system of interpersonal traits, because it appeared to be the most explicit system described in the literature, an entire book being devoted to the topic. Two colleagues [1] and I thoroughly familiarized ourselves with the system and its theoretical framework. Working as a team, we considered each of the approximately 800 terms previously classified as interpersonal and attempted to classify each within one of the 16 interpersonal vectors. Adjectives that could not be classified within the 16 dimensions and that did not appear to belong in one of the other five a priori categories were temporarily set aside for further analysis. There were remarkably few such adjectives. With considerable effort we were able to distribute 567 adjectives across the 16 categories with unanimous agreement among three raters.

Selection of Preliminary Markers

Social desirability scale values for all adjectives were available from a previous study by Norman (Note 2), and mean self-endorsement frequencies were available from a more recent study by Goldberg (Note 2). On the basis of these itemmetric data, items were selected to serve as preliminary nuclear clusters within each of the 16 interpersonal categories. These items were selected in such a way as to represent the category unambiguously across the range of endorsement frequencies and desirability values. Approxi-

[1] I am grateful to James M. Kilkowski and Alexander Galvin for their help and support in this enterprise.

mately 12 items were selected for each category.

In Goldberg's sample of 70 male and 117 female University of Oregon undergraduates, self-ratings were obtained on a 9-place scale for each of the 1,710 trait descriptors. Using Goldberg's data, our next step was to obtain the intercorrelations among terms selected to be preliminary markers of each of the categories of interpersonal behavior. By a clustering procedure, a smaller and more homogeneous subset of items was selected within each category, with approximately six items per cluster.

The next step was to consider all items that had been rationally classified as falling within the 16 categories as potential candidates for addition to or deletion from each nuclear interpersonal cluster. For this purpose, we scored the preliminary clusters as 16 scales and examined the correlations of individual items with these scales with an eye toward their circumplex patterning. Thus, we examined the correlation of each of the 567 items with the 16 preliminary clusters ordered by their hypothetical position in the circumplex. Items were sought that had positive correlations with adjacent clusters, zero correlations with orthogonal clusters, and negative correlations with opposite clusters. For most items, this circumplex patterning across 16 clusters was far from perfect. Item selection was further complicated by the fact that some of the initial nuclear clusters were clearly out of place on the circumplex, so that we constantly had to keep in mind the inadequacies of our original clusters. By this bootstrap procedure, we selected a set of eight adjectives for each of the 16 vectors that, we hoped, would increase the homogeneity of their vector and correct the previous shortcomings of that vector.

We assembled the 128 adjectives chosen to mark the 16 vectors into a test format and administered it to a small group of students from the University of British Columbia. The students were requested to rate the self-applicability of the adjectives on a 9-place Likert scale. Item responses were summed for both octant and sixteenth variables. Intercorrelations among the variables were factored by the method of principal components. Figure 3 displays the pattern of octant variables loading on the first two factors in the University of British Columbia sample. The pattern of loadings for 16 interpersonal variables is considerably less orderly than this, due no doubt to the fact that the vector scores are based on fewer items in a relatively small sample of subjects. But the pattern of octant scores clearly displays some of the difficulties we encountered in our attempt to map out Leary's (1957) system.

The most striking feature of Figure 3 is the lack of variables in the upper right-hand quadrant. This gap in the Leary (1957) circumplex has been previously noted by several authors including Stern (1970), who felt that the failure of Octants PA and NO to "close" raises the question of whether the Leary system is a circumplex. Lorr and McNair (1965) noted this gap also and attempted to close it with additional substantive variables. Another notable departure from expectations is the location of Octant NO, which not only fails to appear in the upper quadrant but appears *below* Octant LM. In a sense, then, we partially succeeded in replicating the Leary system with trait-descriptive adjectives, but in so doing we carried over the faults of the system as well.[2]

Development of Bipolar Clusters

At this point it occurred to us that a 16-variable circumplex can be rather easily constructed from eight genuine *bipolar* dimensions. One of the conceptual difficulties we experienced in working with the Leary system was a decided lack of bipolarity between vectors that appeared opposite each other on the circle. In particular, the following implicitly bipolar contrasts did not make a great deal of sense: success versus masochism, narcissism versus conformity, rebellion versus

---

[2] Juris I. Berzins (personal communication, March 8, 1977) plotted the loadings of the original Leary (1957) Interpersonal Checklist (ICL) octants on the first two principal components in samples of 685 high school students and 1,109 college students. The plots for both samples are indistinguishable from Figure 3. (See also Rinn, 1965, p. 458, Figure 6.)

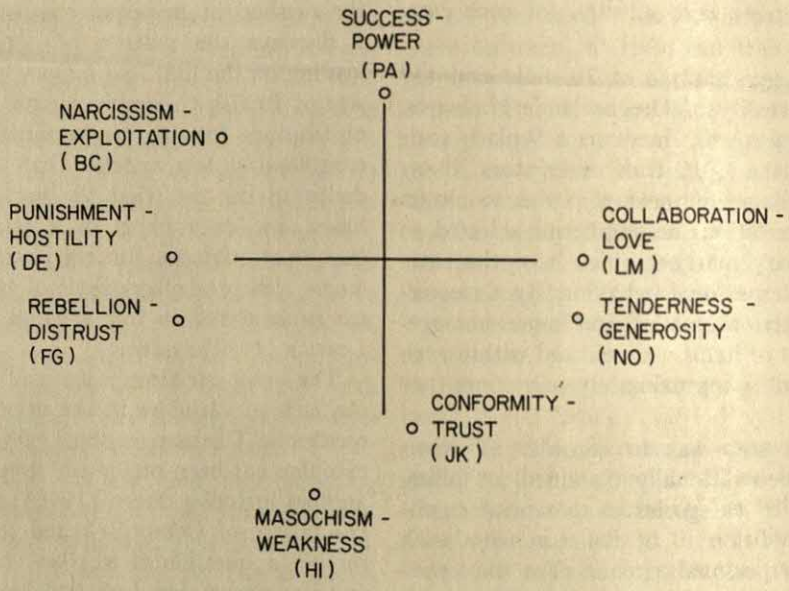


Figure 3. Rotated components of Leary (1957) interpersonal variables.

tenderness, distrust versus generosity, and punishment versus collaboration.

In Leary's (1957) original system the tenderness–generosity octant (NO) was conceptualized as the bipolar contrast to the rebellious–distrustful octant (FG). This likely accounts for the noticeable gap in the upper right quadrant of the system (Figure 3). Tenderness and generosity are not bipolar contrasts to rebellious and distrustful behavior. Tenderness and generosity are too weak and loving to be placed this high on the circle. As Lorr and McNair (1965) noted earlier, the NO octant, which falls between dominance and love, reflects a socially exhibitionistic style of behaving (gregarious–extraverted).

With a little effort, our revised representation of the 16 interpersonal variables can be read from Figure 2. Each variable represents a bipolar contrast to the variable appearing opposite it on the circle. The contrasts are dominant–submissive, arrogant–unassuming, calculating–ingenuous, cold–warm, quarrelsome–agreeable, aloof–gregarious, introverted–extraverted, and ambitious–lazy.

In developing our bipolar clusters, we attempted to select items that were highly negatively correlated with their opposite cluster and that had zero correlations with their theoretically orthogonal clusters. Thus, a good dominant item should have a high positive correlation with a dominant cluster, a high negative correlation with a submissive cluster, and essentially zero correlations with quarrelsome and agreeable clusters. We used a tentative set of 16 four-item bipolar clusters as markers to select items having the properties just described. This enabled us to revise and expand our tentative clusters into eight-item variables. Factor analysis of these eight-item variables in Goldberg's (Note 2) sample of 187 subjects revealed the clearest circumplex structure we had yet encountered in our own work or in the literature. This result was replicated in an additional sample of 119 subjects from the University of British Columbia.

Development of Final Scales

The items in our eight-item bipolar interpersonal clusters were developed from the pool of 567 adjectives classified as interpersonal in our initial rational sorting of the total list of 1,710 adjectives. The possibility clearly existed that some adjectives from among the 1,143 not considered would be appropriately placed within the 16 interper-

sonal categories. Consequently, a program was written to examine the relationships between all 1,710 adjectives and the 16 bipolar interpersonal clusters. The program correlates each adjective with the 16 bipolar clusters, selects the cluster with which the adjective is most highly correlated, and then prints out an ordered list of variables within each of the 16 categories. This listing provided the correlation of each item with the cluster with which it was most highly correlated. It also provided the correlation with the bipolar opposite cluster and with the two orthogonal clusters. This procedure produced 16 lists that located all 1,710 adjectives with respect to the interpersonal cluster with which they were most highly correlated.

Two of us then went through these lists and examined the correlation of each item with the cluster from the category to which it had been assigned, as well as the correlation of the item with its opposite and orthogonal clusters.[3] Items that clearly belonged in the interpersonal category to which they had been assigned were retained. Items that did not appear to belong in the interpersonal category to which they were assigned were placed in one of eight other taxonomic categories that then existed (e.g., temperament, character, material traits, etc.).

At this point, an attempt was made to develop more refined subcategories within the broader taxa. Preliminary groupings of clusters were formed within the domains of temperament, character, attitudes, mental predicates, and social roles. This classification added 38 new clusters to the 16 interpersonal clusters, making a total of 54 categories within the preliminary taxonomy. These 54 categories were scored as scales by computing, for each subject, the sum of his or her ratings to the items in that category. In the case of the 16 interpersonal categories, the 25 best items, as determined from item–cluster correlations, form the reference scale for each category.

The next stage of analysis was designed to determine more exactly the classification of doubtful adjectives in the preceding taxonomic sort. This included adjectives that were interpersonal but that seemed to be in the wrong interpersonal category, adjectives that were unclassifiable in any category, and adjectives that were candidates for deletion on the grounds of ambiguity, obscurity, or lack of personological relevance. In addition, all adjectives in the interpersonal category other than the 400 (16×25) marker adjectives were classified as doubtful for the purpose of this analysis. There were 768 such adjectives designated as questionable by this criterion. Each of the 768 adjectives was correlated with all of the 54 taxonomic category scales. Two of us then examined this pattern of correlations for all 768 words and assigned them to one of the 54 categories on the basis of both conceptual and empirical (correlational) considerations. Some items were retained in their original interpersonal category, others moved to other categories, and some deleted from further consideration. By these procedures, a revised 54-category taxonomy was developed. Within this revised taxonomy, 864 adjectives were classified as interpersonal.

The selection of items for inclusion in the final interpersonal trait scales involved one more cycle in our iterative procedures The 25 "best" items previously identified in each of the 16 interpersonal categories were scored as reference clusters. These reference clusters provided a broader representation of the content of the final interpersonal variables than did our earlier clusters. Within each category, each item was correlated with its own 25-item cluster and with its 25-item opposite and orthogonal clusters. The items were then ordered within each category by their correlations with their own clusters.

Working independently, three of us selected the "best" eight items in each of the 16 interpersonal categories. One rater made selections strictly on the basis of a summary numerical index based on item correlations with same, opposite, and orthogonal clusters, a procedure referred to as *empirical*. Another rater attended to the same item correlations, but attempted to choose the more "meaningful" adjectives from pairs that had roughly similar empirical characteristics, a procedure

---

[3] I would like to acknowledge the substantial help of Ana Holzmuller in this and many other phases of the project.

designated *quasi-empirical*. The third rater selected sets of eight adjectives that were "psychologically cohesive" in terms of his conception of the constructs under investigation, the *rational* procedure.

The empirical and quasi-empirical item-selection procedures yielded sets of 16 scales with highly similar psychometric properties. For a specific sixteenth, it was possible to identify one or the other of the two scales as being preferable, either on grounds of a higher alpha coefficient or on the grounds of a better position in the 16-variable circumplex plot. Consequently, a *combination* set of 8 variables was formed from the best of the empirical and quasiempirical sixteenths.

An additional procedure for item selection was based on a more fine-grained analysis of the circumplex properties of individual *items*. Each item was correlated with the 16 25-item clusters that served as markers of the interpersonal variables. This *pattern* of correlations was then inspected for the number of departures from perfect circumplex ordering that occurred. Items with the smallest number of departures from this pattern were retained in a set of scales labeled "item ordering."

The final selection of a 128-item set of interpersonal variables was based on an index of circumplexity that permits comparison among competing scale sets (Wiggins & Marston, Note 4). Very briefly, two factors were extracted from the intercorrelations among interpersonal variables in a given scale set. A hypothetical factor matrix was constructed to represent a perfect circumplex solution for variables of this level of reliability (as estimated from communalities). The sum of the absolute differences between elements in the hypothetical factor matrix and elements in the obtained factor matrix was taken as an index of circumplexity. Each of the five sets of interpersonal scales was then compared on this index. Both octants and sixteenths were evaluated in Goldberg's (Note 2) sample of 187 University of Oregon undergraduates and in Norman's sample of 123 University of Western Australia undergraduates. An overall index of circumplexity, based on both 8- and 16-variable solutions in American and Australian samples, suggested that the *combination* set of 128 adjectives had a slight edge over the item-ordering, empirical, and quasi-empirical procedures and a substantial advantage over the rational procedure, which fared rather badly. This combination set of adjectives, which has been employed in all subsequent investigations, is listed in Table 2.

Generalizability of Circumplex Structure

The present taxonomy of the interpersonal domain is based on an explicit structural model (Guttman, 1954) that follows from a facet analysis of cognitive categories of social perception (Foa & Foa, 1974). On the basis of both theoretical and psychometric considerations, a set of eight 16-item scales were developed as marker variables of the principal vectors of this system. In addition to being potentially useful personality assessment measures in their own right, these scales enable one to classify any interpersonal trait descriptor by establishing its location within the circumplex space. However, the meaningfulness of such classification depends heavily on the generalizability of the present circumplex structure to samples other than those involved in the derivation of the eight interpersonal scales. The scales were derived primarily from Goldberg's (Note 2) sample of American students and cross-validated at various stages in samples of Canadian students and in Norman's sample of Australian students. Although relatively diverse samples of students were employed in scale derivation, evidence for the generalizability of the circumplex structure must come from other sources. Four samples of subjects were tested for this purpose.

Cross-Validation Samples

Sample A. Subjects were recruited by advertisements to participate in a "personality study" that involved two separate testing sessions of approximately 2 hours each (Marston, Note 5). Subjects were paid $10 each for their participation and were provided individual feedback if requested. The 128 interpersonal adjectives were embedded in a

Table 2
Final Set of Interpersonal Adjective Scales

P. **Ambitious** (Success)	A. **Dominant** (Power)	B. **Arrogant** (Narcissism)	C. **Calculating** (Exploitation)
Persevering	Dominant	Bigheaded	Sly
Persistent	Assertive	Boisterous	Tricky
Industrious	Forceful	Conceited	Wily
Self-disciplined	Domineering	Boastful	Cunning
Organized	Firm	Overforward	Overcunning
Deliberative	Self-confident	Swellheaded	Crafty
Stable	Self-assured	Cocky	Calculating
Steady	Un-self-conscious	Flaunty	Exploitative
D. **Cold** (Hate)	E. **Quarrelsome** (Hostility)	F. **Aloof** (Disaffiliation)	G. **Introverted** (Withdrawal)
Warmthless	Impolite	Antisocial	Silent
Unsympathetic	Uncordial	Unneighborly	Shy
Ironhearted	Discourteous	Impersonal	Introverted
Uncharitable	Ungracious	Unsociable	Bashful
Coldhearted	Disrespectful	Distant	Inward
Hardhearted	Uncooperative	Dissocial	Unrevealing
Cruel	Ill-mannered	Unsmiling	Unsparkling
Ruthless	Uncivil	Uncheery	Undemonstrative
H. **Lazy** (Failure)	I. **Submissive** (Weakness)	J. **Unassuming** (Modesty)	K. **Ingenuous** (Trust)
Unproductive	Self-doubting	Nonegotistical	Uncunning
Lazy	Self-effacing	Undemanding	Uncalculating
Unthorough	Timid	Unvain	Uncrafty
Unindustrious	Meek	Unwild	Unwily
Inconsistent	Unbold	Unargumentative	Unsly
Disorganized	Unaggressive	Boastless	Guileless
Unbusinesslike	Forceless	Pretenseless	Undevious
Impractical	Unauthoritative	Conceitless	Undeceptive
L. **Warm** (Love)	M. **Agreeable** (Collaboration)	N. **Gregarious** (Affiliation)	O. **Extraverted** (Outgoingness)
Tenderhearted	Courteous	Friendly	Outgoing
Gentlehearted	Charitable	Genial	Extraverted
Tender	Well-mannered	Neighborly	Vivacious
Kind	Respectful	Companionable	Jovial
Emotional	Cordial	Approachable	Enthusiastic
Sympathetic	Cooperative	Congenial	Cheerful
Softhearted	Accommodating	Good-natured	Perky
Appreciative	Forgiving	Pleasant	Unshy

list of 605 adjectives whose order was randomized separately for each subject. Four personality inventories were administered in varying orders along with the total list of adjectives. The 152 subjects (51 men and 101 women) included both undergraduate and graduate students from a variety of academic majors at the University of British Columbia.

Sample B. Representative samples of volunteer and professional workers were recruited from three different social service agencies in the Greater Vancouver area (Merritt, Note 6). The 128 interpersonal adjectives were administered in a single booklet along with a value survey and a standardized personality inventory. Testing was accomplished individually, in small groups, or on a take-home basis. One hundred subjects (29 men and 71 women) completed the interpersonal adjective form.

Sample C. Students in the second term of an introductory psychology class at the Uni-

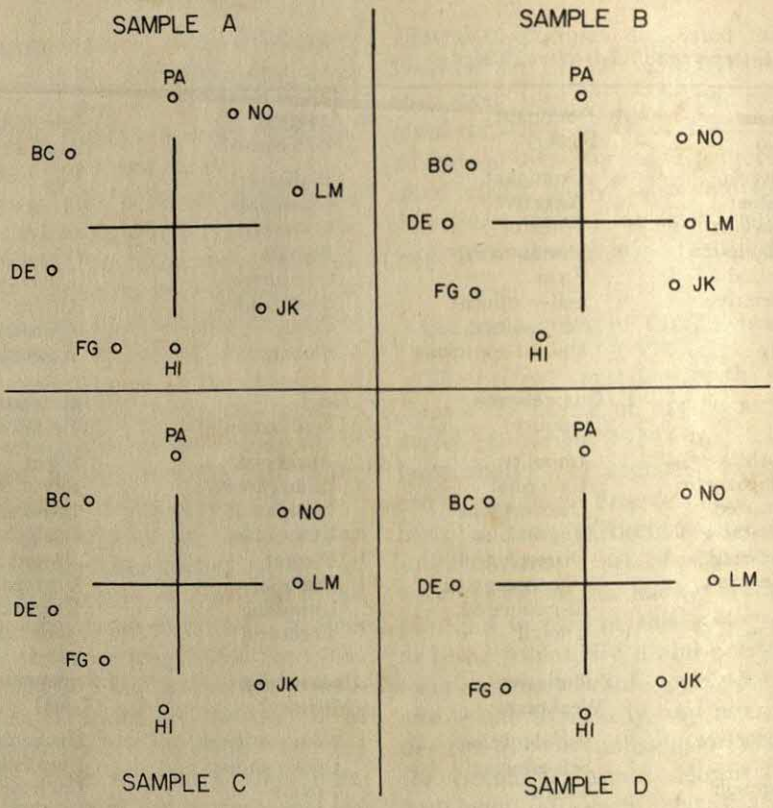


Figure 4. Structure of final interpersonal variables in four samples. (PA = ambitious-dominant; BC = arrogant-calculating; DE = cold-quarrelsome; FG = aloof-introverted; HI = lazy-submissive; JK = unassuming-ingenuous; LM = warm-agreeable.)

versity of British Columbia were administered the 128 adjectives, embedded in a larger list of adjectives, during a regular class period. The students were given a guest lecture on personality testing in exchange for their participation. Data were collected for 132 subjects (57 men and 75 women).

Sample D. Summer students enrolled in an introductory psychology class at the University of British Columbia were administered the 128 adjectives, embedded in a larger list, during a regular class period. The students later received their own standardized scores on the interpersonal variables and a lecture on interpersonal behavior. Data were collected for 139 subjects (58 men and 81 women).

Circumplex Analyses

Intercorrelations were obtained among the eight interpersonal adjective scales in each of the four samples just described. Two principal components were then extracted from each of the intercorrelation matrices. In the respective analyses, these two components accounted for 76.1% (Sample A), 67.9% (Sample B, 72.0% (Sample C), and 74.8% (Sample D) of the total variance. For comparability of inspection, the second component of each solution was hand rotated to pass through the ambitious–dominant vector (PA). Figure 4 depicts the loadings of the eight interpersonal scales on two hand-rotated principal components in each of four samples.

Perfect, evenly spaced circumplexes (Figure 2) are not expected in real data, because of measurement error, and hence the structures in Figure 4 are more properly described as quasi-circumplexes (Guttman, 1954). Although the amount of "quasiness" one is prepared to tolerate in empirical data is to some extent a matter of taste, it seems fair to say that the quasi-circumplex structures in

Table 3
Psychometric Characteristics of Interpersonal Adjective Scales

Scale label	Scale M and SD[a]						Social desirability rating (N = 100)		Internal consistency (coefficient alpha)	
	Men (n = 236)		Women (n = 374)		Total (N = 610)				Total sample (N = 610)	Range of values from the 4 subsamples
	M	SD	M	SD	M	SD	M	SD		
PA	5.87	.95	5.72	1.01	5.79	.99	6.77	1.10	.855	.831–.880
BC	4.30	1.02	3.60	1.07	3.87	1.05	3.66	.98	.870	.845–.885
DE	2.95	.89	2.48	.80	2.66	.84	2.31	.81	.877	.862–.889
FG	4.28	1.19	3.87	1.17	4.03	1.17	3.54	.71	.891	.887–.894
HI	3.95	.94	4.13	1.01	4.06	.99	3.37	.61	.809	.769–.838
JK	4.64	.93	5.14	1.00	4.95	.97	5.73	.95	.801	.743–.840
LM	6.64	.80	7.08	.76	6.91	.77	7.53	.73	.865	.841–.879
NO	6.17	.96	6.54	.96	6.40	.96	7.52	.66	.897	.887–.917

Note. Scale labels stand for ambitious-dominant (PA); arrogant-calculating (BC); cold-quarrelsome (DE); aloof-introverted (FG); lazy-submissive (HI); unassuming-ingenuous (JK); warm-agreeable (LM); and gregarious-extraverted (NO).
[a] Scale values range from 1 (extremely inaccurate description) to 9 (extremely accurate description).

Figure 4 are among the clearest reported in the personality literature to date. The circumplex ordering obtained in scale derivation samples is generalizable across relatively diverse samples of subjects tested in differing contexts. Hence, the eight scales may prove useful both as measuring instruments and as reference points for the classification of items and scales in the interpersonal domain.

Psychometric Characteristics

Sex Differences

Table 3 presents normative and psychometric data for each of the eight interpersonal adjective scales. The first six columns contain means and standard deviations in a combined sample of 610 North American university students. Inspection of the separate means for men and women reveals clear-cut, and to some extent predictable, sex differences in self-report. In comparison with women, men present themselves as more ambitious–dominant, arrogant–calculating, cold–quarrelsome, and aloof–introverted. Women present themselves as more lazy–submissive, unassuming–ingenuous, warm–agreeable, and gregarious–extraverted. All of these mean differences are statistically reliable. Surprisingly, the smallest differences occur on ambitious–dominant ($p < .03$) and lazy–submissive ($p < .01$); all other differences are highly reliable ($p < .0001$).

Sex differences in self-report on the interpersonal circumplex provide a graphic representation of sex role stereotypes in North American society (Wiggins & Holzmuller, 1978). These differences are sample dependent and may possibly decrease in future years. Samples of university students are not uniformly "stereotyped." They contain differing proportions of complexly "sex-reversed" subjects, such as men who score high on lazy–submissive and on aloof–introverted and women who score high on ambitious–dominant and on gregarious–extraverted (Wiggins & Holzmuller, Note 7).

Social Desirability

Social desirability ratings of the individual adjectives were obtained from Norman's monograph (Note 3). One hundred University of Michigan undergraduates (50 men and 50 women) rated the adjectives on a 9-place social desirability rating scale. The means and standard deviations for the 16 adjectives in each scale appear in columns 7 and 8 of Table 3. Comparing the total-sample mean self-report scale scores with corresponding mean social desirability ratings, it is clear that the well-known relationship between endorsement and desirability is found within the interpersonal domain (Edwards, 1957a). Moreover, the similarity between mean endorsement and mean desirability with respect to the *patterning* or ordering of interpersonal

variables is open to alternative interpretations of the circumplex structure in terms of stylistic response variables.[4]

A variety of procedures have been suggested for coping with the endorsement–desirability confound in self-report personality data. On the assumption that in principle, responses within all interpersonal vectors should be equiprobable, LaForge and Suczek (1955) tried to write desirable phrases for undesirable dimensions ("Can be strict if necessary") and undesirable phrases for desirable dimensions ("Spoils people with kindness"). Jackson's (1970) differential reliability index removes desirability variance and enhances content saturation at the stage of item selection. Norman (Note 9) has described analysis of variance procedures for removing desirability variance from the Subject × Item response matrix. Finally, several individual difference measures of social desirability response style have been constructed to permit the assessment of desirability responding for each subject (e.g., Crowne & Marlowe, 1960; Edwards, 1957b; Wiggins, 1959).

The extent to which an investigator may feel the need to "correct for" the ubiquitous endorsement–desirability confound will vary with the purpose of the investigation and with the investigator's theoretical stance on the meaning of the evaluative dimension of affective meaning. A set of interpersonal variables that did not differ in desirability (or did not reveal sex differences) would be a feeble representation of real-life categories of social perception. On the other hand, basic investigations of the process of responding to personality inventory items may require variance decomposition procedures such as those of Norman (Note 9). In any event, interpretations of individual or group scores on the interpersonal scales should be made with reference to *normative data* of the kind provided in Table 3. Whether a person or a group scores high on warm–agreeable or low on cold–quarrelsome can be judged only with reference to the scores of others.

Internal Consistency

In the selection of item sets to mark the vectors of the interpersonal domain, the principal itemmetric criterion was that of structural fidelity (Loevinger, 1957). By a variety of procedures, we attempted to ensure that each item selected was properly located with reference to a circumplex that we had adopted as the most appropriate measurement model for representing the relationships among categories of interpersonal perception. Items were also selected that had relatively high correlations with appropriate preliminary scales. Final scales thus constructed are likely to be internally consistent, although this must be demonstrated rather than assumed.

The next-to-last column of Table 3 presents coefficient alpha (Cronbach, 1951) estimates of internal consistency for each of the interpersonal scales in a combined sample of university students. The final column of the table indicates the range of alpha coefficients obtained in the four samples that constituted the combined student group. Although some of the interpersonal scales are more reliable than others, all of them meet a reasonably stringent requirement of acceptable internal consistency ($\alpha > .80$). The range of alpha coefficients in the different samples tends to be small, and this, together with the representativeness with which the universe of content was sampled, makes it reasonable to assume that the mean alpha values reported are good estimates of the characteristic degree of interitem structure for the different interpersonal variables (Loevinger, 1957). In this respect, the extraversion–introversion dimension (NO–FG) is the most internally cohesive, and the variables of unassuming–ingenuous (JK) and lazy–submissive (HI) are the least.

Discussion

The long-range goal of the research reported in this article was the development of a psychological taxonomy that would encompass the approximately 4,000 relatively familiar trait-descriptive terms identified by

[4] For example, Douglas N. Jackson (personal communication, November 27, 1975) is convinced that the present interpersonal circumplex structure, as well as others reported in the literature, can be accounted for in terms of his threshold theory of desirability responding (Rogers, 1971; Jackson, Note 8).

Norman (Note 3). The research strategy adopted for this task relied heavily on certain a priori distinctions between different domains of human characteristics. In the present work, the domain of interpersonal traits was defined in such a way as to distinguish it from other domains such as temperament, character, moods, and cognitive traits. These, or similar, distinctions have been made by most personality theorists concerned with taxonomic issues (e.g., Allport, 1937; Cattell, 1950; Murray, 1938). However *within* the interpersonal domain, the adoption of a two-dimensional real-space circumplex model (Benjamin, 1974) to capture the interrelationships among trait terms represents a personal methodological preference that is perhaps not as widely shared. On the basis of the work reported in this article, it seems fair to conclude that the domain of interpersonal trait descriptors *can* be classified, quite precisely, with reference to a circumplex model (a conclusion that was not obvious when we began. Since there are, no doubt, many other ways in which the interpersonal domain could be classified, under different definitions and different models, the specific advantages of the present framework must be discussed in more detail.

Two-dimensional circular orderings of interpersonal variables have been reported in the literature for more than 40 years (Schaefer, 1961), and the general conceptual model underlying this structure can be traced as far back as Galen (Roback, 1928). Remarkable "convergences" in conception and structure have been noted for studies varying widely in variables, populations, and measurement procedures (Foa, 1961; Schaefer, 1961; Wiggins, 1968). These convergences do not stem from the similarity of generative mechanisms postulated by different theorists (needs, interests, dynamisms, etc.) to account for surface patterns. Instead, the convergences reflect a set of semantic categories that impart a common meaning to social events observed by a variety of procedures. The categories distinguished by the ordinary language of personality are not sharply demarcated, and they are perhaps best thought of as "fuzzy sets" (Zadeh, Fu, Tanaka, & Shimura, 1975) rather than as precise logical distinctions. The circumplex model appears to be particularly well suited for representing elements (in the present case, adjectives) whose class membership is continuous rather than discrete.

Although often working in apparent isolation from one another, numerous investigators have proposed highly similar models of interpersonal behavior for the study of parents (Chance, 1959; Roe & Siegelman, 1963; Schaefer, 1959), children (Baumrind & Black, 1967; Becker & Krug, 1964; Schaefer & Bayley, 1963), parents and children (Benjamin, 1974; Foa, Triandis, & Katz, 1966), normal and abnormal adults (Leary, 1957; Lorr & McNair, 1963), psychotics (Lorr, Klett, & McNair, 1963), and college students (Stern, 1970). Within many of these diverse programs of research, the circumplex model has provided a nomological network that has enhanced the meaning and significance of the separate interpersonal variables employed (Schaefer, 1961). Unfortunately, the potential of the circumplex as an integrative conceptual model has not been as widely recognized by those whose research paradigms have focused on the intensive study of single dimensions in personality and social psychology.

In a recent book that presents summaries of research on the major dimensions of personality, the editors state:

> There obviously has been no overarching plan or theory, implicit or explicit, guiding the selection of topics for trait researchers. Indeed, the editors were forced to organize the book by means of the unsophisticated tactic of simply placing the chapters in alphabetical order. (London & Exner, 1978, p. xiv)

But surely the major research topics in contemporary personality and social psychology such as achievement and power (PA), Machiavellianism (BC), aggression (DE), introversion (FG), obedience (HI), interpersonal trust (JK), altruism and helping behavior (LM), affiliation and extraversion (NO)—can be related to each other by schemes more sophisticated than the alphabet. This is not to claim an established isomorphism between the research topics just enumerated and the present interpersonal taxonomy; rather, it is to illustrate the kinds of questions whose answers might bring conceptual clarity to

both taxonomic and experimental research. For example, once it is recognized that the dimensions measured by Bem's (1974) masculinity and femininity scales are the orthogonal interpersonal dimensions of ambitious–dominant (PA) and warm–agreeable (LM), then the issue of the "independence" of masculinity and femininity, so conceived, is seen in a new light (Wiggins & Holzmuller, 1978).

The alphabetical list of human "needs" provided by Henry Murray (1938) has proved to be an exceptionally fertile source of personality dimensions. Within the framework of Murray's personology, several of these dimensions have been studied in considerable depth (e.g., Atkinson, 1958; McClelland, 1961; Winter, 1973), although little attention has been given to the interrelationships among the needs in the overall system, perhaps because an alphabetical taxonomy provides no guidance in this matter. Needs selected from Murray's list have formed the basis for two multiscale inventories (Edwards, 1959; Jackson, 1967) and for scoring keys on an adjective checklist (Gough & Heilbrun, 1965). There appears to be a consistent factor structure associated with the dozen needs these instruments happen to share in common, but attempts to interpret this structure have not been particularly enlightening from a substantive point of view (Huba & Hamilton, 1976). In contrast, the strangely neglected work of Stern (1958, 1970) presents convincing evidence that a circumplex model provides a meaningful representation of the full range of Murray's need variables.

Two-dimensional circumplex models may have utility outside the realm of ordinary language-trait description associated with personality inventories and adjective checklists. The circular ordering of intercorrelations among clinical scales from the MMPI has been noted (Schaefer, 1961), and it has been argued that these scales can be interpreted within a two-dimensional circular model of personality (Kassebaum, Couch, & Slater, 1959). Given the manner in which the MMPI clinical scales were constructed, this argument appears rather farfetched on initial consideration. However, a recent study by Plutchik and Platman (1977) suggests a possible reason why this might be the case. Twelve trait-descriptive adjectives were selected to represent an interpersonal circumplex (Schaefer, 1961). Psychiatrists were then asked to rate seven diagnostic categories (e.g., paranoid, schizoid, hysterical) with respect to these adjectives. The mean adjective profiles of the seven diagnostic categories were intercorrelated, and two factors were extracted. The circular ordering of the diagnostic categories (Plutchik & Platman, 1977, p. 421) bears a striking resemblance to the circular ordering of corresponding MMPI scales reported elsewhere (Schaefer, 1961, p. 135). The constellations of traits implied by diagnostic labels can be represented by a two-dimensional interpersonal circumplex (Schaefer & Plutchik, 1966). To the extent that the MMPI clinical scales reflect conventional diagnostic labeling, they too would be expected to exhibit a circular ordering.

The present taxonomy of interpersonal traits may prove useful to investigators who employ single adjectives as stimulus materials in studies of interpersonal perception, impression formation, and trait attribution. The representativeness of the stimulus materials is always an issue in such studies, and the present taxonomy provides a systematic basis for sampling the entire domain of interpersonal traits as an alternative to, for example, the less differentiated strategy of sampling "favorable" traits. Elaborate sampling procedures for the study of one or more specific dimensions of interpersonal behavior are made possible by the availability of a list of 864 terms that have been classified within 16 categories. The position of an adjective within a given category is indexed by the correlation of that adjective with its own, its opposite, and its orthogonal clusters. In addition, the earlier itemmetric work of Norman (Note 3) provides information on these same adjectives with respect to such characteristics as social desirability, difficulty level, self-endorsement, and attributions to liked, indifferent, and disliked others. Although such itemmetric data are useful primarily for calibration of stimulus materials, they may themselves be profit-

ably studied for the light they shed on basic issues of social perception (Goldberg, 1978).

Reference Notes

1. Wiggins, J. S. *In defense of traits.* Invited address to the Ninth Annual Symposium on Recent Developments in the Use of the MMPI, Los Angeles, February 28, 1974.
2. Goldberg, L. R. *Language and personality: Developing a taxonomy of trait-descriptive terms.* Invited address to the Division of Evaluation and Measurement at the 86th Annual Convention of the American Psychological Association, San Francisco, August 27, 1977.
3. Norman, W. T. *2,800 personality trait descriptors: Normative operating characteristics for a university population.* Unpublished manuscript, University of Michigan, 1967.
4. Wiggins, J. S. & Marston, B. J. *An index for evaluating circumplexity in two-dimensional real-space models.* Unpublished manuscript, University of British Columbia, 1976.
5. Marston, B. J. Unpublished materials, University of British Columbia, 1976.
6. Merritt, N. G. *Personality and value characteristics of volunteers and professionals at three different social service agencies.* Unpublished master's thesis, University of British Columbia, 1977.
7. Wiggins, J. S., & Holzmuller, A. *Further evidence on androgyny and interpersonal flexibility.* Unpublished manuscript, University of British Columbia, 1978.
8. Jackson, D. N. *A threshold model for stylistic responding.* Paper presented at the 76th Annual Convention of the American Psychological Association, San Francisco, September 1968.
9. Norman, W. T. *On decomposing personality ratings.* Paper presented at the annual meeting of the Society of Multivariate Experimental Psychology, London, Canada, November 7, 1974.

References

Allport, G. W. *Personality: A psychological interpretation.* New York: Holt, 1937.

Allport, G. W., & Odbert, H. S. Trait names: A psycho-lexical study. *Psychological Monographs,* 1936, *47*(1, Whole No. 211).

Atkinson, J. W. (Ed.). *Motives in fantasy, action and society.* Princeton, N.J.: Van Nostrand, 1958.

Baumrind, D., & Black, A. E. Socialization practices associated with dimensions of competence in preschool boys and girls. *Child Development,* 1967, *38,* 291–328.

Becker, W. C., & Krug, R. S. A circumplex model for social behavior in children. *Child Development,* 1964, *35,* 371–396.

Bem, D. J., & Allen, A. On predicting some of the people some of the time: The search for cross-situational consistencies in behavior. *Psychological Review,* 1974, *81,* 506–520.

Bem, S. L. The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology,* 1974, *42,* 115–162.

Benjamin, L. S. Structural analysis of social behavior. *Psychological Review,* 1974, *81,* 392–425.

Carson, R. C. *Interaction concepts of personality.* Chicago: Aldine, 1969.

Cattell, R. B. *Personality: A systematic, theoretical, and factual study.* New York: McGraw-Hill, 1950.

Cattell, R. B. *Personality and motivation structure and measurement.* Yonkers-on-Hudson, N. Y.: World Book, 1957.

Chance, E. *Families in treatment; from the viewpoint of the patient, the clinician and the researcher.* New York: Basic Books, 1959.

Cronbach, L. J. Coefficient alpha and the internal structure of tests. *Psychometrika,* 1951, *16,* 297–334.

Crowne, D. P., & Marlowe, D. A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology,* 1960, *24,* 349–354.

Edwards, A. L. Social desirability and probability of endorsement of items in the Interpersonal Check List. *Journal of Abnormal and Social Psychology,* 1957, *55,* 394–396. (a)

Edwards, A. L. *The social desirability variable in personality assessment and research.* New York: Holt, Rinehart & Winston, 1957. (b)

Edwards, A. L. *Edwards Personal Preference Schedule.* New York: Psychological Corporation, 1959.

Foa, U. G. Convergences in the analysis of the structure of interpersonal behavior. *Psychological Review,* 1961, *68,* 341–353.

Foa, U. G. New developments in facet design and analysis. *Psychological Review,* 1965, *72,* 262–274.

Foa, U. G., & Foa, E. B. *Societal structures of the mind.* Springfield, Ill.: Charles C Thomas, 1974.

Foa, U. G., Triandis, H. C., & Katz, E. W. Cross-cultural invariance in the differentiation and organization of family roles. *Journal of Personality and Social Psychology,* 1966, *4,* 316–327.

Goldberg, L. R. Differential attribution of trait-descriptive terms to oneself as compared to well-liked, neutral, and disliked others: A psychometric analysis. *Journal of Personality and Social Psychology,* 1978, *36,* 1012–1028.

Gough, H. G., & Heilbrun, A. B. *The Adjective Check List manual.* Palo Alto, Calif.: Consulting Psychologists Press, 1965.

Guilford, J. P. *Personality.* New York: McGraw-Hill, 1959.

Guttman, L. A new approach to factor analysis: The radex. In P. R. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences.* Glencoe, Ill.: Free Press, 1954.

Heider, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

Homans, G. C. *Social behavior: Its elementary form.* New York: Harcourt, Brace, 1961.

Huba, G. J., & Hamilton, D. L. On the generality of trait relationships: Some analyses based on Fiske's paper. *Psychological Bulletin*, 1976, *83*, 868–876.

Jackson, D. N. *Personality Research Form manual*. Goshen, N.Y.: Research Psychologists Press, 1967.

Jackson, D. N. A sequential system for personality scale development. In C. D. Spielberger (Ed.), *Current topics in clinical and community psychology* (Vol. 2). New York: Academic Press, 1970.

Kassebaum, G. G., Couch, A. S., & Slater, P. E. The factorial dimensions of the MMPI. *Journal of Consulting Psychology*, 1959, *23*, 226–236.

Kelly, G. A. *The psychology of personal constructs* (Vols. 1 & 2). New York: Norton, 1955.

LaForge, R., & Suczek, R. The interpersonal dimension of personality: III. An interpersonal check list. *Journal of Personality*, 1955, *24*, 94–112.

Leary, T. *Interpersonal diagnosis of personality*. New York: Ronald Press, 1957.

Loevinger, J. Objective tests as instruments of psychological theory. *Psychological Reports*, 1957, *3*, 635–694 (Monograph No. 9).

London, H., & Exner, J. E., Jr. (Eds.). *Dimensions of personality*. New York: Wiley, 1978.

Lorr, M., Klett, C. J., & McNair, D. M. *Syndromes of psychosis*. New York: Macmillan, 1963.

Lorr, M., & McNair, D. M. An interpersonal behavior circle. *Journal of Abnormal and Social Psychology*, 1963, *67*, 68–75.

Lorr, M., & McNair, D. M. Expansion of the interpersonal behavior circle. *Journal of Personality and Social Psychology*, 1965, *2*, 823–830.

McClelland, D. C. *The achieving society*. Princeton, N.J.: Van Nostrand, 1961.

Mischel, W. *Personality and assessment*. New York: Wiley, 1968.

Murray, H. A. *Explorations in personality*. New York: Oxford University Press, 1938.

Osgood, C. E. Speculation on the structure of interpersonal intentions. *Behavioral Science*, 1970, *15*, 237–254.

Plutchik, R., & Platman, S. R. Personality connotations of psychiatric diagnoses. *Journal of Nervous and Mental Disease*, 1977, *165*, 418–422.

Rinn, J. L. Structure of phenomenal domains. *Psychological Review*, 1965, *72*, 445–466.

Roback, A. A. *The psychology of character*. New York: Harcourt, Brace, 1928.

Roe, A., & Siegelman, M. A parent–child questionnaire. *Child Development*, 1963, *34*, 355–369.

Rogers, T. B. The process of responding to personality items: Some issues, a theory and some research. *Multivariate Behavioral Research Monographs*, 1971, *6*, No. 2.

Schaefer, E. S. A circumplex model for maternal behavior. *Journal of Abnormal and Social Psychology*, 1959, *59*, 226–235.

Schaefer, E. S. Converging conceptual models for maternal behavior and child behavior. In J. C. Glidewell (Ed.), *Parental attitudes and child behavior*. Springfield, Ill.: Charles C Thomas, 1961.

Schaefer, E. S., & Bayley, N. Maternal behavior, child behavior, and their intercorrelations from infancy through adolescence. *Monographs of the Society for Research in Child Development*, 1963, *28*(3, Whole No. 87).

Schaefer, E. S., & Plutchik, R. Interrelationships of emotions, traits, and diagnostic constructs. *Psychological Reports*, 1966, *18*, 399–410.

Stern, G. G. *Preliminary manual: Activities Index-College Characteristics Index*. Syracuse, N.Y.: Syracuse University Psychological Research Center, 1958.

Stern, G. G. *People in context: Measuring person-environment congruence in education and industry*. New York: Wiley, 1970.

Sullivan, H. S. *The interpersonal theory of psychiatry*. New York: Norton, 1953.

Webster's new international dictionary (2nd ed., unabridged). Springfield, Mass.: Merriam, 1924.

Webster's third new international dictionary of the English language unabridged. Springfield, Mass.: Merriam, 1961.

Wiggins, J. S. Interrelationships among MMPI measures of dissimulation under standard and social desirability instructions. *Journal of Consulting Psychology*, 1959, *23*, 419–427.

Wiggins, J. S. Personality structure. *Annual Review of Psychology*, 1968, *19*, 293–350.

Wiggins, J. S., & Holzmuller, A. Psychological androgyny and interpersonal behavior. *Journal of Consulting and Clinical Psychology*, 1978, *46*, 40–52.

Wiggins, J. S., Renner, K. E., Clore, G. L., & Rose, R. J. *The psychology of personality*. Reading, Mass.: Addison-Wesley, 1971.

Winter, D. G. *The power motive*. New York: Free Press, 1973.

Zadeh, L. A., Fu, K.-S., Tanaka, K., & Shimura, M. (Eds.). *Fuzzy sets and their applications to cognitive and decision processes*. New York: Academic Press, 1975.

Received April 14, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 3, 413–423

Self-Awareness, Psychological Perspective, and Self-Reinforcement in Relation to Personal and Social Standards

Ed Diener and Thomas K. Srull
University of Illinois at Urbana-Champaign

The present study was designed to assess whether subjects would be more likely to judge their own behavior from a social perspective when they were self-aware than when they were non-self-aware. Participants reinforced themselves after receiving feedback about a performance task that indicated they had surpassed their own standard, a social standard, both standards, or neither standard. Subjects did not alter their reinforcement as a function of the self-awareness manipulation when they surpassed both standards or neither standard. However, when subjects learned that they had surpassed their own standard but not the social standard, they rewarded themselves significantly more ($p < .01$) and felt more satisfied with their performance ($p < .01$) when they were non-self-aware. When subjects surpassed the social standard but not the personal standard, they reinforced themselves significantly more ($p < .01$) and felt more satisfied ($p < .01$) when self-aware. It was also found that self-awareness decreased feelings of choice in administering self-reinforcement ($p < .001$). The possibility that self-awareness may create a social perspective from which one's own behavior is evaluated and the implications this possibility has for future theory development are discussed.

Duval and Wicklund (1972) outlined a theory of self-awareness that has been extended and refined by Wicklund (1975, in press). The theory maintains that attention may be focused either outward or upon the self as an object. One major emphasis of the theory of self-awareness has been on standard-related behavior. The theory maintains that self-awareness should increase adherence to normative standards, because the self-aware person will be highly aware of any deviations between his or her behavior and these standards. Moreover, since perceived discrepancies between one's behavior and salient standards will cause negative affect, the self-aware person is more likely to avoid antinormative behavior.

Several studies have demonstrated the impact of self-awareness on normative behavior. For example, Diener and Wallbom (1976) found a large reduction in cheating when college student subjects were made self-aware. It has also been found that objective self-awareness increases aggression when subjects think aggression is normative (Carver, 1974), but decreases aggression when they believe it is antinormative (Rule, Nesdale, & Dyck, 1975; Scheier, Fenigstein, & Buss, 1974). The present experiment concerns two ambiguities present in current conceptualizations of self-awareness. The first is whether self-awareness necessarily increases adherence to all types of standards. The second is whether self-awareness produces a purely introspective examination of oneself or whether, under some conditions, self-awareness may create a social perspective from which one's own behavior is evaluated.

The present research was supported by U.S. Public Health Service Grants MH-14257 and MH-15140 to the second author. Appreciation is extended to Fran DeNinno and Dennis Heim for serving as experimenter and assistant, respectively. We also thank Robert Wicklund, Robert S. Wyer, and two anonymous reviewers for their valuable comments on an earlier draft of the manuscript.

Requests for reprints should be sent to Ed Diener, Department of Psychology, University of Illinois, Champaign, Illinois 61820.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3703-0413$00.75

Types of Standards

The major purpose of the present study was to explore the question of whether self-awareness increases adherence to all types of norms, only to social norms, or only to personal norms. According to the original theory of objective self-awareness, compliance with both personal and social standards should increase in the state of objective self-awareness. In other words, the theory maintains that *any* salient standard will exert more influence if the person is objectively self-aware, regardless of whether the standard is social or personal. It is possible, however, that self-awareness operates mainly upon one or the other type of standard. Alternatively, Fenigstein, Scheier, and Buss (1975) have identified two types of self-awareness, public and private, and it is possible that adherence to one type of standard will depend on which type of self-awareness is manipulated. It is clear that social and personal standards will often covary to some degree, and no effort has been made in past research to examine their independent effects.

Demonstrations of increased norm adherence have relied mostly on socially imposed standards (cf. Carver, 1974; Scheier et al., 1974). It is difficult, however, to draw any firm conclusions regarding the relative influences of personal and social standards from this literature. In most situations, including those used in past self-awareness research, social and personal standards are undoubtedly intertwined. For example, Carver (1975) found that self-reported belief in punishment led to greater administering of electric shock, but only when participants were self-aware. In contrast, Scheier (1976) found that self-awareness, but not personal attitudes toward punishment, led to heightened aggression among angry subjects. However, these findings do not indicate that social or personal standards were necessarily more powerful or were the locus of the self-awareness effects. Note that in each of the studies mentioned above, personal and social standards were both relevant to the behavior in question. Moreover, the social standard had to be filtered by the subject's perceptions. Thus, based on past research, it is difficult to separate the relative influence of the two types of standards. Even in cases where self-awareness has increased compliance to personal standards, it is quite possible that subjects perceived the social standard to be congruent with their own personal standard. Indeed, there is evidence to show that in the absence of contradictory information, people assume a close relationship between self- and social standards (Ross, Greene, & House, 1977).

Psychological Perspective and Attention to Standards

Another ambiguity associated with past research is suggested by Mead's (1934) self theory, upon which the self-awareness model was originally based. According to Mead, self-awareness is the ability to look at oneself as others do, to adopt an outside social perception of oneself. This suggests that the self-aware person might become more concerned with the opinions of others and more empathic with their viewpoint and therefore conform more to social expectations. On the other hand, the self-aware person might be more aware of personal standards because of self-focused attention and therefore conform more to them. This raises the further question of whether the relative salience of personal and social standards is the same. More specifically, it is possible that the salience of personal and social standards varies dramatically across different situations, and this factor has also been neglected in past research. In the present study, personal and social standards were both made highly explicit and varied independently. This procedure allowed an examination of the relative impact of personal and social standards when they are similar as well as when they are dissimilar.

It is undoubtedly the case that adherence to personal and social standards is not mutually exclusive, and one of the most important immediate problems is to delineate those situations in which the relative impact of one type of standard is greater than the other. By making both standards highly salient, and thereby controlling for differences in attention, the present study is concerned with one such important class of theoretically relevant situations.

In sum, it is suggested that self-awareness has two principal components: attention to aspects of the self and a social perspective of oneself. It is possible that these two components underlie the two individual difference patterns of self-awareness (private and public) identified by Fenigstein et al. (1975). It is hypothesized that when social and personal standards are independently varied, and *attention to standards is held constant*, self-awareness will have a greater impact on adherence to social than personal standards, because of the social perspective subjects have of themselves. The differential reliance on standards was assessed in a self-reinforcement paradigm. Self-reinforcement was adopted because it reflects in a concrete way an individual's evaluation of his or her behavior. In addition, Diener (in press) has recently stressed the interplay between self-awareness and self-reinforcement in developing a theory of deindividuation. Thus, the use of self-reinforcement as a dependent measure provides a potential operational bridge between the theories of self-awareness and deindividuation.

Method

Overview

Participants made a series of 12 perceptual judgments and generated a standard of accuracy within which they would consider an individual's performance to be successful. Later they were given systematically varied bogus feedback concerning their own performance levels, their self-standards of success, and social standards of success ostensibly generated by their peers. Trials on which the subject surpassed both standards, one or the other standard, or neither standard were thereby created. On the basis of the social and self-standards and their own performance levels, participants were allowed to administer reinforcement to themselves under self-aware and non-self-aware conditions.

Participants

Participants were 24 male and 24 female University of Illinois introductory psychology students. Participation in the experiment partially fulfilled a course requirement.

Procedure

Subjects were run in individual sessions. Each participant was greeted by the female experimenter, who introduced her male assistant. Both the experimenter and assistant were blind to the experimental hypothesis. The assistant requested the subject to recite his or her name and hometown into a microphone. This information was recorded on a tape loop for later use.

The subject was told that he or she was participating in the development of a perceptual judgment task. The experimenter explained that she was interested in the maximum performance of which people are capable under actual testing conditions. She stated that there are a variety of distractions, such as the presence of proctors and stray noises, present in testing situations and that concentration is important in determining how well one performs. She explained that there would therefore be a variety of distractions present during the experiment. She stated that as an incentive to perform well, the subject would later be able to reward him- or herself by taking from 0 to 9 "points" on each trial. Furthermore, at the end of the experiment, the total number of points accumulated would be converted to money.

A slide projector was used to provide an example of the perceptual task that subjects would be asked to complete. After displaying a practice slide, the experimenter explained that each of the test slides would contain a large number of dots that would be shown for approximately 4 sec. The subject's task consisted of two parts. First, he or she was to estimate the number of dots on the slide. Second, the subject was told:

> Obviously you will not be able to guess the exact number of dots after seeing the slide for only four seconds. Therefore, I would also like you to tell me within what range you think you should be for it to be a successful performance. That is, within how many dots should your estimate be for you to consider it a success? For example, you might estimate this slide to have, say, 160 dots and think you should be within 25 dots to have a successful estimate.

The subject was then told that he or she would go through a series of test trials. Later, the experimenter would review each trial and report the participant's actual performance level, his or her self-standard for a successful performance, and a social standard of success that was based on the performance of 250 introductory psychology students from the previous semester. The subject would then be able to reward him- or herself for each trial.

After any questions had been answered, the first test slide was presented. With the onset of the first slide, white noise began to filter into the room, ostensibly as the first distractor. Throughout this procedure, the experimenter's assistant could be seen sitting at a desk, recording the subject's responses and performing a number of calculations on a calculator. Although the actual performance of subjects on these tasks was of no import to the study, obtaining the estimates was necessary to make credible the later feedback about their performance and self-standards.

At the end of the 12th slide, the experimenter explained that her assistant had been calculating how well the participant had performed on each trial in terms of percentage of wrong answers. She then took a booklet from her assistant and began reviewing the 12 slides. For each trial, the subject was told his or her own "performance level," his or her stated "self-standard" of success, and the "social standard" of success. These three figures were represented in supposed percentages and shown on a large visual display to facilitate comparison. Obtaining estimates and standards in terms of raw numbers and later representing them as percentages allowed the experimenter to provide the false feedback without suspicion. After receiving the feedback for each trial, the subject took from zero to nine marbles from a container, placed them in an empty bowl as a reward, and told the experimenter how many marbles were taken. The number of marbles taken each time was recorded by the experimenter. The subject then made several short ratings, and the experimenter went on to the next trial.

At the end of the 12th trial, subjects were completely debriefed. Participants were told that the feedback and standards were bogus. The experimenter carefully explained why this was a necessary part of the experiment and said that since she had promised that the marbles would be converted to money, she was paying each participant $2. The experimenter fully discussed any concerns of the subject and was alert to any signs of upset about the deception.

Self-awareness manipulation. Each subject was exposed to both self-aware and non-self-aware manipulations and was led to believe that these were different types of distractors. The self-aware condition was created by having the subject's image displayed on a television screen. It has been shown that seeing one's own image increases self-focused attention (Carver & Scheier, 1978; Geller & Shaver, 1976). In addition, the subject's tape-recorded voice was played on a continuous loop, repeatedly saying, "My name is _____ and I am from _____." Thus, subjects in the self-aware condition were exposed to a voice (ostensibly similar to a person talking) and a televised face (ostensibly similar to a proctor watching them). In the non-self-aware condition, an exciting automobile chase scene was played on the television. The videotape included both picture and sound and was presented as the type of distractor that might occur while a person is studying. The order of the self-aware and non-self-aware conditions was counterbalanced across subjects. Each subject rewarded him- or herself on Trials 1 through 6 while in one condition and on Trials 7 through 12 in the other condition.

Feedback conditions. As noted above, systematic feedback was given to each subject for his or her "performance level," "self-standard" of success, and "social standard" of success. A total of six feedback conditions was created by having the subject's ostensible performance on different trials surpass either (a) both the self- and social standard, (b) the self- but not the social standard, (c) the social but not the self-standard, (d) neither standard, (e) the self-standard, with no information about the social standard, or (f) the social standard, with no information about the subject's previous self-standard. Each subject received the six feedback conditions in a different random order. The feedback was constructed in such a way that the percentage figures given subjects varied considerably from trial to trial in how close they were to the actual number of dots presented, in order to make the feedback more believable. This maneuver seemed reasonable, since some trials appeared very easy, while others appeared very difficult. However, the *discrepancy* between the subject's ostensible performance and either of the two standards was, although not exact, close for each trial. These discrepancies were controlled by making them equivalent when summed over subjects in any given condition.

Self-report measures. A variety of self-report measures were obtained to explore the psychological processes underlying subjects' behavior in self-aware and non-self-aware conditions, to check on the validity of the manipulations, and to check for possible suspicion. After the subject rewarded him- or herself on each trial, two rating scales were completed. One asked, "How satisfied were you with your performance on the last task?" and the other asked, "How much choice did you feel you had in deciding how many points you should take?" At the end of the 6th and 12th trials (i.e., at the end of the self-aware and non-self-aware conditions), subjects were also asked the extent to which they considered their own standards of success and the extent to which they considered the way they might appear to others in rewarding themselves over the previous 6 trials. All of the ratings were made on 11-point scales. The intrusion of additional rating scales after the 6th and 12th trials was not as conspicuous as it might seem, because they occurred at the same time the assistant was changing from one set of "distractors" to another. At the end of the 12 trials, subjects were asked to guess the hypothesis of the study, to say whether they had heard about the experiment beforehand, and to say what thoughts they had about the feedback they received.

Experimental Design

Subjects were exposed to each feedback condition under both self-aware and non-self-aware conditions. The feedback conditions were presented in a different randomized order for each subject. Thus, the experiment was a 2 (sex) × 2 (order of self-aware and non-self-aware conditions) × 2 (self-aware vs. non-self-aware) × 6 (feedback conditions) factorial design with repeated measures on the last two factors.

Table 1
Mean Number of Marbles Taken and Ratings of Personal Satisfaction Under Self-Aware and Non-Self-Aware Conditions as a Function of Type of Feedback

Condition	Surpass self- and social standard	Surpass self- but not social standard	Surpass social but not self- standard	Surpass neither standard	Surpass self- standard only	Surpass social standard only	Total
			Number of marbles				
Own TV image and voice	6.81	3.44	5.75	.88	4.15	6.73	4.63
Chase scene	6.73	5.50	3.65	.98	6.10	4.13	4.51
Total	6.77	4.47	4.70	.93	5.13	5.43	4.57
			Satisfaction				
Own TV image and voice	7.21	3.81	6.27	2.48	5.27	7.50	5.42
Chase scene	7.42	6.00	4.50	2.56	6.65	5.25	5.40
Total	7.31	4.90	5.39	2.52	5.96	6.38	5.41

Note. The possible range of marbles was 0 to 9 (higher numbers indicating greater self-reinforcement), and the possible range of satisfaction ratings was 0 to 10 (higher values indicating greater satisfaction).

Results

The order in which subjects experienced the self-aware (own televised image and own voice) and non-self-aware (televised chase scene) conditions did not produce a significant difference on any of the dependent variables. Sex of subject was significant to the extent that, across conditions, males took more marbles ($M = 4.95$) than females ($M = 4.19$), $F(1, 44) = 9.28$, $p < .01$. The order of the self-awareness manipulation and sex of subject were thus eliminated as variables, and the design was considered a 2 (self-awareness vs. non-self-awareness) × 6 (feedback conditions) within-subjects factorial for all further analyses. It should also be noted that no subject had heard about the experiment beforehand, suspected the feedback was false, or was able to guess the hypothesis of the study. These possibilities were assessed both in the written questionnaire and in verbal interviews.

Self-Administered Reward and Personal Satisfaction

The mean number of marbles taken by subjects in the 12 conditions is presented in Table 1. There was a significant main effect for feedback condition, $F(5, 235) = 167.04$, $p < .001$, and, consistent with the hypothesis, a highly significant interaction between feedback condition and the self-awareness manipulation, $F(5, 235) = 60.62$, $p < .001$. Specific sets of comparisons are particularly germane to the original hypothesis. Those trials in which subjects received feedback indicating that they surpassed their self-standard but not the social standard, or surpassed the social standard but not their self-standard, should indicate the relative weight subjects put on each type of standard in administering reward under self-aware and non-self-aware conditions. Similarly, those trials in which subjects surpassed only the self- or only the social standard will indicate the relative importance of these standards across self-aware and non-self-aware conditions.

Post hoc comparisons using procedures suggested by Tukey (see Keppel, 1973) were therefore computed on selected pairs of means. These comparisons revealed that when subjects surpassed their own but not the social standard, they rewarded themselves significantly more highly in the non-self-aware than in the self-aware condition ($p < .01$). In con-

Table 2
Mean Ratings of Choice Under Self-Aware and Non-Self-Aware Conditions as a Function of Type of Feedback

Condition	Surpass self- and social standard	Surpass self- but not social standard	Surpass social but not self- standard	Surpass neither standard	Surpass self- standard only	Surpass social standard only
Own TV image and voice	6.19	5.29	5.31	4.60	5.42	5.63
Chase scene	7.90	7.60	7.69	7.56	7.73	7.48
Difference	1.71	2.31	2.38	2.96	2.31	1.85

Note. The possible range of ratings was 0 to 10, with higher values indicating a greater feeling of choice.

trast, when subjects surpassed the social but not the self-standard, they rewarded themselves more highly in the self-aware than in the non-self-aware condition ($p < .01$). Further, an identical pattern appears when subjects were only given feedback with respect to one of the standards. Specifically, when subjects surpassed their self-standard but were not told the social standard, they rewarded themselves significantly more highly in the non-self-aware than in the self-aware condition ($p < .01$). However, when subjects surpassed the social standard but were not told their original self-standard, they rewarded themselves more in the self-aware than in the non-self-aware condition ($p < .01$). Finally, it is also important to note that there was no difference in self-reinforcement as a function of self-awareness when subjects surpassed both standards or neither standard.

Subjects were asked how satisfied they were with their performance following the feedback on each trial. It was thought that these ratings would also reflect the relative importance subjects placed on surpassing their own and the social standards in the two conditions. The mean ratings of satisfaction, which are also presented in Table 1, parallel those for self-reinforcement quite closely. There was a main effect for feedback condition, $F(5, 235) = 66.06$, $p < .001$, and a significant interaction between feedback condition and the self-awareness manipulation, $F(5, 235) = 25.95$, $p < .001$.

Post hoc comparisons revealed a pattern identical to that for self-reinforcement. Specifically, subjects rated themselves more satisfied with their performance in the non-self-aware than in the self-aware condition when they surpassed their own but not the social standard ($p < .01$). When they surpassed the social but not their personal standard, however, subjects rated themselves as more satisfied in the self-aware than in the non-self-aware conditions ($p < .01$). Similarly, when subjects surpassed their own standard but did not know the social standard, they rated themselves as more satisfied in the non-self-aware than in the self-aware condition ($p < .01$). However, this finding was again reversed when subjects surpassed the social standard but were not given their original self-standard. Here, subjects rated themselves as more satisfied in the self-aware than in the non-self-aware condition ($p < .01$). Consistent with the hypothesis, there was again no difference as a function of self-awareness when subjects surpassed both standards or neither standard.

One might expect ratings of satisfaction and amount of self-administered reward to be very redundant measures. However, the correlations between these two variables, within each of the conditions, ranged from .13 to .59, with a mean of .41. This result suggests that each of the measures contributed some independent information.

Additional Ratings

Following each trial, subjects rated how much choice they felt they had in deciding how many points to take. The mean ratings of subjects for each of the 12 conditions are presented in Table 2. There was a significant feedback condition main effect, $F(5, 235) =$

2.98, $p < .02$. Examination of the means reveals that, as one would expect, subjects felt that they had the most choice when they surpassed both standards and the least choice when they surpassed neither standard, with intermediate values in the other conditions. More interesting was a strong main effect for the self-awareness manipulation, $F(1, 47) = 50.94$, $p < .001$. Post hoc comparisons indicate that subjects felt that they had more choice when non-self-aware (overall $M = 7.66$) than when self-aware (overall $M = 5.41$) in *every one* of the six feedback conditions (all at least at $p < .05$).

Each subject was asked to make several additional ratings after the 6th and 12th trials (i.e., following the entire block of self-aware and non-self-aware conditions). One question asked to what extent subjects considered their own standards of success in rewarding themselves over the previous 6 trials. Results indicate that subjects considered their own standards of success more in non-self-aware ($M = 7.98$) than in self-aware ($M = 6.13$) conditions, $t(47) = 5.64$, $p < .001$. A second question asked to what extent subjects considered the way they might appear to others in rewarding themselves over the previous six trials. Here, subjects considered the way they might appear to others more in self-aware ($M = 7.77$) than in non-self-aware ($M = 5.25$) conditions, $t(47) = 7.63$, $p < .001$. These data provide a check on the intended manipulation of perspective and add further support to the general contention that subjects see themselves as social objects and are more concerned with social standards when they are made self-aware.

Discussion

Summary of Findings

The major finding of the present study was that subjects relied more on their personal standards of accuracy for evaluating their performance when they were non-self-aware, whereas they relied more on the social standards of accuracy when they were self-aware. Even though the social group from whom the public standard was derived was not present, self-awareness increased reliance upon its standard. Several different measures all pointed to the fact that self-aware subjects were more concerned with their social selves and felt more pressure to comply with the social standard. The self-reinforcement and satisfaction ratings proved to be totally consistent in demonstrating the self-aware person's heavy reliance on social standards. In addition, self-reinforcement was very similar across self-aware conditions in those situations in which subjects surpassed both standards or neither standard. This finding suggests that self-awareness has no effect when both personal and social standards are highly salient and are consistent with each other.

The increased importance of social standards to the self-aware subjects was clearly confirmed in every one of the four conditions in which subjects received feedback that they had surpassed one or another of the standards. In addition, whether subjects received feedback on only one or two standards was unimportant; the social standard produced the large impact on self-aware participants in all conditions. It is important to note that since this was an entirely within-subjects design, these differences represent a dramatic shift in the relative reliance on one standard or the other as a function of the self-aware conditions.

The responses to questions asked at the end of the self-aware and non-self-aware periods confirmed the greater weighting of social standards by self-aware subjects. When asked how concerned they were with their personal standards, non-self-aware subjects reported being significantly more concerned than objectively self-aware subjects. However, when asked about their concern for how they would appear to others, objectively self-aware subjects were the most concerned. Finally, self-aware subjects reported feeling less freedom in choosing their level of self-reward, indicating that they felt less self-directed and more constrained by the situation. Thus, multiple measures across a variety of conditions all demonstrated the objectively self-aware person's greater reliance on social standards. These findings clearly demonstrate that when standards are manipulated in a social situation and attention to the standards is held constant, self-aware persons will heavily weight the social stan-

dard. It would appear that self-aware persons take on a social perspective of themselves in a social situation and judge their performance from this external vantage point.

A close examination of the self-reinforcement and satisfaction data reveals an interesting finding in addition to the clear shift noted above. A number of comparisons all indicate that subjects in one condition or the other weighted a certain standard more heavily, but did not ignore the other standard. That is, all subjects took both standards into account, but simply shifted the relative weight placed on them. When one compares the cells in which only one standard was given with the comparable surpassed-one-but-not-the-other cells, it can be seen that the former are higher for every comparison. This indicates that the standard that was given less weight nevertheless had a definite effect. The fact that the two surpass-both-standard cells have higher values than all others also indicates that one standard was never ignored completely. Thus, self-awareness shifted the weight given the two types of standards, but both standards were always considered.

It is informative to examine more closely the data concerning subjective feelings of choice. In the non-self-aware conditions, subjects felt that they had free choice in determining their reward even when they surpassed neither standard. Thus, even when they gave themselves little reward, they felt that they had a great deal of choice. When self-aware, however, subjects felt much less freedom when they surpassed neither standard. Even more impressive is the fact that self-aware subjects felt that they had less choice in choosing their reward when they surpassed *both* standards than they did in any of the non-self-aware conditions. In fact, the self-aware participants' feelings of choice were highly affected by their performance, whereas this was much less true for non-self-aware subjects. Thus, self-awareness appears not only to decrease feelings of free choice overall but also tends to produce a greater feeling of constraint on the person who has performed poorly. When a person is non-self-aware, he or she appears to feel free in rewarding and punishing him- or herself, suggesting a feeling of internal self-direction.

One interesting finding of the present study was that self-awareness had no impact on self-reinforcement *or* feelings of satisfaction when subjects surpassed both standards or failed by both standards. This indicates that at least when standards are clear and explicit self-aware subjects do not judge themselves more harshly when wrong or reward themselves more liberally when successful. If attention to standards were allowed to vary, however, it is possible that the self-aware person would be more aware of behavior–standard discrepancies and thus reinforce himself or herself differently. Indeed, this is what is typically found in the self-awareness literature (Wicklund, 1975). However, the present data indicate that when attention to standards remains high, self-aware subjects do not judge themselves differently across all types of standards, but simply rely more heavily on the social standard.

Psychological Perspective and Attention to Standards

As noted earlier, one recent line of evidence in the objective self-awareness literature suggests that there are two types of trait self-awareness, private and public (Fenigstein et al., 1975). Private self-awareness is said to be a focus of attention on the covert, internal aspects of oneself, whereas public self-awareness is what is commonly thought of as "self-consciousness"—an awareness of the self as a social object. A person who is privately self-aware should experience a heightened awareness of his or her own beliefs and feelings, whereas a publicly self-aware person should be more aware of his or her social appearance and the impression he or she makes on others. This distinction in the conceptualization of dispositional self-awareness suggests a parallel two-factor approach to the underlying processes involved in state self-awareness—one that emphasizes both perspective and attention. This distinction raises several interesting possibilities that are noted below.

There are several possible explanations for the apparent divergence between the present findings and past results showing greater ad-

herence to personal standards among self-aware subjects. A strong possibility is that personal standards are often adhered to more closely by self-aware subjects because of their close perceived correspondence with social norms (cf. Ross et al., 1977). Another possibility is that self-aware subjects are more aware of both social and personal standards than are non-self-aware subjects, but where the two are clearly discrepant, they weight social standards more heavily than usual because of their social perspective of themselves.

One explanation for the absence of increased adherence to personal standards in the present study concerns the difference pointed out earlier between attention and perspective. In most studies, self-awareness is manipulated, and attention is free to vary, focusing either on aspects of the self or on the environment. Thus, if the typical self-awareness paradigm had been employed, attention would have been free to focus on the external environment or on social or private standards. In such a case, the person might *attend* more to both personal and public standards under self-aware conditions and thus be influenced more by both of them. This interpretation is supported by the findings of Rule et al., (1975), who found that self-aware subjects complied with their own beliefs more when in a situation where no standards had been made explicit.

The procedure used in the present study, however, presented standards in such a way that they were explicit and easily comparable. Thus, attention to the two types of standards was likely to be high in both self-aware and non-self-aware conditions. However, the viewpoint or perspective of the subject was likely to change across the two conditions. The self-aware person was likely to see him- or herself as a social object, and this apparently led subjects to weight the social standard much more heavily. In the non-self-aware condition, subjects were also likely to be equally aware of both types of standards, but here they gave greater weight to the personal standard, apparently because of their egocentric perspective.

This analysis suggests an important distinction between the attention given to different types of standards and psychological perspective in the analysis of self-awareness affects. In order to determine the importance of this distinction, future research will need to examine adherence to the two types of standards when they vary in explicitness, as a function of self-awareness. Since attention to various standards may vary in many situations, a theory explaining which standards will be most salient (and hence most magnified in importance by self-awareness) when standards are not highly explicit is needed. Wicklund (in press) has offered the concept of "end points" (roughly analogous to goals) to explain which standards the self-aware person is likely to attend to, and this approach should be vigorously explored.

Possible Limitations

One factor may limit the generality of the present findings and suggests the need for additional study. It is possible that the present results are due to the fact that subjects did not have a firm basis for their self-standard and thus quickly abandoned it in favor of the social standard when made self-aware. If the person had a strong relevant personal standard, self-awareness might serve to enhance its influence.

Another potential limitation of the present findings is that they may be applicable only to *public* self-awareness (also referred to as self-consciousness). That is, it is possible that the self-awareness manipulation combined with the interaction with the two experimenters induced public self-awareness. Preliminary analyses of a subsequent study in which the self-reinforcement situation was less social in nature indicate that the present pattern of findings may not emerge in a non-social situation. Hence, the findings of increased reliance on social standards may be most applicable to public self-awareness or self-awareness produced in a social situation. Indeed, the process underlying public self-awareness may be the psychological perspective effect varied in the present design. It is interesting to speculate that the effects of private self-awareness depend on attentional differences that were intentionally held constant in the present study.

Future Research

Although it is clear that self-aware subjects relied more upon the social standard, it is also apparent that non-self-aware subjects relied more upon their personal standard. It seems likely that subjects approached the situation with an initial bias toward weighting their personal standard more heavily, possibly because of cultural values related to independence and self-reliance. One possible explanation for the reliance on personal standards in the non-self-aware condition is that the non-self-aware person may have an egocentric orientation that favors personal goals and standards when he or she is aware of them. This possibility deserves further research. The different effects of self-awareness on adherence to performance standards (e.g., those oriented toward reaching a specific goal) and norms (e.g., those used to judge moral or value-related behavior) should also be explored.

Finally, the distinction drawn earlier between the person's perspective or viewpoint and the focus of attention in the self-aware state also provides an intriguing avenue for future investigation. It will be important to examine whether self-aware persons are equally aware of both social and personal standards. If so, it will also be important to determine the precise mechanism through which one type of standard is given more weight than the other. It is possible that *public* self-awareness predisposes one to take the view of others and thus to emphasize social standards. However, *private* self-awareness may focus one's attention on internal factors and thus increase adherence to personal standards. This suggests that the public and private individual difference dimensions may rest on two psychological processes—focusing attention on aspects of the self and taking a social perspective of oneself—that together comprise what has become known as "self-awareness." Moreover, it may be possible to vary these two factors independently. Thus, future research should explore the relation of perspective and attention to public and private self-awareness. One important direction for future theoretical development will be to discover the effects that different types of self-awareness have on adherence to personal and social standards.

Conclusions

The present experiment indicates that self-aware persons may rely more heavily on social standards when judging their performance than non-self-aware persons do. However, the present findings may be limited to variations in social self-awareness. Results also indicated that self-aware subjects felt less freedom in choosing the level of reinforcement for their behavior, suggesting that they felt more constrained by a social perspective of themselves. Focused attention to various aspects of oneself and taking a social perspective of oneself were postulated to be two distinct components of self-awareness. It is suggested that these two components parallel the public and private individual difference dimensions of self-consciousness reported by Fenigstein et al. (1975).

References

Carver, C. S. Facilitation of physical aggression through objective self-awareness. *Journal of Experimental Social Psychology*, 1974, *10*, 365–370.

Carver, C. S. Physical aggression as a function of objective self-awareness and attitudes toward punishment. *Journal of Experimental Social Psychology*, 1975, *11*, 510–519.

Carver, C. S., & Scheier, M. F. Self-focusing effects of dispositional self-consciousness, mirror presence, and audience presence. *Journal of Personality and Social Psychology*, 1978, *36*, 324–332.

Diener, E. Deindividuation: The absence of self-awareness and self-regulation in group members. In P. Paulus (Ed.), *The psychology of group influence*. Hillsdale, N.J.: Erlbaum, in press.

Diener, E., & Wallbom, M. Effects of self-awareness on antinormative behavior. *Journal of Research in Personality*, 1976, *10*, 107–111.

Duval, S., & Wicklund, R. A. *A theory of objective self-awareness*. New York: Academic Press, 1972.

Fenigstein, A., Scheier, M. F., & Buss, A. H. Public and private self-consciousness: Assessment and theory. *Journal of Consulting and Clinical Psychology*, 1975, *43*, 522–527.

Geller, V., & Shaver, P. Cognitive consequences of self-awareness. *Journal of Experimental Social Psychology*, 1976, *12*, 99–108.

Keppel, G. *Design and analysis: A researcher's handbook*. Englewood Cliffs, N.J.: Prentice-Hall, 1973.

Mead, G. H. *Mind, self and society*. Chicago: University of Chicago Press, 1934.

Ross, L., Greene, D., & House, P. The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, 1977, *13*, 279–301.

Rule, B. G., Nesdale, A. R., & Dyck, R. Objective self-awareness and differing standards of aggression. *Representative Research in Social Psychology*, 1975, *6*, 82–88.

Scheier, M. F. Self-awareness, self-consciousness, and angry aggression. *Journal of Personality*, 1976, *44*, 627–644.

Scheier, M. F., Fenigstein, A., & Buss, A. H. Self-awareness and physical aggression. *Journal of Experimental Social Psychology*, 1974, *10*, 264–273.

Wicklund, R. A. Objective self-awareness. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 8). New York: Academic Press, 1975.

Wicklund, R. A. Group contact and self-focused attention. In P. Paulus (Ed.), *The psychology of group influence*. Hillsdale, N.J.: Erlbaum, in press.

Received May 11, 1978 ■

Ad Hoc Reviewers

The Editor acknowledges with thanks the advice and consultation provided by the following persons, who served recently as ad hoc reviewers of manuscripts. (The number of manuscripts reviewed is in parentheses, if more than one.)

Lyn Abramson
Robert S. Baron
Michael Birnbaum
Leonard Brickman
Timothy Brock
Jerome M. Chertkoff
Richard Crandall
Faye Crosby
John Darley
Deborah Davis
James H. Davis
William deJong
Morton Deutsch
Robert Eisenberger
James Esser
Jeffrey D. Fisher
Howard Friedman
Jacquelyn Gaebelein
Martin Gold
Lewis Goldberg
Barbara E. Hanusa
Alexander Heingartner
Edward Helmes
Martin Hoffman
Stan Kaplowitz
Robert Kidd
Samuel Komorita
Vladimir Konecni
Michael Kuhlman
John Lacy
Ellen Langer
Melvin J. Lerner
John Loehlin
Margaret Matlin
Albert Mehrabian
Dale Miller
Richard Moreland
Teruko Morton
Paul Obrist
Mark Pancer
Martin Patcher
Albert Pepitone
Lawrence Pervin
E. Jerry Phares
Bertram Raven
William Ray
Harvey Reed (2)
Timothy Rogers
Julian Rotter
Susan Saegert
Glenn Sanders
Shalom Schwartz
Steven J. Sherman
Marybeth Shinn
Paul Sticha
Richard Sorrentino
Stuart Taylor
Belinda Tucker
Bernard Weiner
Mark Zanna
Miron Zuckerman

Journal of Personality and Social Psychology
1979, Vol. 37, No. 3, 424-432

Inferior Performance as a Selective Response to Expectancy: Taking a Dive to Make a Point

Roy F. Baumeister, Joel Cooper, and Bryan A. Skib
Princeton University

An experiment investigated factors that determine whether persons conform to what is expected of them or not. Female subjects took a bogus personality test that established either publicly or privately that they had a particular (fictional) trait, which was presented either as a very desirable or a very undesirable trait. Subjects then attempted an anagram-solving task after being informed that persons with this trait usually perform poorly on such tasks. Subjects for whom the expectancy was presented as derived from the desirable trait, when it was publicly established that they had this trait, solved significantly fewer anagrams than subjects in the other three conditions, implying that self-presentational concerns are important in determining the extent to which an expectancy influences behavior. Moreover, analyses of the differences between the number of anagram solutions subjects reported during the experiment and the number of actually correct solutions suggested that when the expectancy was linked to an undesirable trait, subjects actively sought to disconfirm it.

Past research on expectancies has explored whether expectancies increase the likelihood of the expected behavior (e.g., Rosenthal & Fode, 1963). In particular, studies dealing with task performance have sought to determine whether persons will tend to adjust their performance levels to agree with whatever level they have been led to expect. The main emphasis of this research has been to determine whether the subject's cognitions of task and expectancy seem to generate some internal processes (such as drives toward cognitive consistency) that induce him or her to perform at the expected level. Thus, Aronson and Carlsmith (1962) proposed that if one expects poor performance, one will avoid good performance because the latter would cause "cognitive dissonance" (Festinger, 1957). Subsequent research, however, has not consistently found this "rejection of success" nor established what cognitions or situational factors produce it (Brock, Edelman, Edwards, & Schuck, 1965; Cottrell, 1965; Lowin & Epstein, 1965; Marecek & Mettee, 1972; Mettee, 1971; Ward & Sandvold, 1963).

Archibald (1974) reviewed the literature dealing with such "self-fulfilling prophecies," that is, with events that tend to occur *because* someone expects them. Archibald raised several unresolved questions regarding the nature of such expectancy effects and closed with a vigorous call for further research. Specifically, he proposed several possible internal cognitive dynamics (such as dissonance, anxiety, and defensiveness) that might account for the self-fulfilling prophecy effects.

The present investigation was motivated by our contention that the effects of an expectancy on behavior are *not* necessarily or exclusively a product of the internal dynamics of the individual—especially when the expectancy originates (as it usually does) outside the individual. Rather, the desire to convey a particular impression of oneself to others in the situation may in some cases be a more important determinant of the individual's response to an expectancy than the individual's private drives toward maintaining a certain self-concept or achieving cogni-

Requests for reprints should be sent to Joel Cooper, Department of Psychology, Princeton University, Princeton, New Jersey 08540.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3703-0424$00.75

tive equilibrium (Tedeschi, Schlenker, & Bonoma, 1971).

The Public Context: Self-Presentation

It is generally assumed that persons try to maintain favorable evaluations of themselves and that one basic criterion of self-esteem is how successfully a person feels he or she has managed to live up to an "ideal self" (e.g., Cohen, 1959). Separate from one's desire to *be* the ideal self, however, is a motivation to convince others that one's personality resembles one's ideal self. Schlenker (1975) has provided evidence that the motivation to present a highly favorable image of oneself is *not* dependent on having such a favorable self-image. Our point is that behavior in response to expectancies furnished by others [1] is more likely to be motivated by the latter (communicative or self-presentational) interest than by internal goals of protecting one's self-esteem or achieving cognitive balance.

Our investigation therefore emphasized the effect of the context of an expectancy rather than a self-fulfilling effect of the expectancy itself. Depending on the context, the same expectancy could conceivably elicit different behaviors. This brings up the distinction between the *content* and the *basis* of an expectancy. The content is what the target person is expected to do. The basis is whatever attributional reasoning led to the formation of the expectancy. For example, if a man is believed to be a lazy person, one might expect him to spend his spare time watching television rather than playing tennis. The content of this expectancy is the behavioral preference for watching television; the basis is his laziness.

It is plausible that the target's response to an expectancy will depend not only on how he evaluates the expected behavior but also on how he feels about the supposedly underlying trait that is attributed to him. One reason for a target person's sensitivity to the observer's *basis* for expectancy may be that it changes the significance of the target's eventual behavior. The target person must take into account that for him to conform to the behavioral expectancy will confirm the observer's attribution to him of the supposedly underlying trait. Even if the target person believes there is no relation between the trait and the behavior, even if he would deny that he has that trait, he must acknowledge that the observer will interpret the expected behavior in relation to the attributed trait. After all, if the target person (for whatever reason) behaves consistently with the observer's expectancy, the observer will in all probability feel simply that the hypothesis has been confirmed.

Predictions

The present experiment was designed to investigate whether differences in responses to a given expectancy as a function of its context are due to public, self-presentational concerns or to private, internal motivations. Our subjects were informed that they were expected to perform poorly based on a trait they (supposedly) had, which was presented either as a desirable or an undesirable trait. Half our subjects were led to believe that they alone knew that they had the underlying trait and were therefore expected to do poorly, the remainder of the subjects were identified publicly as having the (good or bad) trait that predicted poor performance. We predicted that the expectancy would successfully elicit poor performance only when the expectancy was public and only when it was publicly linked to the desirable trait. However, a viewpoint based on the belief in the internal dynamics of the individual as the essential or exclusive mediators of expectancy effects would predict that the results would be the same in the public as in the private treatment conditions.

Relation to Past Work

Although previous research has largely ignored the issue of how responses to expectancies are determined by the public or private

[1] We are not dealing with expectancies that originate within the actor. Such expectancies may well operate independently of public or self-presentational interests.

nature of their contexts, various findings can be interpreted as suggesting a connection between self-presentational concerns and the beliefs and expectancies of others. For example, Zanna and Pack (1975) found that women will present themselves consistently with the sex role attitudes of an attractive male. Similarly, Deutsch and Gerard (1955) demonstrated that conformity to an obviously false majority judgment is greatly reduced if subjects express their judgments anonymously.

A study by Sigall, Aronson, and Van Hoose (1970) is relevant to the present discussion. Their subjects copied lists of telephone numbers after being told that they were expected to perform at a level either below or above that of the subjects' own "practice round" level. In a third condition, subjects were given the expectancy of performance lower than their own *and* told that subjects who performed well on this task had obsessive-compulsive personalities. Sigall et al. found that this third condition produced poorer performance (i.e., fewer numbers copied) than a no-expectancy control condition, whereas the first two conditions elicited superior performance.

However, Sigall et al. were primarily interested in a methodological issue, and their results (although suggestive) do not provide conclusive evidence about reactions to expectancies. In the first place, subjects performed better than controls after an expected level of performance was communicated to them. But it made no difference whether this expected performance level was set above *or* below the subject's pretest level. Both produced improved performance. Sigall et al. suggest the possibility that subjects did not perceive the expectancy in relation to their pretest performances. Therefore, they may not have known whether an increase or a decrease was expected.

In fact, Sigall et al. (1970) reported a second experiment in which the expectancy manipulation was omitted altogether, and improved performance was found as a result of simply telling the subject more about the study (and informing the subject that a hypothesis existed, without telling the subject what it might be). These results therefore seem to imply only that experimental subjects work harder if the experimenter makes a greater effort to explain the purpose and direction of the research than if the experimenter simply tells them what to do. Although this is an interesting finding, it does not have a direct bearing on how persons are affected by what others expect them to do.

The results of the obsessive-compulsive treatment also can be understood without reference to the expectancy about the subject's performance level. Subjects were told that only obsessive-compulsive personalities performed well on that task. The task was thus presented as the only indication of whether the subject was an obsessive-compulsive. Rather than presenting the subject with an expectancy based on an already made inference about his or her personality, the subject was presented with this task as a means of determining whether the subject was an obsessive-compulsive or not.

Thus, although the procedures and results of Sigall et al. raise a number of issues germane to the present investigation, they cannot be considered to provide clear or definitive answers to these questions. Of course, the most important difference between the procedures of Sigall et al. and ours is that their work provides no indication as to whether any such motivation toward inferior performance derives essentially from a private desire of the individual to protect self-esteem or from self-presentational interests. The present investigation was directly concerned with this issue.

Method

Subjects

Forty-one undergraduate women volunteered for a study on "personality and behavior." Data from one subject were discarded because she inadvertently undermined a manipulation by announcing her pretest results aloud when they were supposed to be known only to her.

Subjects participated in the study individually, although a male confederate, who pretended to be another subject, was present for all sessions. All subjects were offered and paid $2.50 for their participation. They were assigned at random among the four treatment conditions.

Female subjects were used exclusively because of availability and to ensure a reasonably homogeneous sample. The cover story and operations were adapted so as to be suitable for an exclusively female sample.

Procedure

Pretest. When the subject arrived at the experimental room, the male confederate was not yet present. The female experimenter[2] said that the first procedure involved a personality inventory that was designed for the female subjects. Since the other participant scheduled for that session was male, the present subject might as well begin the personality measure without waiting for him to arrive.

The pretest was in fact bogus. Its only function was to provide a vehicle for convincing the subject that she scored high on the bogus personality trait from which the behavioral expectancy would later be derived.

In order to facilitate the confidentiality of the subject's score and in order to avoid suspicion concerning the results of the test, subjects scored their own tests. The tests were structured so that all subjects would score high on the trait in question.

The male confederate waited until the female subject was nearly finished taking her test. He entered with a confused apology for being late. The experimenter instructed him to sit across a large table from the subject. A large barrier precluded visual contact.

Manipulation of trait desirability. The subject heard one of two tape-recorded briefings. Both began by informing subjects that the preceding personality test was a measure of the trait of "surgency." Our use of the term *surgency* followed Ickes, Wicklund, and Ferris (1973), who suggested that the term is well suited to denote a bogus personality trait because it sounds like authentic psychological jargon but does not have a meaning that is familiar to most subjects.

Both initial briefings stated that it was expected that many Princeton women would be high in surgency because of their unusual developmental backgrounds. The importance of the research on surgency was stressed. Moreover, the recorded voice said that even if the subject did not score high on surgency, her participation was valuable in order to furnish a comparison with the high-surgency group. In addition, a male control group was to be included.

There followed the explication of surgency. In the "good" condition, the recording protrayed high-surgency women as unusually mature and well-adjusted, sensitive, intelligent, and insightful. A speculative developmental scenario suggested that the introduction of some aspects of traditionally male socialization resulted in an optimal "best of both worlds" personality development in the high-surgency woman.

In the "bad" condition, high-surgency women were depicted as immature and ill-adjusted, shallow, insensitive, defensive, and insecure. The implication of the developmental argument was that exposing the young female to some aspects of traditionally male socialization made her insecure about her femininity.

Manipulation of public/private. At this point, the subject was asked to score her own answer sheet for surgency. She was given a scoring template and a red pen and instructed to mark it. In the "public" condition, the experimenter asked the subject aloud what her score was, stated publicly, "That is above 13, so you are 'high' in surgency," and made a show of recording the subject's name, score, and "high" in a data record book. Although the confederate made no comment, it was obvious that he heard all the proceedings. The experimenter then collected the subject's answer sheet, after making certain that the subject had written her name on it.

In the "private" condition, the subject was instructed not to write her name on the answer sheet (or on any of the experimental materials). She was assigned an arbitrary three-digit "subject number" by which her data would later be assembled. When it was time for her to score her answer sheet, the experimenter gave her the template and explained how to score it and how to determine if she was in the high-surgency group. The experimenter then left the subject alone to do the scoring and recording, after saying that she (the experimenter) was not supposed to know the subject's score. When the subject was finished, she folded her answer sheet and stuffed it into a box full of other answer sheets, closed the data book, and returned the materials to the experimenter.

The criterion "13 high/12 low" was written on the blackboard and was thus apparent to all subjects, although until they scored the tests it presumably had no significance for them.

Anagram task. The tape-recorded instructions announced that the next procedure would involve solving anagrams. A brief statement of the rationale for including this measure in the experiment contained the critical prediction that high-surgency women were expected to perform poorly on the anagram task.

In the "good" condition, it was explained that anagram solving is a simple, mechanistic cognitive task and that because the talents of the high-surgency woman are geared toward tasks involving meaning, creative tasks, and tasks involving insight into and understanding of other persons, she would not perform well on exercises like anagram solving. In the "bad" condition, it was asserted that the defensive vigilance of the high-surgency woman makes her reluctant to immerse herself in a simple, mechanistic cognitive task, because she prefers to keep her attention on interpersonal matters "where the source of her insecurities lies."

---

[2] The authors wish to thank Elaine Seligman for her very capable performance as the experimenter.

Table 1
Mean Number of Anagrams Reported and Actually Solved

Condition	Trait	
	Bad	Good
Private		
Reported	21.5	23.2
Actual	20.3	22.8
Public		
Reported	22.7	14.4
Actual	21.8	14.3

Note. $n = 10$ per cell; maximum possible number of anagrams = 40.

Following the prediction of poor performance among high-surgency women, the tape recording said that performance among low-surgency women was "considerably less uniform." Low-surgency women might therefore do well or poorly. Specific instructions for the task were then given: It would involve two 8-minute sessions, each with a list of 20 anagrams; all sets of letters could be unscrambled to make English words (this was true); the subject should solve as many as possible in the alloted time.

Anagram performance was to be public for all subjects. The initial instructions said that after each list, the experimenter would ask each subject how many anagrams he or she had solved. Following the first list, the experimenter asked first the subject and then the confederate how many anagrams (out of 20) they had solved. The confederate invariably reported the same number as the subject had. Following the second list, this was repeated, except that the confederate reported having solved two fewer than the subject. The number of anagrams solved by the subject was the main dependent measure.

Finally, a postexperimental questionnaire contained manipulation checks and solicited reactions to or feelings about the procedures. A careful debriefing followed in which the nature and necessity of all deceptions were fully explained.

Results

Manipulation Checks

An item on the postexperimental questionnaire asked subjects to "Indicate how you felt about the results of the personality test." Four subjects left this item blank. Subjects in the desirable trait condition felt much better about having high surgency than did subjects for whom surgency was presented as an undesirable trait, $F(1, 32) = 6.84$, $p < .02$. Neither the effect for the public/private variable ($F < 1$) nor the interaction ($F = 1.66$) approached significance. A second item asked whether the other subject (i.e., the confederate) had been aware of the results of the subject's personality test. The appropriate answer would of course be "yes" for the public conditions and "no" for the private conditions. All but three subjects answered this item correctly. We did not discard the data from these subjects. However, omission of their data would not reduce the significance levels of the analyses.

Anagram Solutions

The experimenter recorded how many anagrams the subjects reported having solved. Another experimenter later scored all the problem sheets for the actual number of correct solutions. Because there were occasional discrepancies between the number of anagrams actually solved and the number reported aloud, separate analyses were performed on these two measures.

Mean anagram scores for both reported and actual solutions are presented in Table 1. Both analyses of variance revealed significant interactions between the two independent variables: for reported scores, $F(1, 36) = 8.00$, $p < .01$; for actual scores, $F(1, 36) = 7.86$, $p < .01$. In addition, the main effect for public–private was significant, $F(1, 36) = 4.62$, $p < .05$, for the reported scores, but not for the actual scores. It is clear from Table 1, however, that the significant effects are attributable to the disparity between the performance of the good trait/public group and the other three conditions. Using Tukey's honestly significant difference (HSD; see Kirk, 1968) technique, the good trait/public cell mean was found to be significantly lower than each of the other means for the reported scores ($p < .05$) and lower than each of the others except the bad trait/private cell mean for the actual scores. The difference between the good trait/public and the bad trait/private cell means for the actual scores was, however, significant ($p < .05$) according to Fisher's LSD (least significant difference) technique. No other pairwise difference between means was significant. It seems, thus,

that the good trait/public treatment was much more effective then the others in eliciting behavior consistent with the expectancy.

Discrepancy Between Reported Score and Actual Score

Some subjects reported scores that were different from the actual number of anagrams they had solved correctly. When this difference occurred, the reported scores were invariably greater than the actual correct scores, implying that these were not simply random, honest mistakes. We shall use the term *enhancing* to refer to a subject's reporting a score greater than the correct one. A measure of enhacing was obtained by subtracting the actual number of anagrams each subject solved correctly from the number she reported during the experiment. Analysis of variance on these enhancing scores revealed only one significant effect, and that was the main effect for the desirability of the trait, $F(1, 36) = 6.44$, $p < .02$. Subjects in the bad trait conditions enhanced significantly more than subjects in the good trait conditions.

That this difference was a general trend and not due to a few isolated individuals can be verified by considering the number of subjects who enhanced. Only 2 of the 20 subjects in the good trait conditions enhanced. In the bad trait conditions, however, 12 of the 20 subjects enhanced. The chi-square analysis of this difference is highly significant, $\chi^2(1) = 10.99$, $p < .001$.

Difference Between the Two Anagram Lists

Subjects worked two lists (labeled A and B) of anagrams. All subjects did List A first. The difference between the performance levels on the two lists was not influenced systematically by the independent variables.

However, the patterns of enhancing on the two lists *did* vary among the different conditions. For each subject, the difference between her enhacing on the first (A) test and on the second (B) test was obtained by subtracting the enhancing score (reported minus actual anagram solutions) on the B list from that on the A list. Analysis of variance on these differences showed only one significant effect: an interaction between the public–private and the good–bad trait variables, $F(1, 36) = 4.53$, $p < .05$. Because enhancing was negligible in the good trait conditions, the interaction arises chiefly from the bad trait cells. The bad trait/private condition subjects enhanced more on the A list than on the B list, whereas the bad trait/public condition subjects enhanced more on the B list than on the A list. The difference between the bad trait/private and the bad trait/ public means on this measure is significant ($p < .05$) according to Tukey's HSD technique.

Final Questionnaire

Several questions on the postexperimental questionnaire asked subjects to rate on 5-point scales how hard they had tried to do well on the anagram task and how hard they had tried to conform to the level of performance that was expected of them. No significant differences were found on these measures.

Discussion

Experimental subjects typically perform tasks to do the best of their ability, motivated presumably by implicit demand characteristics, habit, and concern with self-esteem and public esteem. Our subjects systematically violated this practice and performed at a level significantly below that of their capabilities when they were publicly expected to do poorly and this expectation reflected a desirable quality in their personalities.

There is no evidence from our self-report (or other) data of any conscious intention to achieve inferior performance. Still, it seems apparent that a person is at some level sensitive to how others will interpret her behavior and that she will modify her performance level in order to convey the desired message. Because the inferior performance levels were found when the expectation about the subject was public but not when it was private, it is clear that the public nature of the expectation was indispensable in eliciting conformity. We interpret this result as meaning that

modification of performance level was due to essentially communicative interests. It has been shown previously that the presence or orientation of an audience can raise concerns that may interfere with performance (Brown, 1968, 1970; Brown & Garland, 1971; Cottrell, Wack, Sekerak, & Rittle, 1968; Deutsch & Gerard, 1955; Henchy & Glass, 1968). In the present study, the performance was reduced to a mere tool for the communication of information that had no analytical relation to the task itself. We submit that the concerns raised by the audience expectancies in the present study were not merely distractions that interfered by competing with the task for the subject's attention (for such concerns would have been equally if not more potent in the public expectancy/*bad* trait condition) but actually preempted the desire to do well.

Illicit Enhancement

Illicit performance enhancement did seem to reflect some internal motivation, in that it occurred in both the public and the private versions of the situation in which the expectancy of poor performance was linked to an undesirable (bad) trait. When poor performance was expected on the basis of an undesirable trait, our subjects were eager to do well. Although they performed no better than subjects in the private expectancy/good trait condition, they resorted more to illicit techniques such as reporting nonwords as solutions and simply miscalculating their score in their favor. This suggests that they were strongly motivated to achieve high scores, possibly in order to discredit the attribution of the bad trait to them by disconfirming the prediction based on it.

An alterative interpretation of this finding can be derived from the work of Aronson and Mettee (1968). These authors found that subjects were more willing to engage in dishonest behavior after receiving an unfavorable evaluation than after receiving a favorable one. By cheating at a game, subjects could demonstrate that they apparently had extrasensory perception (ESP), a possibly very desirable trait that had not been mentioned in the unfavorable evaluation. The cheating of these subjects may therefore have represented a self-enhancement designed to compensate for the bad evaluation (Baumeister & Jones, 1978). It is possible that if one knows that one is evaluated unfavorably by others, this knowledge effectively reduces the inhibitions against using illicit means in order to look good. The illicit performance enhancement of our subjects following the bad trait treatment can be seen as an instance of this possibility.

Moreover, there was a difference in the patterns of illicit performance enhancement between the public and the private versions of the condition in which the expectancy was based on an undesirable trait. When the expectancy was private, subjects resorted to illicit enhancement more on the *first* of the two anagram tasks, but when it was public, they did so more on the *second* test. The only difference in the situation between the two tests was that after the first test, the subject and the confederate reported a score equal to that of the subject. Thus, if the subject enhanced her performance on the first list and not on the second list, she was apparently content with performing just as well as the confederate; whereas if the subject enhanced her score more on the second test, she was apparently not content with matching the confederate but wanted to do better than he.

We did not predict this finding, but we venture a post hoc speculative interpretation. In the private condition, the subject's internal desire to discredit the notion that she had high surgency was satisfied by doing as well as the confederate. In the public condition, however, she had the additional concern of discrediting the attribution (of high surgency to her) in the eyes of the others present, and she could not assume that they would be as easily convinced as she herself might be. In other words, she could not be certain that the others would consider the expectancy of poor performance disconfirmed if she merely performed at the same level as the confederate; therefore, after learning that her score on the first test equaled his, she was tempted to resort to illicit enhancement on the second test in the hope of exceeding his score.

Subjects in the private expectancy/good trait condition showed no sign of trying either to raise or to lower their performance level from the level appropriate to their ability. There is no reason to infer that they endowed the task with any surplus meaning. That is, there is no evidence that they believed poor anagram performance would convince the audience that they had high surgency. In this respect they were unlike the subjects in the other three conditions.

Origin of Expectancy

A possible boundary condition of the present findings is implied by the fact that the subject herself provided the basis for the expectancy by her own behavior (that is, the expectancy was based on her test responses). The findings of Gurwitz and Topol (1978), although not based on behavioral measures, suggest a relation between whether a person has given evidence in support of an expectancy and how the person will subsequently intend to behave vis-à-vis that expectancy.

Because our subjects were women, our findings might be considered in the context of literature that has suggested that women perform below the level of their abilities because of a "fear of success" (Horner, 1972). Our study was not designed to address this issue. Still, it might be noted that in our sample, any motivation to perform below the level of one's ability was derived from immediate *public* expectancies and self-presentational interests rather than from internalized stereotypes or dispositional fears of success.

Conclusions

The present findings do *not* mean simply that women perform poorly when bad performance means something good. Rather, our evidence indicates that an expectancy exerts effective power over a woman [3] toward eliciting behavior confirming the expectancy *if* the following two conditions are satisfied: The expectancy is based on a desirable rather than an undesirable aspect of the target's personality, and other persons present in the situation are aware that this trait (on which the expectancy is based) is attributed to the target person. We do not deny that internal expectancies can affect behavior. Nonetheless, our findings imply that in order to understand responses to overt public expectancies, one must take into account not only the internal cognitive dynamics of the individual but also the individual's self-presentational interests. The influence of expectancy on behavior is in part determined by the expectancy's derivation and context, which give extra meaning to the behavior in the face of the expectancy and which therefore affect the self-presentational strategy.

---

[3] Our conclusions are stated in reference to women because our experimental sample consisted of women. We do not, however, wish to imply that our hypotheses would not be true for men.

References

Archibald, W. P. Alternative explanations for self-fulfilling prophecy. *Psychological Bulletin*, 1974, *81*, 74–84.

Aronson, E., & Carlsmith, J. Performance expectancy as a determinant of actual performance. *Journal of Abnormal and Social Psychology*, 1962, *65*, 178–182.

Aronson, E., & Mettee, D. Dishonest behavior as a function of differential levels of induced self-esteem. *Journal of Personality and Social Psychology*, 1968, *9*, 121–127.

Baumeister, R., & Jones, E. E. When self-presentation is constrained by the target's prior knowledge: Consistency and compensation. *Journal of Personality and Social Psychology*, 1978, *36*, 608–618.

Brock, T., Edelman, S., Edwards, D., & Schuck, J. Seven studies of performance expectancy as a determinant of actual performance. *Journal of Experimental Social Psychology*, 1965, *1*, 295–310.

Brown, B. The effects of need to maintain face on interpersonal bargaining. *Journal of Experimental Social Psychology*, 1968, *4*, 107–122.

Brown, B. Face-saving following experimentally induced embarrassment. *Journal of Experimental Social Psychology*, 1970, *6*, 255–271.

Brown, B., & Garland, H. The effects of incompetency, audience acquaintanceship, and anticipated evaluative feedback on face-saving behavior. *Journal of Experimental Social Psychology*, 1971, *7*, 490–502.

Cohen, A. Some implications of self-esteem for social influence. In C. Hovland & I. Janis (Eds.), *Personality and persuasibility*. New Haven, Conn: Yale University Press, 1959.

Cottrell, N. Performance expectancy as a determinant of actual performance: A replication with a new design. *Journal of Personality and Social Psychology*, 1965, *2*, 685–691.

Cottrell, N., Wack, D., Sekerak, G., & Rittle, R. Social facilitation of dominant responses by the presence of an audience and the mere presence of others. *Journal of Personality and Social Psychology*, 1968, *9*, 245–250.

Deutsch, M., & Gerard, H. A study of normative informational social influences upon individual judgment. *Journal of Abnormal and Social Psychology*, 1955, *51*, 629–636.

Festinger, L. *A theory of cognitive dissonance*. Stanford, Calif.: Stanford University Press, 1957.

Gurwitz, S., & Topol, B. Determinants of confirming and disconfirming responses to negative social labels. *Journal of Experimental Social Psychology*, 1978, *14*, 31–42.

Henchy, T., & Glass, D. Evaluation apprehension and the social facilitation of dominant subordinate responses. *Journal of Personality and Social Psychology*, 1968, *10*, 446–454.

Horner, M. Toward an understanding of achievement-related conflicts in women. *Journal of Social Issues*, 1972, *28*, 157–176.

Ickes, W., Wicklund, R., & Ferris, C. Objective self-awareness and self-esteem. *Journal of Experimental Social Psychology*, 1973, *9*, 202–219.

Kirk, R. *Experimental design: Procedures for the behavioral sciences*. Belmont, Calif.: Wadsworth, 1968.

Lowin, A., & Epstein, G. Does expectancy determine performance? *Journal of Experimental Social Psychology*, 1965, *1*, 248–255.

Marecek, J., & Mettee, D. Avoidance of continued success as a function of self-esteem, level of esteem certainty, and responsibility for success. *Journal of Personality and Social Psychology*, 1972, *22*, 98–107.

Mettee, D. Rejection of unexpected success as a function of the negative consequences of accepting success. *Journal of Personality and Social Psychology*, 1971, *17*, 332–341.

Rosenthal, R., & Fode, K. Three experiments in experimenter bias. *Psychological Reports*, 1963, *12*, 491–511.

Schlenker, B. Self-presentation: Managing the impression of consistency when reality interferes with self-enhancement. *Journal of Personality and Social Psychology*, 1975, *32*, 1030–1037.

Sigall, H., Aronson, E., & Van Hoose, T. The cooperative subject: Myth or reality? *Journal of Experimental Social Psychology*, 1970, *6*, 1–10.

Tedeschi, J., Schlenker, B., & Bonoma, T. Cognitive dissonance: Private ratiocination or public spectacle? *American Psychologist*, 1971, *26*, 685–695.

Ward, W., & Sandvold, K. Performance expectancy as a determinant of actual performance: A partial replication. *Journal of Abnormal and Social Psychology*, 1963, *67*, 293–295.

Zanna, M., & Pack, S. On the self-fulfilling nature of apparent sex differences in behavior. *Journal of Experimental Social Psychology*, 1975, *11*, 583–591.

Received May 15, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 3, 433-446

Sex Differences in the Allocation of Pay

Charlene M. Callahan-Levy
New College, University of South Florida

Lawrence A. Messé
Michigan State University

The present research examined the possibility that relative to males, females perceive less of a connection between their work and monetary rewards. In Study 1, females and males determined either their own pay or the pay of another person. The results supported the hypothesis, in that (a) females paid themselves less than did males, and (b) females paid themselves less than other people (males or females) paid females. Moreover, the results suggested that subjects were more generous when they paid females. In Study 2, sex differences in self-pay were examined in samples of first-, fourth-, seventh-, and 10th-grade subjects. Results replicated those of Study 1, in that at every grade level, females paid themselves less than did males. In addition, the extent to which females allocated pay the way their male counterparts did was found to be highly related to the masculinity of their career goals. Some implications of these results for the understanding of inequitable consequences of traditional sex roles are discussed.

One of the more consistent findings generated by studies of reward allocation behavior is that American female allocators pay themselves less than do males (Lane & Messé, 1971; Leventhal & Lane, 1970; Mikula, 1974; Messé & Lichtman, Note 1). While this sex difference has been demonstrated a number of times, there have been few empirical investigations of the numerous speculations that have been offered to account for this phenomenon (e.g., Leventhal, Note 2). Thus, the present research examined one possible explanation, derived from the position of Mednick and Tangri (1972) and Chesler and Goodman (1976), that relative to males, females perceive that money is less salient a reward for the work that they perform.

Mednick and Tangri (1972) and Chesler and Goodman (1976) speculated that females learn that performances of the sex role-related tasks traditionally assigned to them (e.g., mothering, housekeeping, etc.) usually are not rewarded monetarily in any direct way. It is possible that the lack of a relationship between money and the traditional tasks for women generalizes to some extent to other work situations as well. Moreover, this perception may be reinforced indirectly by sociocultural biases about females' relative "needs" for monetary rewards. For example, differences in pay as a function of sex often have been "justified" by assertions that it is acceptable for women to earn less than men because women usually do not provide the sole means of economic support for their families. Males, on the other hand, might learn to distinguish between family-oriented tasks, for which no money is expected, and "work," which they perform for pay.

Males also may come to see money not only as a means to acquire goods and services but also as a symbol of how their work is evaluated and, thus, as a measure of their worth. Females, however, may not perceive

This research was supported by National Institute of Mental Health Grant MH24550, "Sex Differences in the Distribution of Rewards," and the Computer Institute for Social Science Research, Michigan State University. Study 2 was a modification of the first author's doctoral dissertation, submitted to the psychology department at Michigan State University. The authors are grateful to Michael Albrecht, Larry Bozyk, Sharon Daugharty, Larry Hall, Cheryl Isaac, Dave Kritt, Deborah Levine, John MacFarlane, Sandy Pritchard, and Deborah Shulman for their assistance in data collection.

Requests for reprints should be be sent to Lawrence A Messé, Department of Psychology, Michigan State University, East Lansing, Michigan 48824.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3703-0433$00.75

their pay to have this evaluative component, since the traditional tasks at which women are "supposed to do well" and to which their self-concepts are supposed to be tied (e.g., mothering, cleaning, cooking, etc.) are not "jobs" for which pay is expected. Chesler and Goodman (1976) speculate that women might use money as an index of self-worth to the extent that they can acquire it without working for it in the traditional sense, that is, when it is "magic money."

Given that females might have internalized less of a connection between work and pay and between pay and self-esteem, it follows that females' views of money as pay would be influenced by a number of factors that males would perceive as less relevant. For example, females could be concerned with appearing "nice," "not greedy," or "not concerned with money," and so forth, whereas males could be concerned almost entirely with getting paid at least what they deserve.

This article reports two studies that examined the implications of this argument. In Study 1, we explored the extent to which sex differences in reward distribution found in past studies were due, at least in part, to the tendency of females to pay themselves less than males pay themselves. In Study 2, we attempted to replicate some of the major findings of Study 1 and to extend them by examining self-pay allocation in other subject populations and by relating this behavior more directly to sex role socialization.

Study 1

The "less-of-a-connection" hypothesis can be viewed within the perspective of equity theory (Berkowitz & Walster, 1976). In one version of the theory (e.g., Lane & Messé, 1972; Weick, 1966), persons are assumed to have, among other criteria of equity, an internal standard by which they judge the adequacy of the compensation that is available for a given level of work inputs. This standard can be applied both to one's own rewards ("own equity") and the rewards of other persons ("other equity"). It could be that males are at least as likely to apply this internal standard of fair pay to outcomes that are potentially available for their own inputs as they are to apply it to the potential outcomes of others. Females, on the other hand, could be more likely to apply the internal standard to the outcomes of others than they are to their own outcomes. In Weick's (1966) terms, the more tenuous connection between work and pay that is hypothesized to be found in females could be manifested in a weaker sense of own equity. Further, as discussed above, this hypothesized weaker sense of own equity would make females more susceptible to the influence of such factors as concerns with impression management, which would tend to affect adversely their judgment about what is appropriate pay for themselves.

An alternative interpretation, again cast within the framework of equity theory, is that persons, both males and females, apply different standards of fairness to the pay of females and males. Thus, it could be that both males and females believe that women should be paid less than men for a given amount of work. This difference in the perception of what is fair pay would mean that the "double standard" of different pay for the same work by males and females would have the status of a general norm.

Both explanations involve a sex difference in the connection between work and pay, and both would account, at least in part, for the reward allocation findings of past research. In these studies (e.g., Leventhal & Lane, 1970; Mikula, 1974; Messé & Lichtman, Note 1), allocators typically were asked to divide anonymously a monetary reward between themselves and another person. In this zero-sum situation, it could be that females gave the other person more in part because they conformed to a norm that specified that women (including themselves) deserved less. On the other hand, it also could be that the female subjects in these studies allocated relatively more reward to the other person in part because they might have perceived less of a connection between their own work and the money that they could give themselves, and thus, they could have been more susceptible to being influenced by concerns with the other person's welfare, by the norm of equality, and so forth.

However, the procedure that typically was used in these past studies—where subjects divided a reward between themselves and another person—does not separate the impact on allocation behavior of people's internal standards from the impact of their interpersonal concerns. The present study, on the other hand, used a procedure that reduced the salience of interpersonal concerns and thereby permitted a more direct exploration of the hypothesis that females perceive less money as an appropriate reward for themselves than do males. In addition, it also permitted a direct examination of the extent to which the tendency in females to pay themselves less is a result of a weaker sense of their own equity or of a general norm that persons apply to all women.

In the present research, female and male subjects were asked to pay either themselves *or* another female or male for a given period of work. If females do, in fact, perceive less of a connection between work and money and, as a result, are more influenced by additional concerns, they should pay themselves less than do males. Moreover, given that the hypothesized less-of-a-connection is reflected in a weaker sense of own equity, then females should pay themselves less than both females and males pay other females. However, if the sex difference in reward allocation is based on a general norm, then females should pay themselves less, and both females and males should pay females less than they pay males.

Method

Subjects and Recruitment

One hundred twenty-six undergraduates, 66 females and 60 males, served as subjects in this experiment. These persons were chosen by chance from approximately 400 respondents to an advertisement in the Michigan State University student newspaper that solicited subjects for pay.

Design

Three dichotomous independent variables were examined via a factorial design. The variables were (a) the target of the allocation (the allocator or another person), (b) the sex of the target person, and (c) (included as a control to complete the design) the sex of the nontarget person (same as or different from the target). When the subject was choosing pay for herself (himself), the sex of the nontarget merely referred to the sex of the people in the other group (see below). When the subject was paying the other person, the sex of the nontarget referred to the sex of the allocator. Viewed from a simpler, more straightforward perspective, the design involved the sex of the allocator, the sex of the people in the other group, and the target of the allocation (the allocator or a person in the other group). However, the comparisons of greatest interest, which were subsumed under the Target × Sex of Target interaction, could be examined most directly when the design was cast in the more complicated format.

Procedure

Subjects were examined in 16 groups of six or eight persons.[1] Four of the groups were composed entirely of females, four entirely of males, and the remainder were composed so that half of the participants were male and half were female. In two same-sex groups and four mixed-sex groups, each subject allocated money only to herself (himself) after she (he) had completed a pretask; in the remaining groups, half the subjects each allocated pay to another person, who also had completed the pretask but who did not allocate payment.

When subjects arrived for the experiment, they were divided between two large tables in a rather spacious room. In mixed-sex groups, females were always assigned to one table and males to the other. Two experimenters, one male and one female, handed out the pretask materials and explained what the subjects were to do. They always started by explaining that the subjects were participating in a study that attempted to simulate an industrial setting. This rationale has been used a number of times in past reward distribution studies (e.g., Leventhal & Michaels, 1969).

The pretask consisted of a number of essay questions pertaining to relevant campus-related issues.[2] Subjects were asked to spend approximately 5 minutes writing their own opinions on each question or topic. It was stressed that this task was nonevaluative, in the sense that there were no right or wrong answers to the questions.

---

[1] It should be noted that a group of subjects performed only the initial task in the same room. As explained in detail below, subjects always allocated pay and completed the postallocation questionnaire in separate cubicles.

[2] This task appears to be an appropriate choice for a nonevaluative work input, given that 223 males and females, when asked to work on and evaluate it, generated no sex differences in their perception of their absolute or relative enjoyment, skill, effort, and quality of performance; multivariate $F(8, 214) = .90$, $p < .5$.

At the end of 50 minutes, the experimenters collected the finished questionnaires; then they told the subjects that the remainder of the session would be devoted to paying them. In the self-allocation conditions, each subject was led to a cubicle, where she or he was given $6 in bills and change, two envelopes—one marked "My pay" and the other "Remainder"—and a short questionnaire that asked (a) the amount of money allocated, (b) the reason, in the subject's own words, why this amount was chosen, (c) the amount of money (up to $6) that the subject thought would be fair pay for this work, and (d) the sex of the person in the "other group." The experimenter explained that each subject should pay herself (himself) any amount, up to the maximum of $6, by placing the desired amount in the "My pay" envelope. This amount was hers (his) to keep. The subject was to take the remainder and place it in the appropriately marked second envelope. He (she) was informed that the amount that was left over would be returned to the general research fund, and thus, his (her) allocation would not affect the payment of any other person. After allocating the money, she (he) was then to complete the questionnaire, fold it, place it in the "Remainder" envelope, seal the envelope, and deposit it in a box that was placed at the exit to the room.

In the other-allocation conditions, only half the subjects (always sitting at the same table) were led to the cubicles, while the remainder stayed at their original places at their table. Once in the cubicles, subjects were told that they were to allocate up to $6 to pay one of the other subjects who was left in the original room. The specific person who would be the recipient of the allocation was to be selected by chance, and her (his) identity would never be disclosed to the allocator, nor vice versa. (However, in a matter-of-fact manner the experimenter stated verbally, and the written instructions indicated, whether the recipient was female or male.) Again, each subject in the cubicles was handed two envelopes—this time, however, one was marked "Her (or His) pay" and the other was marked "Remainder"—and a similar short questionnaire. The experimenter stressed that the allocator's behavior in no way influenced his or her own pay, which had been determined in advance but would not be revealed or dispensed until this task was completed. The procedure for allocating pay to the other person was essentially the same as that used in the self-pay conditions, except that the allocator placed both envelopes in the box before she (he) was given a sealed envelope containing $2.50. The experimenter did, in fact, take the pay envelopes and distribute them at random to the subjects who were waiting in the original room.

Results

There were two payment scores obtained from each allocator. One score was the amount of money that subjects (on the postallocation questionnaire) reported giving to the target of the allocation; the other was the amount of money that the allocator indicated was "fair" pay.[3] These scores were subjected to separate 2 (target: allocator or another person) × 2 (sex of target) × 2 (sex of nontarget) unweighted-means analyses of variance. The results of these analyses are reported below in terms of their implications for the hypotheses.[4]

Both the less-of-a-connection hypothesis and the double standard hypothesis predict that females should underpay themselves relative to males. In addition, however, females should pay themselves less than other people pay females if the lesser connection is manifested in a weaker sense of own equity, whereas both males and females should pay other females less than they pay males if there exists a general, lower standard of fair pay for women. Thus, we expected either a significant Target × Sex of Target interaction (if the less-of-a-connection was expressed in weaker own equity) or a significant sex of target main effect (if the sex difference in pay was a reflection of different standards). The Target × Sex of Target interaction was significant in both analyses, $Fs(1, 66) = 12.50$ and 11.35, $ps < .001$, for actual and

[3] As noted above, an item on the postallocation questionnaire asked subjects to indicate the sex of the other person. Since this variable was part of the design, and it was crucial when subjects allocated pay to the other, the data of nine persons who answered this question incorrectly were excluded from the analysis. In addition, the data from two other subjects were not used: One subject, on his postquestionnaire, indicated that he was told when he had been telephoned and asked to participate that he was to be paid at a rate of $2/hour, which, in fact is what he paid himself. This occurred only once, since he was one of the first subjects that we examined. He was told this in answer to a question, and we took steps to insure that the assistants who were in charge of scheduling answered such questions more appropriately in the future. A second subject gave very bizarre reasons for his allocation (he paid a female $6).

[4] It is important to note, though, that as expected, the sex of the nontarget had little systematic effect on allocations, since no F ratio involving this variable even approached significance. The means of these F ratios were .940 and .739, for actual and fair pay, respectively.

fair pay, respectively. Table 1, which presents the relevant cell means, indicates that the direction of this effect supports the position that the more tenuous connection between work and pay for women is reflected in a weaker sense of own equity.

Inspection of Table 1 and appropriate individual comparisons (Winer, 1971, p. 384) indicate that females tended to pay themselves less than males paid themselves, $t(66) = 1.89$, $p < .03$ (actual pay), and $t(66) = 3.50$, $p < .001$ (fair pay), and less than other people paid females, $t(66) = 3.86$, $p < .001$ (actual pay), and $t(66) = 2.87$, $p < .005$ (fair pay).

Since the weaker own equity interpretation posits that females will pay themselves less than they pay other females, it seemed reasonable, given the results presented above, to perform individual comparisons of these means. These comparisons provided further support for the own equity explanation, since females were found to pay themselves (see Table 1) less than female allocators paid other females: $M = \$5.19$, $t(66) = 2.86$, $p < .005$, for actual pay; and $M = \$3.77$, $t(66) = 2.36$, $p < .02$, for fair pay. Thus, our evidence indicates that females pay themselves less than males, and this sex difference appears due to a weaker sense of own equity in women.

Table 1
Mean Payment Allocated to Females and Males

Sex of target	Target of allocation: Self[a] Amount (in $)	*n*	Other[b] Amount (in $)	*n*
Actual pay allocation				
Female	3.45	22	5.11	17
Male	4.26	17	3.42	18
"Fair" pay allocation				
Female	2.49	22	3.62	17
Male	3.88	17	3.16	18

[a] Values presented in this column are the average amounts that the subjects of each sex allocated to themselves.
[b] Values presented in this column are the average amounts that subjects (both males and females) allocated to other females and males.

In addition, the simple effects analysis of the Target × Sex of Target interaction indicated that allocators tended to pay other males less than they did other females, $F(1, 66) = 9.62$, $p < .005$ (actual pay); $F(1, 66) = 2.98$, $p < .10$ (fair pay).

Although (as noted above) data collected on a separate sample of students indicated that males and females evaluate their own performance on the task similarly, it still could be that in general, writing essays—a task that requires some reasonable degree of verbal fluency to perform well—is perceived as a task more suited for the "typical" female. To explore this possibility, 307 subjects —students in introductory psychology and social psychology classes—completed a "Campus Issues Questionnaire" and evaluated on a 5-point scale either their own performance or the expected performance of other females or males. These data indicated that there was no difference in the evaluation of performance, either one's own or that of others, as a function of the sex of respondent, $F(1, 301) = .15$. However, there was a difference in evaluation of others' expected performance as a function of their sex, $F(1, 301) = 8.24$, $p < .01$; females were expected to do a better job than males. Thus, it appears that the differential pay given to other females and males was due, at least in part, to the differential expectations that allocators might have had about performance as a function of sex.

In addition, both males and females evaluated their own performance more favorably than they did the expected performance of other persons of either sex, $F(1, 301) = 105.44$, $p < .0001$. Of particular interest is the finding that females evaluated their own performance more favorably than they did the expected performance of either other women, $F(1, 301) = 23.92$, $p < .0001$, or other men, $F(1, 301) = 78.64$, $p < .0001$. Table 2 presents the cell means that are relevant to these effects.

Table 2
Evaluation of Performance on Campus Issues Questionnaire

Sex of target of evaluation	Target of evaluation: Respondent Score	Respondent *n*	Other Score	Other *n*
Female	4.52	118	3.96	44
Male	4.66	105	3.48	40

Note. Scores on this measure could range from 1.0 (i.e., poor performance) to 5.0 (i.e., excellent performance).

Discussion

The results of this study indicate that persons who were recruited for pay, who actually worked for a period of time, and who received payment for their work differed in the amount of money that they paid themselves (and considered to be fair) as a function of their sex. These findings lend support to Mednick and Tangri's (1972) and Chesler and Goodman's (1976) contention that females expect less monetary reward than males for the work that they do.

Results also indicated that this more tenuous connection between work and pay is expressed primarily through a weaker sense of own equity in females. Females paid themselves less than other people paid females, although females also evaluated their own performance more favorably than they did the performance of other females. Thus, it appears that females' payment to themselves was not based entirely on the evaluation of their own performance. On the other hand, both females and males paid other females more than they paid other males, and they evaluated the expected performance of other females more favorably than they did that of males. Since this correspondence between evaluation and pay did not occur when females allocated money to themselves, it appears that women were less apt to apply a standard of fair pay to their own behavior than they were to the behavior of others.

It should be noted, however, that these findings do not necessarily invalidate completely other alternative explanations. Some explanations—for example, that females are more influenced by a norm of equality than are males (Leventhal, 1976; Sampson, 1975) —are not antithetical to the less-of-a-connection hypothesis. In fact, the weaker sense of own equity in females could be a necessary condition that permits these other variables to operate. For example, females might subscribe to the norm of equality because their weaker sense of own equity alone does not provide them with an adequate basis for making a decision about an appropriate distribution of rewards.

The data do, however, tend to diminish the plausibility of a number of alternative explanations. For example, one possible alternative explanation is that the female subjects in this study had been victims of wage discrimination in their pasts, which may have lowered their expectations with regard to monetary payment for work performed. A second alternative suggests that females are less experienced than males (and therefore less "skilled") at the task of determining monetary outcomes. Thus, females may have tended to underreward themselves because of a relative lack of experience in dealing with such matters. Neither of these explanations, however, can account for the amount of money women allocated to female targets. That is, any lowered expectations regarding pay for work as a result of past experience (or inexperience) would most likely generalize to paying other females as well and, therefore, should have resulted in female allocators treating other females in the same way that they treated themselves. Results indicated, however, that this was not the case.

Another potential alternative explanation that is rendered more implausible by the results of the present research is that females are more "moral" or socially responsible than males. The data that are summarized in Table 1 suggest that *both* males and females tended to actually take more for themselves than they indicated they felt was appropriate. Thus, while females paid themselves less than did males, they (as did males) still "took advantage" of the situation and paid themselves more than the amount they reported was appropriate immediately after that allocation.

Thus, there is no indication that females acted in a more moral manner than did males.

As noted above, the greater payment made by allocators to other females, compared to that made to other males, was congruent with differential evaluations of performance as a function of sex. However, examination of the relevant means (Tables 1 and 2) suggests that the amount actually paid to other females was excessive, even given an expected difference in performance. Moreover, the difference between what subjects actually paid other females and males was greater than the difference between what they said was fair pay for the two sexes. Thus, a case can be made that allocators in the present study were being "kind" to females. This conclusion is congruent with the findings of Gruder and Cook (1971) that persons are more helpful to females than to males and those of Messé and Callahan-Levy (Note 3) that allocators tend to be kinder to females when distributing a reward between themselves and another person. The notion of being kinder to females is congruent with the passive, dependent aspects of the traditional female sex role. Thus, if females are persons towards whom others are nurturant and protective, there is no need for them to be concerned with acquiring a reasonable amount of money for their work.

The original impetus for the present research came from an attempt to explain, at least in part, the sex differences that were rather consistently found in studies of reward allocation in which subjects were asked to anonymously divide money between themselves and another person. The present findings suggest that one reason why females appear more generous in such situations is their indifference, relative to males, to receiving "appropriate" money for their work. Study 2, presented below, attempted to replicate this result across a different task and a number of age groups and to relate pay allocation to some consequences of sex role socialization.

Study 2

Given the results of Study 1, it seemed reasonable to explore two related questions: (a) whether sex differences in self-pay allocation occur across a number of ages and (b) whether these differences, if they exist, are related to sex role socialization. To this end, in Study 2, we examined self-pay allocation in first-, fourth-, seventh-, and 10th-grade females and males and measured these subjects' preferences for masculine and feminine occupations and activities.

Studies that have examined reward distribution in children have yielded mixed results. For example, Leventhal and Anderson (1970) found that kindergarten-age females allocated less reward to themselves than did males. On the other hand, Lane and Coon (1972), in a replication of the Leventhal and Anderson study, found no sex difference, and Lerner (1974), in a study that used a similar methodology, found that 5-year-old females allocated more to themselves than did 5-year-old males. The reasons for these discrepant findings are not clear. Although the methodologies employed have much in common, they also differed in many potentially relevant ways—for example, the type of reward that the subjects were asked to distribute—and these differences in procedure could have caused the differences in results.

Given the age of the children who were examined in these studies, it could be that these differences in research findings were due, at least in part, to the complex interpersonal nature of the reward distribution task. It is possible that a more straightforward task—for example, self-pay allocation—would yield more consistent, interpretable findings.

Hartley (1960) found that both males and females, who were as young as 5 years old, were capable of making clear distinctions between men's and women's work roles. Her subjects saw women's work to be centered on homemaking duties (which receive no direct monetary pay), whereas men's work was perceived as a number of tasks for which money was paid. Assuming (a) that the weaker sense of own equity in females is connected to sex role socialization regarding appropriate occupations for men and women and (b) that the consequences of the socialization appear at least by the age of 5 years, it follows that a sex difference in self-pay allocation

Table 3
Mean Ages (and Frequencies) of Female and Male Subjects at the Four Grade Levels

	Grade							
	First		Fourth		Seventh		10th	
Sex	Age (years)	*n*	Age (years)	*n*	Age (years)	*n*	Age (years)	*n*
Female	6.28	7	9.33	12	12.57	11	15.40	10
Male	6.14	14	9.33	9	12.45	7	15.60	10

would occur quite early in children's psychological development. However, as children become more experienced with sex role-related activities, it also seems likely that this difference would be increased. Thus, we expected that the sex difference in self-pay allocation would be greater among adolescent subjects than it would be among younger children.

A fundamental premise of the less-of-a-connection hypothesis is that it is related to sex role socialization, especially those areas of sex role that involve occupational choice. In Study 2, we attempted to examine the premise more directly by measuring preferences for sex-typed occupations and activities and exploring in females the relationship between these measures and their self-pay allocation behavior. We expected that girls who expressed a preference for masculine occupations and activities would be more likely to pay themselves as males do, whereas girls who selected more feminine pursuits would be more likely to underpay themselves relative to males.

Method

Subjects

The subjects were male and female students in the first, fourth, seventh, and 10th grades of schools in a district that serves a lower-middle-class to middle-class suburb of Lansing, Michigan. Table 3 presents the frequencies and mean ages of the subjects. In each class the teacher introduced the experimenter to his or her students. The experimenter explained to the class that the students were needed to perform a job—participating in an interview that involved "things that you might like to do." Parental permission was obtained for all those students who agreed to participate. Of all the students who were contacted, only one, a female 10th-grade student, declined to take part.

Table 4
Presentation Order and Femininity Ranking of Items in the Activities and Occupations Scales

	Scale			
	Activities		Occupations	
Order of presentation	Item	Ranking	Item	Ranking
1	Playing football	10	Firefighter	9
2	Sewing or knitting	1	Secretary	1
3	Shoveling snow	9	Football player	10
4	Cooking or baking	2	Nurse	2
5	Going camping	8	Astronaut	8
6	Dancing	3	Teacher	3
7	Going bowling	7	Police officer	7
8	Swimming	6	Parent	4
9	Ice skating	4	Doctor	6
10	Reading	5	Baker	5

Interview Schedule

The instruments used in the study were administered to subjects via an interview protocol, which was the "task" for which subjects were paid. This interview required approximately 25 minutes to complete. A female or male undergraduate experimenter (counterbalanced with respect to sex of subject) administered the interview individually.

The interview began with a number of questions about personal and family demographic characteristics, for example, age of subject, family size, and parents' employment status. Then, subjects were presented with a list of 10 activities. This list was obtained through research that established, via a modified Thurstone procedure, that these activities differed significantly with respect to the degree of femininity and masculinity associated with them. Table 4 presents these activities ranked as a function of their perceived femininity, and it also indicates the order in which the activities were given to the subjects. The experimenter read through the list twice and then asked each subject to indicate her or his three most preferred activities.

Next, subjects were presented with 10 occupations. These occupations were selected to vary along the feminine–masculine dimension and were obtained via the same procedure that we used to construct the Activities scale. The 10 occupations, their rank order in terms of femininity, and their presentation order also are given in Table 4. The Occupations scale was administered to subjects in exactly the same manner as was the Activities scale.

Following the presentation of the Occupations scale, the experimenter administered a short form (10 items) of the Bialer-Cromwell Locus of Control Scale for Children (Bialer, 1961). This scale consists of statements that require a yes–no answer.

Next, subjects were given instructions concerning their self-pay allocation (see below), which they performed. Then, the experimenter asked three questions about each subject's evaluation of his or her performance. The exact wording of these questions is presented in Table 5. Finally, the experimenter asked the subject what she or he thought was the "fair amount" of pay for doing this work.[5]

Pay

Prior to Study 2, via extensive pilot testing, we obtained a measure of what children considered to be appropriate pay for participation in an experiment of this type. This evaluation was necessary so that subjects could be presented with an amount of money with which to pay themselves that was in great excess of a consensus standard of what was fair. This pilot research revealed that children considered appropriate pay to be less than $1, and thus, $3 in dimes was the amount of money presented to each subject.

Table 5
Mean Responses of Female and Male Subjects to Questions About Their Own Performance

Interview question	Females (n = 40)	Males (n = 40)
1. How well do you feel you did at this "job"?	2.52	2.70
2. Do you think you did better than most boys your age would do?	2.15	2.18
3. Do you think you did better than most girls your age would do?	2.10	2.30

Note. Subjects answered all questions using a 3-point scale. Thus, scores could range from 1–3, with higher scores indicating a higher evaluation.

Also, there was some question about the appropriateness of money as a reward for 6-year-old first graders. Pilot testing indicated that most first graders were confused about the value of coins, but that they were interested in, and understood, the value of Hershey "kisses." This pretesting also indicated that children in first grade saw less than 10 "kisses" as fair pay for being interviewed. Thus, we presented subjects in this grade level with 30 Hershey "kisses," so that they could take what they wanted as their pay.

Variables and Experimental Design

The design of Study 2 was both experimental and correlational. Sex of subject and grade level were examined through a 2 (female, male) × 4 (first, fourth, seventh, and 10th grades) factorial design with two dependent measures: (a) the amount of reward that subjects actually allocated to themselves and (b) the amount of reward that subjects indicated was fair pay. In addition, the two pay measures were correlated with the two measures of sex role preference, scores on the locus of control measure, and the 12 biographic–demographic questions (e.g., age of subject, number of brothers, employment status of the father, etc.).

[5] The interview was worded exactly the same for all subjects. Thus, care was taken to insure, via pilot testing, that the younger children could understand the meaning of the questions.

Procedure

A major goal in designing the present research was to replicate as closely as possible the essential features of the design and procedure used in Study 1.[6] Study 2 was conducted in large classrooms at the schools that the subjects normally attended. Approximately 10 subjects, both male and female, were interviewed at the same time, but each was interviewed by a different experimenter, and they were always well separated within the room. Subjects were seated so that they faced the experimenter, who had her or his back to the wall, and so that no subject could see or communicate with other subjects in the room.

At first, subjects were given general instructions that informed them that they were about to complete a "job" for a class of students at Michigan State University. They were asked to take this job as seriously as possible and to do their best to answer the interview questions honestly and accurately, although there would be no right or wrong answers to whatever they were asked. They also were reminded that as promised earlier, they were to be paid for their time after they completed the job. Then the experimenter administered the interview schedule to the subject, a procedure that took approximately 25 minutes to complete.

At the completion of the interview, the subject was given the opportunity to pay herself or himself. The subject was told that her or his job was completed and that it was time to be paid. The rationale that was given for this self-payment procedure was as follows:

> We have some money to pay you. The only problem is, we don't know what fair pay is for a job like this for someone like you. So we've decided to let each person take what they feel is a fair amount of money for their work. This means we want you to take what you feel you deserve.

The experimenter also stressed that there was no "correct" amount and that the subject would actually keep the amount that she or he chose.

After the subject received these instructions, the experimenter placed 30 dimes (or candies, for first-grade subjects) on a large cardboard and gave the subject a blank envelope and an envelope marked "Leftover." Subjects were instructed to put what they wanted in the blank envelope, which they would keep, and to put the remainder in the envelope marked "Leftover." The experimenter stressed that she or he would turn around and not look or, if possible, leave the room, so that no one would know how much money (or candy) the subject actually kept. The subject also was told that any leftover money (or candy) would be returned to Michigan State University to be used for other studies.

Results

As in Study 1, there were two measures of allocation behavior: (a) the amount of money that subjects actually allocated to themselves as pay and (b) the amount that subjects indicated they felt was fair pay on a postallocation question. Summary data for the self-allocation dependent measures are presented in Table 6, which presents the mean number of payment units (candy in the first grade, dimes in the fourth, seventh, and 10th grades) actually kept and indicated as fair pay by males and females at the four grade levels.

The amount of reward subjects actually kept and the amount they indicated was fair pay on the relevant postallocation question were subjected to 2 (sex of subject) × 4 (grade level) unweighted-means analyses of variance (ANOVAs). As in Study 1, the two ANOVAs produced essentially the same results. The ANOVAs revealed two significant effects: the main effect for sex of subject, $F(1, 72) = 15.54$, $p < .001$ (actual pay), and $F(1, 72) = 9.22$, $p < .005$ (fair pay), and the main effect for grade level, $F(3, 72) = 13.39$ (actual pay) and $F(3, 72) = 8.37$ (fair pay), both ps $< .001$. As expected, Table 6 shows that the main effect for the sex of subject reflected a tendency in males to allocate greater pay to themselves than did females and to perceive that more pay was fair. Table 6 also shows that as the grade level increased, the amount of reward that they felt was fair pay tended to decrease. Comparisons performed on the two scores between adjacent grade levels across both sexes indicated that there was a very significant decrease in the amounts kept and thought to be fair between the fourth and seventh grades, $F(1, 72) = 27.64$

[6] Study 1, however, recruited college-age subjects for pay through an advertisement in the campus newspaper. Although the subjects in Study 2 were essentially volunteers who were also recruited for pay, the act of volunteering required very little effort on their part. In fact, many of the elementary, junior high, and senior high students appeared to welcome the opportunity to be excused from class for the time required by the study. Thus, one essential difference between the two studies is the motivation of subjects for the pay offered. It is not possible to determine the degree to which this difference might have affected the interpretation of the results, but it is assumed that these differences in motivation would be uniform across sex, within grades.

Table 6
Mean Payment (in Units of Pay) That Females and Males Allocated to Themselves at Each Grade Level

Sex of subject	Grade level: First Amount	n	Fourth Amount	n	Seventh Amount	n	10th Amount	n
			Actual pay allocation					
Female	10.86	7	12.92	12	3.73	11	2.40	10
Male	16.57	14	18.56	9	6.57	7	10.70	10
			"Fair" pay allocation					
Female	10.14	7	9.17	12	4.64	11	3.00	10
Male	14.86	14	15.44	9	5.29	7	8.50	10

actual pay) and 12.43 (fair pay), both ps < 001.

Analysis of variance tests the significance f interval rather than ratio differences, and ıese interval differences tend to be con- trained by the total score over the cells of a iven comparison. Both male and female sub- ects in the higher grades took much less than id their younger counterparts; in fact, by ıe seventh grade, the mean self-allocation as only about five units, across sex. This latively small amount of pay had the effect f constraining the size of the interval differ- nces between the conditions of sex within ıe older grades and, thus, would tend to inimize the magnitude of the obtained Sex Grade Level interaction ($F < 1.00$, in both nalyses). Therefore, in an attempt to ex- mine the possibility that sex, in fact, did in- uence allocation of pay differentially across ade levels, actual and "fair" female scores ere transformed to proportion scores, which represented the females' proportion of the mean males' score at each grade level. For example, each first grade female's actual and fair scores were divided by the respective actual and fair means of the first grade males' scores. In this way, an analysis could be performed on ratio scores, rather than on the more constrained interval units.

These female proportion scores were subjected to one-way (grade level) analyses of variance. These analyses revealed that the effect of grade level was significant for actual pay, $F(3, 36) = 4.94$, $p < .01$, but not for fair pay, $F(3, 36) = 1.20$, $p > .1$. The means of these proportion pay scores over the four grade levels are presented in Table 7. Planned comparisons (Winer, 1971, p. 384) performed between adjacent grade levels revealed that as predicted, the actual pay proportion scores for the 10th-grade females were significantly less than the scores for seventh-grade girls, $t(34) = 2.55$, $p < .02$. This also was the case

able 7
lean Proportion of Average Male Pay That Females aid Themselves in Each Grade

Pay score	Grade level: First (n = 7)	Fourth (n = 12)	Seventh (n = 11)	10th (n = 10)
Actual	.66	.69	.57	.22
Fair	.68	.59	.88	.35

for fair pay proportion scores, $t(34) = 1.88$, $p < .05$, even though the overall effect of grade level was not significant.

The possibility that the sex difference in self-pay was due to differences in the males' and females' evaluations of their own performances was examined through 2 (sex of subject) × 4 (grade level) analyses of variance, performed on the three questions that were relevant to this issue that were asked during the postallocation portion of the interview.

Table 5, which presents the mean responses to these questions as a function of sex of subject, indicates that there was little difference between males and females in their self-evaluations. The results of the three analyses of variance supported this conclusion, since no effect attributable to sex of subject was statistically significant, (Fs = .08 to 2.60). Thus—as was the case for the findings for Study 1—the sex difference in self-pay allocation does not appear to be due to any underlying difference in performance evaluation.

We expected that scores on the sex role preference scales would correlate with self-allocation behavior. It should be noted that t tests comparing the median rank of the three items that were selected by a male or female subject for both the Activities scale (female $M = 5.85$, male $M = 8.50$) and the Occupations scale (female $M = 3.98$, male $M = 8.43$) were highly significant, ts(78) = 8.35 and 16.70, respectively, ps < .001. The two scales also were significantly correlated with each other, $r(76) = .65$, $p < .001$. As mentioned previously, we expected that scores on the sex role preference scales would correlate with self-allocation behavior in females, such that more masculine preferences would relate to more masculine payment behavior. To test this prediction directly, the proportion of the average male payment in her grade level that each female took for herself was correlated with the median rankings of her selected occupations and activities. The resultant correlation coefficient revealed that scores on the Occupations scale correlated substantially with proportion scores, $r(38) = .66$, $p < .001$, indicating that as expected, the more masculine the occupational orientation, the more masculine the allocation behavior. The correlation coefficient computed for scores on the Activities scale and the proportion payment scores was in the predicted direction, but was not statistically significant, $r(38) = .15$.[7]

The payment scores also were correlated with the locus of control measure and the demographic variables. None of these correlations, however, yielded statistically significant findings.

Discussion

The findings of Study 2 replicate closely those of Study 1. Females—even the females in the first-grade sample—tended to pay themselves less than did males at the equivalent grade level, as did the college student females who were the subjects in the first study. Moreover, the sex differences in self-pay that were found in both studies do not appear to be a function of differential evaluations of task performance. Thus, findings of the two studies taken together provide strong, albeit indirect, support for the proposition that females are socialized to have a weaker sense of their own equity.

There were some slight differences between the findings of the two studies, however. For example, the 10th-grade sample of Study 2 was closest in age to the subjects in Study 1 and yet the sex difference at this grade level was substantially larger than was evidenced in the college student sample. It is likely that this difference in results is attributable to differences between the two studies in terms of such factors as (a) the manner in which subjects were recruited to participate—which probably caused systematic differences in subjects' motivation for money—and (b) the tasks that were used as the basis for allowing

[7] It is of interest to note that the comparable correlations for males (i.e., relating femininity–masculinity of occupational and activity choices to self-pay) were similar to those for females, but of lower magnitude. Specifically, the correlation between occupational preference and self-pay was .24, $p < .075$, while, again, it was considerably lower for activity preference and self-pay ($r = .04$).

subjects to pay themselves. It is possible that these and other factors mediate the magnitude of the sex difference. However, it is noteworthy that the sex difference in self-pay allocation did appear under such different conditions and in somewhat different subject populations.

Given Hartley's (1960) findings that young children to some extent have internalized the traditional division of labor as a function of sex, we expected some degree of sex difference in self-pay behavior in our first-grade sample. However, it seemed reasonable that this sex difference would be greatest when the allocators were closest to the age at which they were expected to assume the adult sex role. Thus, we also expected that the relative difference between male and female self-pay would be greatest for our 10th-grade sample.

Findings supported both these expectations to some extent. In fact, though, the difference in the first-grade sample was somewhat stronger than we had originally expected. The magnitude of this finding suggests that the consequences of that portion of sex role socialization that involves economic activities become evident at a rather early age. The proportional difference in self-allocation as a function of sex was greatest, as expected, in the 10th grade, but, as noted above, this increase was not evident when interval differences were examined. This disparity, as we suggested, could be due to the constraints placed on the seventh- and 10th-grade data due to the relatively small amount of money that both males and females in these grades paid themselves. On the other hand, there is the possibility that the hypothesized increase is not really that strong. Thus, it seems reasonable that future research should examine this issue—and that of the comparability of 10th graders and college students—by asking samples of seventh-grade, 10th-grade, and college student male and female subjects to pay themselves after they work on a more challenging task, one that would provide them with a sound basis for taking a reasonably large amount of money.

We hypothesized that females and males differ in their allocation behavior as a consequence of differences in sex role socialization, and thus, this behavior in females should be related to other manifestations of their sex role. Findings of Study 2 indicated that this was the case, since self-pay allocation in females correlated strongly with the femininity–masculinity of occupation preferences. It is interesting that a less specifically economic index of femininity–masculinity (i.e., activities) did not yield a significant relationship with self-pay allocation. If this specific finding is a valid reflection of this relationship in general, it would suggest that the sense of one's own equity is a consequence of a somewhat narrow range of socialization experiences: those that involve rather specifically the economic area of human endeavors.

Conclusions

The findings of the two studies also appear to have more general implications. The well-documented lack of parity in pay that exists between female and male members of the work force in American society to some extent could be the result of a willingness on the part of women to work for lower wages than men. Mednick and Tangri (1972) and Chesler and Goodman (1976) contend that one factor that has contributed to this willingness is the different degree to which work and money are related in females and males. The present research does provide evidence in support of this assertion.

In addition, the results of the present work, as well as the preponderance of past studies, strongly suggest that women tend not to act in ways that maximally benefit them economically. It could be argued that the "feminine" orientation toward money, in fact, is more psychologically healthy. However, it is nevertheless the case that such an orientation literally is costly to those persons who hold it, especially since it makes them vulnerable to exploitation by persons who have more self-serving ideas about money and work.

These conclusions have important implications for the evaluative examination of traditional sex roles that currently is being undertaken within and outside of psychology. For this reason alone, they merit and require further examination.

Reference Notes

1. Messé, L. A., & Lichtman, R. J. *Motivation for the reward as a mediator of the influence of work quality on allocation behavior.* Paper presented at the meeting of the Southeastern Psychological Association, Atlanta, April 1972.
2. Leventhal, G. S. Reward allocation by males and females. In L. A. Messé (Chair), *Individual differences in equity behavior.* Symposium presented at the annual meeting of the American Psychological Association, Montreal, September 1973.
3. Messé, L. A., & Callahan-Levy, C. M. *Sex and message effects on reward distribution behavior.* Paper presented at the meeting of the Midwestern Psychological Association, Chicago, May 1975.

References

Berkowitz, L., & Walster, E. (Eds.) *Equity theory: Toward a general theory of social interaction. Advances in experimental social psychology* (Vol. 9). New York: Academic Press, 1976.

Bialer, I. Conceptualization of success and failure in mentally retarded and normal children. (Doctoral dissertation, George Peabody College for Teachers, 1960). *Dissertation Abstracts,* 1961, *21,* 2357. (University Microfilms No. 60-05859)

Chesler, P., & Goodman, E. J. *Women, money and power.* New York: Morrow, 1976.

Gruder, C. L., & Cook, T. D. Sex, dependency, and helping. *Journal of Personality and Social Psychology,* 1971, *19,* 290–294.

Hartley, R. E. Children's concepts of male and female roles. *Merrill-Palmer Quarterly,* 1960, *6,* 84–91.

Lane, I. M., & Coon, R. C. Reward allocation in preschool children. *Child Development,* 1972, *43,* 1382–1389.

Lane, I. M., & Messé, L. A. Equity and the distribution of rewards. *Journal of Personality and Social Psychology,* 1971, *29,* 1–17.

Lane, I. M., & Messé, L. A. The distribution of insufficient, sufficient, and oversufficient rewards: A clarification of equity theory. *Journal of Personality and Social Psychology,* 1972, *21,* 228–233.

Lerner, M. J. The justice motive: "Equity" and "parity" among children. *Journal of Personality and Social Psychology,* 1974, *29,* 539–550.

Leventhal, G. S. The distribution of rewards and resources in groups and organization. In L. Berkowitz & E. Walster (Eds.), *Equity theory: Toward a general theory of social interaction. Advances in experimental social psychology* (Vol. 9). New York: Academic Press, 1976.

Leventhal, G. S., & Anderson, D. Self-interest and the maintenance of equity. *Journal of Personality and Social Psychology,* 1970, *15,* 57–62.

Leventhal, G. S., & Lane, D. W. Sex, age, and equity behavior. *Journal of Personality and Social Psychology,* 1970, *15,* 313–316.

Leventhal, G. S., & Michaels, J. W. Extending the equity model: Perception of inputs and allocation of reward as a function of duration and quantity of performance. *Journal of Personality and Social Psychology,* 1969, *12,* 303–309.

Mednick, M. S., & Tangri, S. S. New social psychological perspectives on women. *Journal of Social Issues,* 1972, *28,* 1–16.

Mikula, G. Nationality, performance, and sex as determinants of reward allocation. *Journal of Personality and Social Psychology,* 1974, *29,* 435–445.

Sampson, E. E. On justice as equality. *Journal of Social Issues,* 1975, *31,* 45–64.

Weick, K. E. The concept of equity in the perception of pay. *Administrative Science Quarterly,* 1966, *11,* 414–439.

Winer, B. J. *Statistical principles of experimental design* (2nd ed.). New York: McGraw-Hill, 1971.

Received June 5, 1978

Journal of Personality and Social Psychology
1979, Vol. 37, No. 3, 447-461

Self-Esteem, Self-Consciousness, and Task Performance: Replications, Extensions, and Possible Explanations

Joel Brockner
State University of New York College at Brockport

Previous research has suggested that the task performance of low-self-esteem individuals (low SEs) is impaired under conditions designed to increase self-focused attention. Task-focusing, rather than self-focusing, manipulations have actually bolstered the achievement of low SEs. The present studies were designed to replicate and extend these findings. The results of two experiments demonstrated that the performance of low SEs (but not of individuals with higher self-esteem) was affected by a variety of attentional manipulations. As before, task-focusing instructions enhanced and self-focusing stimuli impaired their performance on a concept formation task. Similar results were obtained for individuals who scored high (but not for those who scored low) on an individual difference measure of self-consciousness. Study 2 also demonstrated that when the task-focusing manipulation worked, it neutralized the adverse effects of the self-focusing stimulus on the low SEs' performance. Supplementary data suggested that the manipulations generally had their intended effects on attentional focus and that attentional focus influenced performance. It is hypothesized that the low SEs' level of self-focused attention influenced performance through the mediating effects of anxiety.

Self-esteem (SE) theorists and researchers have established that relative to high-self-esteem people (high SEs), low-self-esteem individuals (low SEs) are more likely to suffer from a variety of emotional and behavioral problems (e.g., greater anxiety, less happiness; Coopersmith, 1967) and more likely to become alcoholics (Wahl, 1956). Perhaps because low SE has so many undesirable correlates, clinical, social, and personality psychologists have long been concerned with uncovering effective methods of enhancing the self-evaluations of low SEs. Unfortunately, this task has proven quite difficult, since low SEs seem to be trapped in a vicious, self-defeating cycle of low SE. For example, many studies (Hamachek, 1971; Shrauger, 1972) have shown that low SEs do more poorly than high SEs in achievement settings (i.e., situations that have evaluative implications for one's SE). Poor performances, in turn, would seem to set the stage for continued self-criticism and low SE. Thus, of potentially critical importance are the precise reasons for the often deflated task performance of low SEs.

Low Self-Esteem and Self-Focused Attention

One of several possible causes of low SEs' poor task performance may be their focus of attention while engaging in the task. Specifically, there is converging evidence that low SEs are generally more self-conscious or self-focused than high SEs. Turner, Scheier, Carver, and Ickes (1978) discovered a significant negative correlation between individual difference measures of self-esteem and self-consciousness (Fenigstein, Scheier, & Buss, 1975). In addition, recent findings from

The author would like to thank John Eckenrode, Herb Fink, Larry Hjelle, Jeff Rubin, Larry Wallnau, and two anonymous reviewers for their insightful comments on an earlier version of the manuscript. In addition, he is greatly indebted to Blethyn Hulton and Dyan Wallnau for their competent assistance in collecting the data.

Requests for reprints should be addressed to Joel Brockner, Department of Psychology, State University of New York College at Brockport, Brockport, New York 14420.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3703-0447$00.75

the objective self-awareness literature suggest an inverse relationship between self-focused attention and SE. For instance, Ickes, Wicklund, and Ferris (1973) found that subjects who completed a SE inventory in the presence of a self-focusing stimulus (a mirror) rated themselves significantly lower than those in a no-mirror condition. Also, self-focusing stimuli cause people to behave like low SEs. For example, both Duval (1976) and Wicklund and Duval (1971) found greater conformity when subjects were exposed to self-focusing stimuli. In parallel fashion, low SEs have also exhibited higher levels of conformity in some studies (e.g., Janis, 1954), although the evidence here is not totally consistent. Taken together, however, the evidence seems to implicate an inverse relationship between SE and self-focused attention.

Self-Focused Attention and Task Performance

The task performance of low SEs may be impaired by their heightened self-focused attention for at least two related reasons: (a) Given that attentional capacities are finite, low SEs will not be able to devote sufficient attention to the task, and (b) low SEs may well focus on their own negative characteristics (Mischel, Ebbesen, & Zeiss, 1976), leading to anxiety. Anxiety, in turn, may impair performance, particularly on complex tasks (e.g., Spence, Farber, & McFann, 1956).

This is not to say, however, that low SEs will *always* be more self-conscious than high SEs and/or perform more poorly than high SEs. For example, if the task is inherently engaging, one might expect all subjects' self-focused attention to be reduced. Under such conditions there may be no relationship between SE and performance. In fact, Brockner and Hulton (1978), Shrauger (1972), and Shrauger and Terbovic (1976) found that low and high SEs performed equally well on a concept formation task that naturally captured the subjects' attention. Nevertheless, when self-focusing stimuli were introduced into the experimental situation (i.e., an audience in the Shrauger, 1972, study and an audience and mirror in the Brockner and Hulton, 1978, experiment), the performance of the low SEs dropped markedly, whereas the high SEs maintained their high level of performance. Hence, even if low SEs are not already more self-conscious than high SEs, it would seem that their susceptibility to becoming self-focused is considerably greater than high SEs'.

If the task performance of low SEs is impaired by heightened self-focused attention and/or a greater propensity to become self-aware, then any manipulation that enhances task-focused and decreases self-focused attention should improve their performance. Indeed, Brockner and Hulton (1978) discovered that low SEs actually performed better than high SEs when both groups were given a set of instructions designed to increase task focusing. These data may be of considerable practical significance, in that the favorable performance of the low SEs may serve as a starting point for breaking the vicious cycle of low SE. That is, it is entirely possible that with repeated success experiences in achievement settings, low SEs will elevate their self-evaluations to a more positive level.

Questions Addressed by the Present Research

The tendencies for low SEs in the Brockner and Hulton (1978) study to perform poorly in the presence of self-focusing stimuli and favorably when instructed to focus on the task raised a number of theoretical and practical questions. First, could the various findings be replicated? Do self-focusing stimuli consistently impair the low SEs' task performance? This is an important question from a practical or applied vantage point. If self-focusing stimuli consistently impair low SEs, this may help us identify (and perhaps start to change) the factors that might ultimately perpetuate low SE. From a more theoretical viewpoint, experimental social psychologists have witnessed a great resurgence of interest in the notion of self-consciousness. To study the effects of self-consciousness, many investigators have manipulated self-focused attention with mirrors, video cameras, and audiences and tested its effect on a variety of behaviors (see Duval & Wicklund, 1972). The present formulation views self-esteem as a factor related to a person's self-

focusing tendencies. Thus, it would be useful to know how this individual difference variable interacts with situational manipulations of self-consciousness.

Second, could the task instructions results be replicated? Given that the instructions enhanced the low SEs' performance and that enhanced performance might begin to break the vicious cycle of low SE, it would be important to know if these instructions reliably improve the low SEs' performance.

Third, can the results be extended? An important condition missing from the Brockner and Hulton (1978) study is one in which the self-focusing stimuli and task-focusing instructions are combined. What happens when these two manipulations with opposite effects on performance and presumably opposite effects on attention are pitted against each other? Do the two stimuli tend to cancel each other, does one override the other, or do they produce some other kind of interaction effect?

Fourth, if the results can be replicated and extended, what are the appropriate explanations for the data? To what extent are attentional processes influencing performance? Brockner and Hulton (1978) used several manipulation checks of attentional focus, and while the results on these items were generally consistent with performance, the manipulation check data were somewhat weak. The present research included additional manipulation checks that may give a more valid indication of subjects' focus of attention. Two studies were conducted to answer the above questions.

Study 1

Subjects were given a pretest measure of their self-esteem and completed the identical concept formation task used by Brockner and Hulton (1978). One third of the subjects completed the task in the presence of self-focusing stimuli. The stimuli used in this study were a video camera and a mirror. Hence, this condition will be referred to as the video–mirror condition. Note that a different procedure was used by Brockner and Hulton, who had subjects work on the task before an audience that was observing them from behind a one-way mirror. To strengthen a possible attentional interpretation of the results, a different operationalization of self-focused attention was employed in the present experiment. One third of the subjects were provided with the identical task-focusing instructions used by Brockner and Hulton (the task condition). The other third of the subjects received both the self-focusing *and* task-focusing manipulations (the combined condition). In sum, a two-factor (SE × Attentional Treatment) between-subjects experiment was conducted. It was expected that the manipulations of attentional focus would have the most pronounced effects on the low SEs' task performance. Following the results of Brockner and Hulton (1978), it was predicted that compared to high SEs, low SEs would perform significantly worse in the video–mirror condition and significantly better in the task condition. Given the seemingly greater salience of the self-focusing versus the task-focusing manipulation, it was expected that the low SEs would perform somewhat worse than high SEs in the combined condition.

Method

Participants

The 88 volunteer subjects (53 male, 35 female) were summer school students at a Vermont state university. Each subject received $3 for participating.

Materials

Self-esteem inventory. The inventory was identical to the one used by Brockner and Hulton (1978) and Shrauger (1972). The scale measured the subjects' perceived competence across 16 various situations (e.g., academic, athletic, social). Subjects had to indicate the percentage of time a particular behavior or outcome applied to them (e.g., "When meeting new people for the first time, what percentage of the time are you able to impress them favorably and form good relations?"). Nine items were worded so that higher percentages indicated higher SE, whereas seven were phrased so that lower percentages indicated higher SE. For the latter items, the subjects' percentage ratings were subtracted from 100. The average score for all 16 items was computed such that higher scores represented higher SE. The mean score for all subjects was 70.36 ($SD = 9.40$). Subjects were classified as high or low SEs on the basis of a median split (high SEs: $M = 77.27$, $SD = 5.91$; low SEs: $M = 63.14$, $SD = 6.49$). Subjects

were randomly assigned to predetermined attention conditions. As a result there was an unequal distribution of subjects across conditions. Nevertheless, *within* each SE level the mean SE scores across conditions were virtually identical.

Concept formation task. The task was identical to the one employed by Brockner and Hulton (1978). Specifically, 32 squares were attached to a 14 × 25 × .25 inch (36 × 64 × .6 cm) wooden board. Each square had five characteristics: number (one or two), size (large or small), color (black or white), shape (circle or square), and position (left or right). Beneath each square a small hole had been drilled in the wooden board.

The subject was told that the concept would consist of one, two, or three characteristics. The experimenter selected a square that was an example of the correct concept. The subject's task was to determine the concept by lifting up other squares, one at a time. If the square was an example of the correct concept, the word *yes* was written on a piece of oak tag that had been inserted below the board. If the square was not an example of the correct concept, the word *no* had been written below. When subjects believed that they had ascertained the correct concept, they wrote it on a slip of paper provided to them. At no time were subjects ever informed of the accuracy of their responses. Subjects were told that while "we are interested in the speed and accuracy with which you determine the concepts, it is more important to be accurate." There were seven trials, which all participants completed in an identical random order. After subjects completed each trial, the experimenter removed the underlying piece of oak tag for that trial and inserted the piece containing the feedback for the next trial.

Procedure

All participants were run one at a time. Upon entering the laboratory they were led to a cubicle and were told that the experiment would consist of a number of parts. In the first part they were asked to complete the self-esteem inventory. After completing this measure they were led to another cubicle where they found the concept formation task and the accompanying instructions. Subjects were asked to read the instructions to themselves while the experimenter simultaneously read them aloud. The instructions informed subjects how to determine the concepts. In addition, subjects were provided a practice trial to familiarize them further with the task. During the practice trial the experimenter directed the subject toward the correct concept whenever necessary. However, subjects always had to determine the concept themselves on the practice trial.

After the practice trial was completed, the focus-of-attention manipulations were introduced. In the video–mirror condition, the experimenter placed a large mirror (24 × 36 inches—61 × 91 cm), which had been hitherto unexposed, directly in front of the subject. Because the mirror was so large, subjects could see not only themselves but also the experimenter observing them as they worked on the task. Also, the experimenter uncovered a video camera and pointed it directly at the subjects, saying:

> Since we are interested in how people go about forming concepts, a video camera has been placed to your right. We will be making a videotape of you and your concept formation strategies. This tape will aid us greatly, as we will be able to observe you an unlimited number of times in the future when we play back the tape. Furthermore, you will notice the mirror in front of you. This will help the experimenter observe you as you are working on the task.

The experimenter then switched on the video-recording unit. In the task condition the experimenter told the subjects:

> OK, now that you are familiar with the task, let me just give you the following advice. This task is not *extremely* difficult, although it *can* be somewhat tricky. Therefore, I can't emphasize strongly enough just how important it is that you maintain your complete and undivided attention on the task at all times. People who have done this experiment in the past frequently mentioned the importance of concentrating on the task at hand. So just remember to keep focusing on the task as much as possible. Again, I can't emphasize this strongly enough.

In the combined condition subjects received both of the manipulations used in the conditions mentioned above. The order in which the self-focusing versus task-focusing stimuli were introduced was counterbalanced. Since the order variable produced no effects on performance, it will not be considered further.

After the focus-of-attention manipulations had been introduced, subjects began to work on the task. The experimenter was seated adjacent to the subject, recording the amount of time that she or he took and the number of squares looked at for each trial. While the experimenter was clearly aware of the subjects' focus-of-attention conditions, he was not aware of their self-esteem scores. When the seventh trial was completed, the experimenter asked subjects to complete a postexperimental questionnaire measuring their attentional focus and felt anxiety while working on the task. In addition, subjects were asked to write their perceptions of the purposes of the study. None of the subjects was able to guess the true nature of the study, and their few suspicions were unrelated to the actual concerns of the experiment. Finally, participants were carefully debriefed, thanked for their participation, and requested not to speak to future subjects about the study.

Table 1
Task Performance, Study 1

Measure	Self-esteem: High *n*	High *M*	High *SD*	Low *n*	Low *M*	Low *SD*
Performance in each condition						
Focus of attention						
Video–mirror	14	1.86	1.59	16	2.81	2.55
Combined	17	2.23	2.31	12	3.92	2.10
Task	14	3.14	2.45	15	3.20	2.79
Internal analysis						
Task concentration						
High	23	2.35	2.03	20	2.05	1.75
Low	22	2.45	2.41	23	4.30	2.69

Note. Performance was measured by errors. Scores ranged from 0 to 9. Thus, the higher the score, the poorer the performance.

Results and Discussion

Performance [1]

Both errors of commission and errors of omission were summed over the seven trials to determine task performance. Thus, if the subject had written a characteristic that was not part of the concept, it was scored as an error. If the subject had omitted a characteristic that was part of the concept, an error was assessed. For example, if the subject had written "one" when the concept was "left square," three errors were assessed.

A preliminary t test yielded no effect as a function of sex of subject. Thus, all data were analyzed with a two-factor unweighted-means analysis of variance. The only effect to approach significance was the self-esteem main effect, $F(1, 82) = 2.98$, $p < .10$, with low SEs making more errors. However, from Table 1 it can be seen that this effect was only obtained in the video–mirror and combined conditions (i.e., when the self-focusing stimuli were present). Several preplanned t tests revealed that low SEs made more errors than high SEs in the presence of the mirror and camera, $t(57) = 2.08$, $p < .025$, one-tailed,[2] with no difference between groups in the absence of these stimuli (i.e., in the task condition; $t < 1$).

Focus-of-Attention Manipulation Checks

Contrary to the results of Brockner and Hulton (1978), the task-focusing instructions had no ameliorating effect on the low SEs' performance in the task condition. However, the focus-of-attention measures will also show that the task-focus instructions did not have their predicted effects (i.e., the manipulation did not work). To test for their attentional focus during task performance, subjects completed the following items on the postexperimental questionnaire (all questions were completed using 41-point scales, with higher scores representing greater endorsement of the statement): *"While you were completing the task,* (a) how self-conscious did you feel, and (b) how hard were you concentrating on the task?" On the self-consciousness item,

[1] For both studies only the error data will be reported. The amount of time subjects took and the number of squares they looked at for each trial were also recorded. However, individual subject variability was quite high on both measures. As a result there were no significant effects for either measure in the two studies (all Fs < 1).

[2] Winer (1971) states that t tests are permissible for meaningful comparisons planned prior to the inspection of the data. Since an a priori directional hypothesis had been proposed, the one-tailed test was employed.

the only effect to emerge was a focus-of-attention main effect, $F(2, 82) = 4.02$, $p < .025$. All subjects were *more* self-conscious in the task condition than in the other two. Likewise, on the task concentration item, there was no evidence that the task-focusing instructions enhanced the low SEs' task-focused attention. In fact, low SEs were least task focused in the task condition.

In essence, then, the task-focusing instructions were ineffectual in this study, as the performance and attentional focus data suggest. If so, one would expect low SEs to become more self-conscious and perform more poorly than high SEs not only in the video–mirror condtion but also in the combined condition. As already stated, low SEs did perform worse than high SEs in both conditions. Moreover, there is additional evidence that low SEs were more self-conscious in both conditions. In addition to completing the self-consciousness manipulation check, the subjects in the video–mirror and combined conditions only were asked, with the 41-point scale format, "While you were completing the task, (a) how aware of the mirror were you, and (b) how aware of the video camera were you?" Each of these items correlated highly with reported self-consciousness: self-consciousness and mirror awareness, $r(57) = .48$; self-consciousness and video camera awareness, $r(57) = .58$; both $ps < .001$. Therefore, all three items were summed into a self-focused attention index. As predicted, low SEs were significantly more self-focused than high SEs, $t(57) = 1.85$, $p < .035$, one-tailed.

Internal analysis. Although the task-focus instructions did not enhance the low SEs' task concentration, one may still wonder if the low SEs who were concentrating on the task performed better than those who were not. Therefore, subjects were classified, by a median split on their task concentration scores, as high or low in task-focused attention. The summary data for the subsequent two-factor (SE × Task Concentration) analysis of variance are presented in Table 1. Interestingly, an interaction effect was obtained, $F(1, 84) = 4.73$, $p < .05$, demonstrating that low SEs only performed worse than high SEs when subjects were not highly task focused. In fact, when both groups reported high task concentration, there was a nonsignificant tendency for low SEs to perform better than high SEs.

In sum, the present study replicated Brockner and Hulton's (1978) finding that self-focusing stimuli cause low SEs to perform more poorly than high SEs. However, the task-focusing instructions did not work in this study—they did not increase task-focused, relative to self-focused, attention or enhance the performance of low SEs. Therefore, there were several reasons to perform an additional study. First, it was unclear if the task condition results were simply due to a failure of the manipulation. An alternative explanation is that the task-focus instructions also enhanced the low SEs' ego involvement or evaluation apprehension. Quite possibly these processes negated the potentially beneficial effects of the task-focus instruction on the low SEs' performance. For example, Sarason (1972) has convincingly demonstrated that evaluative or ego-involving situations cause severe performance decrements for high test-anxious subjects, a group who greatly resemble low SEs. Thus, it would be useful to know how low SEs would perform if there were independent evidence that the task-focus manipulation was successful. Specifically, if the low SEs' performance also improved, then the task condition results of Study 1 would seem to be due to a failure of the manipulation. However, if their performance did not improve even when there was evidence that the manipulation was successful, then the manipulation failure hypothesis would be less plausible.

Second, in light of the failure of the task condition manipulation, the combined condition results are also somewhat ambiguous. That is, is it the case that the self-focusing stimuli actually overpowered the task-focus instructions? Or, were the task-focus instructions, as in the task condition, not successfully manipulated? It would seem that a more convincing test of the combined effects of self-focusing and task focusing would occur if it were at least shown that each stimulus reliably affected attention when manipulated in isolation.

Study 2

The procedures used were in many ways identical to those employed in Study 1. There were, however, several important differences that should be noted. First, a control condition was included, in which subjects received neither the self-focusing nor the task-focusing manipulations. Three studies (Brockner & Hulton, 1978; Shrauger, 1972; Shrauger & Terbovic, 1976) have shown no SE differences in performance on this task in control conditions. However, the addition of this condition made it possible to combine factorially the task-focusing and self-focusing variables. Thus, this experiment consisted of a three-factor design (SE × Task Focus × Self-Focus).

Second, the specific operationalization of self-focused attention was altered somewhat to provide further converging evidence on the joint effects of SE and self-focused attention on performance. That is, only the mirror, and not the video camera, was used in this study.

Third, in addition to responding to the SE inventory, subjects completed an individual difference measure of self-consciousness—the scale recently developed by Fenigstein et al. (1975)—prior to performing the concept formation task. If the interfering effects of self-focusing stimuli and the enhancing effects of task focusing for low SEs are related to the higher self-consciousness of low SEs, then highly self-conscious subjects (high SCs) should respond like low SEs. Thus, the mirror should impair and the task-focus instructions should bolster mainly the high SCs. Parallel findings for low SEs and high SCs (and for high SEs and subjects low in self-consciousness, low SCs) would provide converging evidence that performance was influenced by attentional processes.

Method

Participants

The 119 participants were drawn from an introductory psychology class at a New York state college. All received extra course credit for their participation. Of the 119 subjects, seven were eliminated for failure to understand the instructions. The proportion of eliminated subjects was not related to any independent variables. The final sample consisted of 112 subjects (50 male, 62 female) who were unequally distributed across the various experimental conditions.

Materials

Self-esteem inventory. The scale was identical to the one used in Study 1. The mean score for all subjects was 68.41 ($SD = 9.57$). Subjects were divided into three groups. The upper third of the distribution were labeled high SEs ($M = 79.05$, $SD = 5.49$); the middle third, medium SEs ($M = 68.21$, $SD = 2.59$); and the bottom third, low SEs ($M = 57.97$, $SD = 4.21$).

Self-consciousness inventory.[3] Subjects were asked to indicate on 6-point scales (minimum = 0, maximum = 5) how much each statement was characteristic of them (e.g., "I reflect about myself a lot," "I'm usually aware of my appearance"). Eighteen items were phrased so that a higher score represented higher self-consciousness, and three items were worded so that lower scores reflected higher self-consciousness. For the latter items each score was subtracted from the maximum rating of 5. The scores for each subject were then summed over the 21 items (minimum = 0, maximum = 105), such that higher scores represented greater self-focused attention. The average score for all subjects was 64.72 ($SD = 11.81$). A median split was employed to classify subjects as high or low SCs (M for high SCs = 73.9, $SD = 6.56$; M for low SCs = 54.87, $SD = 7.46$). As expected, SE and self-consciousness were negatively related, $r(110) = -.30$, $p < .01$.

Concept formation task. The task was the same one used in Study 1.

Procedure

As in Study 1, subjects were told that the study would consist of several parts. In the first part they were taken to a small cubicle and asked to complete the SE and self-consciousness inventories. Afterwards they were led to another cubicle to perform the concept formation task. When the instructions and practice trial were completed, the experimenter introduced the focus-of-attention manipulations. In the two self-focus conditions, subjects were told:

> One more point. Since we are interested in how people go about forming concepts, the experimenter will be watching you as you are working

[3] The content of 2 items from this 23-item scale was very closely related to several procedural aspects of this study. Specifically, they were, "One of the last things I do before I leave my house is look in the mirror," and "I have trouble working when someone is watching me." To prevent undue suspicions from arising, these items were deleted from the questionnaire.

Table 2
Task Performance in Relation to Self-Esteem, Study 2

Self-focus	Task focus	Self-esteem: High *n*	High *M*	High *SD*	Medium *n*	Medium *M*	Medium *SD*	Low *n*	Low *M*	Low *SD*
Present	Present (combined)	7	4.43	2.56	9	2.00	1.76	10	4.40	3.20
	Absent (mirror only)	6	2.50	2.63	14	3.93	2.40	9	7.22	5.81
Absent	Present (task only)	13	3.00	3.04	8	2.88	1.69	8	1.75	1.20
	Absent (control)	11	4.18	1.99	7	3.86	1.81	10	3.60	2.15

Note. Performance was measured by errors. Scores ranged from 0 to 20. Thus, the higher the score, the poorer the performance.

on the task. Also, a mirror will be placed in front of you, making it easier for the experimenter to observe your behavior more closely.

With that, the experimenter placed the large mirror directly in front of the subject. In the two task-focus conditions, subjects were given the same task-focus instructions used in Study 1. Essentially, subjects were randomly assigned to one of four focus-of-attention treatments: (a) mirror condition—subjects were only exposed to the mirror manipulation; (b) task condition—subjects were only exposed to the instructions designed to increase task-focused and decrease self-focused attention; (c) combined condition—subjects received the task-focus instructions first, followed by the introduction of the mirror; and (d) control condition—subjects received neither attentional focus manipulation. As in Study 1 the experimenter was not aware of the subjects' SE or self-consciousness scores.

After completing the task subjects made ratings of their attentional focus and anxiety while performing the task. In addition, they were asked to write their hypotheses and suspicions about the experiment, none of which was related to the true purpose of the study. Finally, all participants were fully debriefed.

Results and Discussion

Performance

Performance in relation to self-esteem. A preliminary t test yielded no effect as a function of sex of subject. Thus, all data were analyzed with a three-factor unweighted-means analysis of variance.[4] The summary data for this measure are presented in Table 2. The analysis revealed a significant main effect for task-focus instruction, $F(1, 100) = 3.94$, $p < .05$, with subjects performing better when instructed to focus on the task. Although the SE × Task Focus interaction was not significant ($p < .15$), simple effects analyses revealed that the task-focus instructions only had a significant impact on the low SEs, $F(1, 100) = 5.52$, $p < .025$. High SEs were unaffected by the task-focus instructions ($F < 1$), and medium SEs only marginally benefited from being told to focus on the task ($p < .20$). As additional evidence that primarily low SEs were aided by the task-focus instructions, there was a marginal *inverse* relationship between SE and performance in the no-task-focus conditions, $F(2, 100) = 2.36$, $p < .10$, whereas no such tendency appeared in the task-focus conditions.

The only other significant effect to emerge was a SE × Self-Focus interaction, $F(2, 100) = 3.93$, $p < .025$. Simple effects analyses revealed that only the performance of the low

[4] As in Study 1 the SE scores (within each SE level) did not differ by condition. Also, two experimenters (one male and one female) collected the data in this study. A preliminary t test showed that subjects made more errors in the presence of the female experimenter, $t(110) = 2.56$, $p < .025$. However, there was no Sex of Subject × Sex of Experimenter interaction. Moreover, since the experimenters ran approximately equal proportions of subjects in each condition, the data were collapsed across the experimenter variable.

Table 3
Task Performance in Relation to Self-Consciousness, Study 2

Self-focus	Task focus	Self-consciousness: High *n*	High *M*	High *SD*	Low *n*	Low *M*	Low *SD*
Present	Present (combined)	11	3.64	3.25	15	3.53	2.50
	Absent (mirror only)	13	7.08	5.01	16	2.69	1.86
Absent	Present (task only)	18	3.00	2.56	11	2.00	1.86
	Absent (control)	16	4.00	1.50	12	3.75	2.55

Note. Performance was measured by errors. Scores ranged from 0 to 20. The higher the score, the poorer the performance.

SEs was adversely affected by the mirror, $F(1, 100) = 9.95$, $p < .01$. Medium and high SEs performed equally well in the self-focus and no-self-focus conditions. In addition, only a relationship between SE and performance was observed in the self-focus conditions, $F(2, 100) = 4.68$, $p < .01$, with low SEs performing significantly worse than medium and high SEs (who did not differ). In the no-self-focus conditions, there was a nonsignificant ($F < 1$) inverse relationship between SE and performance.

Performance in relation to self-consciousness. Summary data for this analysis are presented in Table 3. The analysis of variance revealed a significant main effect for self-consciousness, $F(1, 104) = 6.68$, $p < .025$, with high SCs making more errors than low SCs. A significant triple interaction, $F(1, 104) = 5.27$, $p < .025$, revealed that high SCs did particularly worse than low SCs in the mirror condition.

Of course, the main effect for task focusing was still present, $F(1, 104) = 5.96$, $p < .025$. Moreover, simple effects analyses showed that only the high SCs benefited from the task-focus instructions, $F(1, 104) = 8.21$, $p < .01$, paralleling the earlier findings for low SEs. The performance of the low SCs (like that of the high SEs) was not affected by the task-focus instructions.

Fenigstein et al. (1975) reported that the self-consciousness scale consisted of three separate factors: private self-consciousness, public self-consciousness, and social anxiety.[5] They reported that

> Private self-consciousness was concerned with attending to one's inner thoughts and feelings, public self-consciousness was defined by a general awareness of the self as a social object that has an effect on others, and social anxiety was defined by a discomfort in the presence of others. Public and private self-consciousness refer to a *process* of self-focused attention; social anxiety refers to a *reaction* to this process. (p. 523)

Accordingly, subjects were classified as high or low, with a median split for each of these three components of self-consciousness. Three three-factor analyses of variance were performed, with private self-consciousness, public self-consciousness, social anxiety, and the two focus-of-attention factors as independent variables. Interestingly, individual differences in private and public self-consciousness produced no main effects and did not interact with either of the attentional focus manipulations. However, subjects high in social anxiety (high SAs) performed significantly worse than low-social-anxiety participants (low SAs), $F(1, 104) = 6.61$, $p < .025$. Simple ef-

[5] Although Turner et al. (1978) reported negative correlations between SE and each component of self-consciousness, only a correlation between SE and social anxiety was obtained in the present study, $r(110) = -.54$, $p < .01$.

Table 4
Self-Focused Attention Index, Study 2

Self-focus	Task focus	Self-esteem: High *n*	High *M*	High *SD*	Medium *n*	Medium *M*	Medium *SD*	Low *n*	Low *M*	Low *SD*
Present	Present (combined)	7	37.14	24.87	9	19.11	25.58	10	39.10	26.22
	Absent (mirror only)	6	12.17	31.32	14	24.14	28.05	9	46.00	20.99
Absent	Present (task only)	13	28.15	32.15	8	23.50	26.46	8	20.25	18.72
	Absent (control)	11	27.64	39.77	7	4.57	25.98	10	39.80	26.68

Note. Index was summed over four questions. Scores ranged from −31 to 92, with higher scores reflecting greater self-focused attention.

fects analyses demonstrated that this was particularly true in the self-focus conditions, $F(1, 104) = 6.92$, $p < .025$. When the mirror was not present, high and low SAs did not reliably differ, $F(1, 104) = 1.01$. In addition, a triple interaction effect emerged, the nature of which was somewhat unexpected. Specifically, it was only in the control condition that high SAs did not make more errors than low SAs, $F(1, 104) = 4.42$, $p < .05$.

Focus-of-Attention Manipulation Checks

Subjects completed the following items on the postexperimental questionnaire (once again, all questions were completed on 41-point scales, with higher scores representing greater endorsement of the statement): "*While you are completing the task,* (a) how self-conscious did you feel? (b) How much were you aware of the experimenter's presence? (c) How much did you find yourself thinking about how well or how badly you seemed to be doing? and (d) How much did you find yourself completely concentrating on the task and nothing else? That is, how much would you say you were concentrating on the task to the exclusion of any other thoughts?" All items correlated with Item 1, the most direct measure of self-consciousness ($r_{1 \text{ and } 2} = .52$; $r_{1 \text{ and } 3} = .40$; $r_{1 \text{ and } 4} = -.30$; $df = 110$; $p < .01$ in all cases). Therefore, a self-focused attention index was computed for each subject by summing Items 1, 2, and 3 and subtracting Item 4. Higher scores corresponded to greater amounts of self-focused attention.

A three-factor unweighted-means analysis of variance was performed, for which the summary data can be found in Table 4. From inspecting the data one finds that the scores tend to be aligned with the error data presented in Table 2. The only anomalous finding occurred in the medium SE–control condition. There the self-focused attention scores were unusually low and not at all consistent with the error data observed in that cell. The only significant effect to emerge was a main effect for SE, $F(2, 100) = 3.37$, $p < .05$, with low SEs most and medium SEs least self-focused. However, several simple effects paralleled those found for the performance data. For example, in the no-task-focus conditions low SEs were considerably more self-focused than medium and high SEs, $F(2, 100) = 4.51$, $p < .025$. In the task-focus conditions, there were no differences between SE groups. Also, in the self-focus conditions, low SEs were marginally more self-focused than medium or high SEs, $F(2, 100) = 2.52$, $p < .10$. No such trend emerged in the no-self-focus condition.

Additional evidence of the success of the task-focus variable on the low SEs stems from the results on the measure of task-focused attention (Item 4). Just as the performance data showed that primarily low SEs were af

fected by the task-focus factor, the analyses on this measure revealed that the low SEs were most influenced by this manipulation. While the SE × Task Focus interaction was only a trend ($p < .11$), simple effects analyses showed that in the no-task-focus conditions, low SEs were considerably less task focused than medium and high SEs, $F(2, 100) = 3.40$, $p < .05$, whereas all three groups were equally task focused in the task-focus conditions ($F < 1$). Low SEs also tended to be more task focused in the task-focus than in the no-task-focus conditions, $F(1, 100) = 3.34$, $p < .07$, whereas medium and high SEs were not (both Fs < 1).

As further evidence of the role of attentional processes on performance, there was a significant correlation between subjects' scores on the self-focused attention index and the number of errors they committed, $r(110) = .32$, $p < .01$.

General Discussion

Summary and Replications

Both studies demonstrated that (only) the performance of the low SEs was quite malleable across conditions. In Study 2 the self-focusing stimulus caused low SEs to perform worse than (a) medium and high SEs who were exposed to the same stimulus and (b) low SEs who were not exposed to the mirror. In Study 1 low SEs only performed more poorly than high SEs in the presence of the self-focusing stimuli. These data replicate Brockner and Hulton (1978) and Shrauger (1972) by showing that self-focusing stimuli (operationalized differently here than in the aforementioned studies) can exacerbate or activate performance decrements for low SEs.

Alternatively, task focusing seemed to have positive effects primarily for the low SEs. In Study 2 the task-focus instructions caused low SEs to perform better than low SEs who were not provided with such instructions, whereas medium and high SEs were unaffected by this manipulation. In Study 1 the internal analysis showed that low SEs who reported high levels of task concentration made significantly fewer errors than low SEs whose task concentration was low. That analysis also showed that high SEs did equally well, regardless of their task concentration level. Both studies thus confirm Brockner and Hulton's (1978) finding that low SEs especially benefit from increased task-focused attention. However, there are some important differences between the present results and Brockner and Hulton's data that should be made explicit. Brockner and Hulton discovered that low SEs made significantly fewer errors than high SEs when both groups were instructed to focus on the task. Low SE–task condition subjects, however, did not perform significantly better than low SEs in their control condition, although the means were in that direction. In the present studies the finding that low SEs performed better than high SEs when task focusing was high was not replicated (i.e., in the task condition in Study 2 and the internal analysis of SE and task concentration in Study 1), although the means were in that direction in both studies. However, unlike Brockner and Hulton, both of the studies reported here demonstrated that low SEs who were task focused performed significantly better than less task-focused low SEs.

That the self-focusing and task-focusing stimuli generally produced replications of previous results is all the more striking given the differences in subject populations. As suggestive evidence of the heterogeneity of the samples, the subjects in the present Study 2 overall made 71% more errors than Brockner and Hulton's subjects, while the Study 1 sample made 32% more errors. However, in spite of these *overall* performance differences, the relative differences in performance *between* conditions were in most cases remarkably consistent across all three studies.

Extensions

The combined conditions, which were extensions of the Brockner and Hulton (1978) study, produced inconsistent results. The low SEs' performance in Study 2 demonstrated that when the task-focus instructions work they can actually neutralize the effects of the mirror. From the low SE column of Table 2 it

can be seen that *relative to the control condition,* performance was impaired in the mirror condition and enhanced in the task condition. The low SEs' performance in the combined condition, however, was nearly identical to that found in the control condition. It is interesting that the task-focus instructions could functionally neutralize the mirror, given the relative *objective* salience of the two stimuli. That is, the task-focus instructions consisted of several sentences that were only read to subjects right before they started to work on the task. The large mirror, however, was a very salient self-focusing stimulus that remained present throughout the course of the experiment. One could reasonably argue that the possible ameliorating effects of the task-focus instructions were provided a fairly conservative test when combined with the omnipresent mirror. That the instructions actually neutralized the mirror suggests their potential utility for low SEs in situations that are likely to cause high self-consciousness, poor performance, and perhaps exacerbated low SE.

Of course, the above remarks must be tempered by the combined condition results found in Study 1. However, given that the task-focus instructions produced no beneficial effects for low SEs when administered in isolation (i.e., in the task condition of Study 1), it is perhaps not surprising that they were ineffective when used in conjunction with the mirror and video camera.

Attentional Focus and Performance

One weakness of the Brockner and Hulton (1978) study was its failure to provide highly convincing evidence that the low SEs' performance was influenced by attentional processes. Fortunately, there is now considerable evidence to suggest that for low SEs, (a) the manipulations had their intended effects on attentional focus (except for the task-focus instructions in Study 1) and (b) attentional focus influenced performance. First, both studies showed that self-focusing stimuli had their most pronounced effects on the self-focused attention of low SEs. Manipulation checks from Study 2 (see Table 4) showed that when the data from the mirror and combined conditions were taken together, low SEs were more self-focused than medium and high SEs. The data from Study 1 also revealed that low SEs were more self-focused than high SEs in the presence of the mirror and video camera.[6] Second, there was evidence in Study 2 that the task-focus instructions reduced the low SEs' self-consciousness and increased their task-focused attention. Specifically, when subjects were not told to task focus, low SEs were significantly more self-conscious and less task focused than medium and high SEs. However, when subjects were instructed to concentrate on the task, there were no self-focusing or task-focusing differences between the three SE groups.

Given that the manipulations generally had their desired effects, what evidence was there that attentional focus affected performance? First, recall that in Study 2 there was a significant correlation between subjects' scores on the self-focused attention index and the number of errors they committed. Furthermore, from Study 2 the relationships between the individual difference measure of self-consciousness, the attentional manipulations, and performance were particularly illuminating. As stated earlier, if the performance of the low SEs in the various conditions is related to their higher self-consciousness, then one should find similar performance data for low SEs and high SCs, and for high SEs and low SCs. Both predictions received firm support. Like low SEs, *only* high SCs (a) were impaired by the self-focusing stimulus, particularly in the mirror condition, and (b) improved when instructed to task focus.

In Study 1 the task-focus manipulation did not have its desired effect on the low SEs' focus of attention. Given the present position that attention influences performance, it is not surprising that the task-focus instructions had no appreciable effect on their performance. Therefore, the most parsimonious explanation of the null effect of the task-focus

---

[6] These data serve to extend Carver and Scheier's (1978) recent verification that mirrors cause increased self-focused attention. That is, their hypothesis would seem especially true for low SE individuals.

instructions on the low SEs' performance may be the failure of the manipulation to influence attention as intended, rather than the ego involvement hypothesis posed earlier. Note that in studies showing independent evidence of the task-focus instructions' effect on the low SEs' attention (Study 2; Brockner & Hulton, 1978), the performance of the low SEs improved significantly. Moreover, even when the task-focus instructions did not influence the low SEs' attention (Study 1) the internal analysis suggested that attentional factors affected their performance.

Self-Esteem, Attentional Focus, Anxiety, and Performance

If the low SEs' attentional focus influenced their performance, one still needs to consider the theoretical bases for this relationship. Two hypotheses were proposed earlier. First, if low SEs are too self-preoccupied, they simply will not be able to devote adequate attention to the task. Second, self-focused attention may be especially anxiety provoking for low SEs, leading to performance decrements on complex tasks like the one used in the present studies. With no attempt to discount the first possibility (indeed, the two may be complementary rather than competing hypotheses), evidence from these and other studies suggests that (a) self-focusing stimuli increased the low SEs' anxiety in both studies, whereas the task-focus instructions decreased their anxiety in Study 2, and (b) anxiety, in turn, mediated task performance.

Low SE, self-focused attention, and anxiety. On theoretical grounds there is good reason to believe that the self-focusing stimuli increased the low SEs' anxiety. It seems reasonable that the low SEs entered the test situation with higher levels of anxiety than high SEs. Indeed, researchers have consistently discovered inverse relationships between SE and general or test anxiety (Crandall, 1973). Moreover, recent research by Scheier and Carver (1977) showed that self-focused attention serves to intensify an individual's affective or emotional experience. Thus, it would appear that low SEs should have become even more aware of their anxiety when self-focused. In fact, both studies showed that low SEs reported greater anxiety than high SEs in the presence of self-focusing stimuli.[7] In Study 2 this tendency was expressed as a SE × Self-Focus interaction, $F(2, 100) = 3.32$, $p < .05$, such that in the self-focus, but not in the no-self-focus, conditions there was an inverse relationship between SE and anxiety. In Study 1 the t test showed that low SEs were also more anxious than high SEs in the video–mirror and combined conditions, $t(57) = 1.61$, $p < .06$, one-tailed, but not in the absence of the self-focusing stimuli (task condition, $t < 1$).

Furthermore, recall that Fenigstein et al.'s (1975) social anxiety subscale measures the extent to which people become anxious when self-aware. As additional evidence of the self-focused attention – anxiety link for low SEs, there was a pronounced inverse correlation between SE and social anxiety in Study 2 ($r = .54$).[8]

Self-focused attention, anxiety, and performance. There is also evidence that the anxiety that results from self-focused attention impaired performance on the task used in the present studies. Specifically, it was found that in the self-focus conditions in Study 2, subjects high in social anxiety performed significantly worse than low SAs. However, in the no-self-focus conditions, high and low SAs did not differ. Furthermore, test anxiety researchers (Geen & Gange, 1977; Wine, 1971) have reported that the presence of an audience produces additional performance decrements for highly test-anxious subjects. Given that audiences increase self-focused attention (Carver & Scheier, 1978), it is possible that they increased the already high anxiety awareness level of test-anxious

---

[7] Subjects were asked, "While you were completing the task, how anxious did you feel?" Their answers were recorded on 41-point scales.

[8] While SE and social anxiety are inversely related, the nature of the anxiety produced by the low SEs' self-focused attention is still unclear. One possibility is that the anxiety is due to evaluation apprehension. This interpretation seems plausible, given that subjects high in social anxiety seem to be more concerned or worried about being evaluated by others (Turner, 1977).

subjects (Scheier & Carver, 1977), leading to additional performance decrements.

Just as the self-focusing stimuli seemed to intensify the perceived anxiety levels of low SEs, it is possible that the task-focus instructions reduced felt anxiety. In Study 2 low SEs in the task condition did report lower levels of anxiety than low SE–control condition subjects, $t(16) = 2.26$, $p < .05$, two-tailed. As already noted, low SE–task subjects made fewer errors than low SE–controls. Moreover, the test anxiety literature shows that manipulations designed to increase task-focused and decrease self-focused attention have caused dramatic improvements in the performance of highly test-anxious subjects (Sarason, 1958, 1973, 1975). It may be that their performance increments are accompanied or even caused by decreases in their usually overinflated levels of felt anxiety.

Conclusion

The present studies generally confirm the hypothesis that the task performance of low SEs is inversely related to their self-focused attention. To address an important practical issue, future investigators should consider the effect of the low SEs' performance on their subsequent SE. The implication set forth earlier was that subsequent SE should covary with task performance. However, the relationship is undoubtedly more complex. For example, how much performance improvement is necessary to bolster the self-evaluation of low SEs? How long lasting are such effects? In short, under what conditions will the low SEs' performance affect their subsequent SE? To address an important theoretical issue, future researchers should provide additional evidence of the influence of attentional processes on performance. A promising technique was recently employed by Diener and Dweck (1978), who unobtrusively assessed their subjects' task-focusing and self-focusing behaviors.

References

Brockner, J., & Hulton, A. J. B. How to reverse the vicious cycle of low self-esteem: The importance of attentional focus. *Journal of Experimental Social Psychology*, 1978, *14*, 564–578.

Carver, C. S., & Scheier, M. F. Self-focusing effects of dispositional self-consciousness, mirror presence, and audience presence. *Journal of Personality and Social Psychology*, 1978, *36*, 324–332.

Coopersmith, S. *The antecedents of self-esteem*. San Francisco: Freeman, 1967.

Crandall, R. The measurement of self-esteem and related constructs. In J. R. Robinson & P. R. Shaver (Eds.), *Measures of social psychological attitudes*. Ann Arbor, Mich.: Institute for Social Research, 1973.

Diener, C. I., & Dweck, C. S. An analysis of learned helplessness: Continuous changes in performance, strategy, and achievement cognitions following failure. *Journal of Personality and Social Psychology*, 1978, *36*, 451–462.

Duval, S. Conformity on a visual task as a function of personal novelty on attitudinal dimensions and being reminded of the object status of self. *Journal of Experimental Social Psychology*, 1976, *12*, 87–98.

Duval, S., & Wicklund, R. A. *A theory of objective self-awareness*. New York: Academic Press, 1972.

Fenigstein, A., Scheier, M. F., & Buss, A. H. Public and private self-consciousness: Assessment and theory. *Journal of Consulting and Clinical Psychology*, 1975, *43*, 522–527.

Geen, R. G., & Gange, J. J. Drive theory of social facilitation: Twelve years of theory and research. *Psychological Bulletin*, 1977, *84*, 1267–1288.

Hamachek, D. *Encounters with the self*. New York: Holt, Rinehart & Winston, 1971.

Ickes, W. J., Wicklund, R. A., & Ferris, C. B. Objective self-awareness and self-esteem. *Journal of Experimental Social Psychology*, 1973, *9*, 202–219.

Janis, I. Personality correlates of susceptibility to persuasion. *Journal of Personality*, 1954, *22*, 504–518.

Mischel, W., Ebbesen, E., & Zeiss. A. R. Determinants of selective memory about the self. *Journal of Consulting and Clinical Psychology*, 1976, *44*, 92–103.

Sarason, I. The effects of anxiety, reassurance, and meaningfulness of material to be learned in verbal learning. *Journal of Experimental Psychology*, 1958, *56*, 472–477.

Sarason, I. Experimental approaches to test anxiety: Attention and the uses of information. In C. D. Spielberger (Ed.), *Anxiety: Current trends in theory and research* (Vol. 2). New York: Academic Press, 1972.

Sarason, I. Test anxiety and cognitive modeling. *Journal of Personality and Social Psychology*, 1973, *28*, 58–61.

Sarason, I. Test anxiety and the self-disclosing coping model. *Journal of Consulting and Clinical Psychology*, 1975, *43*, 148–153.

Scheier, M. F., & Carver, C. S. Self-focused attention and the experience of emotion: Attraction, repulsion, elation, and depression. *Journal of Personality and Social Psychology*, 1977, *35*, 625–636.

Shrauger, J. S. Self-esteem and reactions to being observed by others. *Journal of Personality and Social Psychology,* 1972, *23,* 192–200.

Shrauger, J. S., & Terbovic, M. L. Self-evaluation and assessments of performance by self and others. *Journal of Consulting and Clinical Psychology,* 1976, *44,* 564–572.

Spence, K. W., Farber, I. E., & McFann, H. H. The relation of anxiety (drive) level to performance in competitional and noncompetitional paired-associate learning. *Journal of Experimental Psychology,* 1956, *52,* 296–305.

Turner, R. G. Self-consciousness and anticipatory belief change. *Personality and Social Psychology Bulletin,* 1977, *3,* 438–441.

Turner, R. G., Scheier, M. F., Carver, C. S., & Ickes, W. J. Correlates of self-consciousness. *Journal of Personality Assessment,* 1978, *42,* 285–289.

Wahl, C. Some antecedent factors in the family histories of 109 alcoholics. *Quarterly Journal of Studies on Alcohol,* 1956, *17,* 643–654.

Wicklund, R. A., & Duval, S. Opinion change and performance facilitation as a result of objective self-awareness. *Journal of Experimental Social Psychology,* 1971, *7,* 319–342.

Wine, J. Test anxiety and direction of attention. *Psychological Bulletin,* 1971, *76,* 92–104.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1971.

Received June 8, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 3, 462–467

Effects of Contrast and Generalization on the Attitude Similarity – Attraction Relationship

John J. Seta, Lenny Martin, and George Capehart
University of North Carolina at Greensboro

An experiment was performed examining the phenomena of contrast and generalization within the attitude similarity – attraction paradigm. Subjects read of the attitudes of two strangers (A and B) whom they had met and who had both initially agreed with these subjects on 50% of a number of topics. After a brief period of time, each subject again met the two strangers. During this second meeting, Stranger A continued to agree with the subject at a 50% rate, whereas Stranger B agreed with the subject at a 100% rate. In one condition, the two strangers were depicted as members of the same group, and in a second condition, the two strangers were depicted as members of different groups. The results indicated that after the second meeting, there was less of a difference between the attractiveness of Strangers A and B in the same-group condition compared to the different-group condition. In addition, within the different-group condition, Stranger A was liked less after the second meeting than after the first meeting, whereas within the same-group condition, Stranger A was liked more after the second meeting than after the first meeting. These results support the notion that the effects of contrast are accentuated when two individuals being rated are distinct entities (members of different groups), whereas generalization is accentuated when the two individuals are not distinct entities (members of the same group). The above results are interpreted within the Byrne–Clore reinforcement affect model.

In the area of attraction, a great deal of research has supported the notion that attitude similarity affects attraction (e.g., Byrne, 1969, 1971; Byrne & Nelson, 1965; Clore & Baldridge, 1968; Griffitt, 1971; Schachter, 1951). The effects of attitude similarity on attraction can best be described as a "positive linear function of the sum of the weighted similar attitudes divided by the total number of weighted similar and dissimilar attitudes" (Byrne, 1971, p. 71). The Byrne–Clore reinforcement affect model (e.g., Byrne & Clore, 1970) has been the most parsimonious account of the attitude similarity – attraction paradigm. Briefly, they propose (Byrne, 1971) that the affective response elicited by an individual (X) "is a positive linear function of the sum of the weighted positive reinforcements associated with X divided by the total number of weighted positive and negative reinforcements associated with X" (pp. 279–280).

The present article deals with how the weight or value of a positive reinforcer or punisher associated with X is affected by the presence of a second individual. That is, the attractiveness of a particular individual should depend not only on the reinforcers (e.g., agreements) and punishers (e.g., disagreements) obtained from that particular individual but also on the reinforcers and punishers obtained from other individuals in the situation. A situation in which an individual meets two strangers is briefly depicted. Individual X simultaneously meets two strangers (Stranger A and Stranger B) who both initially agree with X on 50% of a number of topics. After

The present investigation was partially supported by a grant from the University of North Carolina at Greensboro.

Requests for reprints should be sent to John J. Seta, Department of Psychology, University of North Carolina, Greensboro, North Carolina 27412.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3703-0462$00.75

a brief period of time, X again meets these two strangers and discusses a number of different, yet equally important, topics. During this second meeting, Stranger A continues to agree with X at a 50% rate, whereas Stranger B now agrees with X on 100% of the topics.

Opposing Predictions

The question that can be asked is "Would the attractiveness of each of the strangers depend on the presence of the other?" Further, would Stranger A be liked more by X after their first or after their second meeting? Likewise, would Stranger B be liked more after their second meeting? Predictions concerning this question can be derived from two sources. The first is the phenomenon of contrast (e.g., Helson, 1964). Since contrast occurs when the differences between two stimuli are accentuated, one would predict that X would like Stranger A less after their second rather than their first meeting, even though Stranger A's proportion of agreements has not changed. Stranger A's rate of agreement after the second meeting should seem less valued, since Stranger B has increased his or her agreements from 50% to 100%. In contrast, Stranger B should be liked more after the second meeting, since not only has the rate of agreement increased but the rate has also increased relative to that of Stranger A. Experiments by Mascaro and Graves (1973) and Griffitt (1971) support the notion of a contrast effect within the attitude similarity–attraction paradigm. Mascaro and Graves (1973) found that a second stranger's attraction was affected by the first stranger's rate of agreement.[1] That is, the second stranger was liked less as the first stranger's rate of agreement increased relative to that of the second stranger.

An opposing prediction can be made from the phenomenon of stimulus generalization. Since generalization occurs when different stimuli evoke a common response, one would predict that X would like Stranger A more after their second meeting than after their first meeting. Stranger B's 100% agreement rate should positively affect the value of Stranger A's 50% rate, since A and B's agreement rate might be averaged if the two individuals are perceived as a unit. An experiment by Wyant, Lippert, Wyant, and Moring (1977) seems to support the concept of generalization within the attitude similarity–attraction paradigm. They found that the effects of one stranger's evaluation generalized to a second stranger. In this experiment, the second stranger was depicted as being similar to the first.

Contrast and Generalization

The phenomena of contrast and generalization may occur under somewhat different conditions. Contrast should predominate when the individuals being judged are seen as distinct entities or members of different groups, whereas generalization should predominate when the individuals being judged are not distinct, being members of the same group. In essence, the attractiveness of a particular individual should depend not only on the reinforcers and punishers obtained from that particular individual but also on the reinforcers and punishers obtained from other individuals. If there is an association, then the affective response associated with each of the individuals should generalize. That is, there should be a tendency to average the reinforcers obtained from each individual and therefore minimize differences between them. However, if the two individuals come from different groups, a contrast effect should be pronounced. That is, the reinforcers and punishers and, hence, the affective response associated with one individual should be compared to the affective response associated with the second individual. Given that there are differences in the absolute magnitude of rein-

[1] In this study, subjects first met one stranger and then at some later time met a second stranger. Specifically, their results supported the notion of what can be labeled *successive contrast*. However, their results do not support the existence of what can be labeled *simultaneous contrast*. Simultaneous contrast would be supported if the subjects met the two strangers simultaneously and a contrast effect was obtained. The present experiment tests the notion of simultaneous contrast within the attitude similarity–attraction paradigm.

forcement obtained from the two individuals, a contrast effect would accentuate these differences.

The present experiment was performed in an attempt to answer the following questions: First, does the evaluation that is given to an individual depend on the percentage of agreements (rewards) given by a second individual? Second, under what conditions would a contrast effect be accentuated and under what conditions might a generalization effect be accentuated? In this experiment a subject initially meets two strangers (A and B) who independently agree with the subject on 50% of the topics. After rating both strangers, the subjects were presented with additional attitudes of the two strangers on six new topics and again were asked to independently rate the attractiveness of the two strangers. During this second meeting, Stranger A continued to agree with the subject at a 50% rate, whereas Stranger B now agreed with the subject at a 100% rate. In the first condition (different-group condition), Strangers A and B were depicted as members of different groups (one stranger was a math major, and the other was a drama major), whereas in a second condition (same-group condition), both strangers were depicted as members of the same group. (Both strangers were math majors or both strangers were drama majors.) As earlier mentioned, the effects of contrast should be accentuated when an individual concentrates on the distinctive properties of others (different-group condition), whereas generalization should be accentuated when an averaging of differences is made (same-group condition). That is, after the second meeting there should be less of a difference between the attractiveness of Strangers A and B in the same-group condition compared to the different-group condition. Differences should be accentuated in the different-group condition (contrast predominates) and minimized in the same-group condition (generalization predominates).

Method

Subjects

Thirty-six students (31 females and 5 males) from two junior-level psychology classes were randomly assigned to either the same- or different-group condition. There were 18 subjects (15 females and 3 males) in the different-group condition and 18 subjects (16 females and 2 males) in the same-group condition. Subjects were run in groups. One group consisted of 24 subjects (12 in the same-group condition and 12 in the different-group condition), and the second group consisted of 12 subjects (6 in the same-group condition and 6 in the different-group condition). All subjects volunteered their participation.

Experimental Design

A one between (same or different) and two within [Persons A or B and trials (first or second meeting)] repeated measures analysis of variance was performed. In the same-group condition, Strangers A and B were depicted as either both math or both drama majors, whereas in the different-group condition one stranger was depicted as a math major and the other as a drama major. The order of depiction was counterbalanced.

Procedure

Consistent with the phantom–other procedure (e.g., Byrne & McGraw, 1964), subjects first filled out a standard questionnaire defining their attitudes (e.g., Byrne & Nelson, 1965). Approximately 3 weeks after filling out the standard questionnaire defining their attitudes, all subjects were presented with typed instructions and the purported attitudes of Strangers A and B. After reading the instructions, subjects in both conditions read of the attitudes of Strangers A and B (both agreed on 50% of the six topics) and were then asked to rate each stranger separately. All subjects were asked to respond to three questions. Each question contained a 7-point scale, where 1 corresponded to like and 7 corresponded to dislike. The three questions were: (a) How much would you enjoy working further with this person? (b) How much did you like the stranger? and (c) How much would you like to continue a relationship with the stranger? After rating the strangers, all subjects were presented with additional attitudes of Strangers A and B. This time Stranger A agreed with the subject on 50% of the six new topics, whereas Stranger B agreed with the subject on 100% of the six new topics. All subjects were then asked by typed instructions to again rate Strangers A and B.

As a manipulation check, a preexperimental questionnaire was given to 12 subjects who were similar in academic standing to the subjects used in the present study. Using only information that was given to them by the experimenter, they were asked to answer two questions: (a) Who do you feel are more similar, two math majors or a math major and a drama major? (b) Who do you feel are more similar, two drama majors or a math major and a drama major? The phrasing of the above questions was counterbalanced. All subjects felt that the two math

majors were more similar than the math major and the drama major and that the two drama majors were more similar than the math major and the drama major. A postexperimental interview was also given to 10 subjects. They were asked if they had read the entire instructions and attitudes of the strangers. In addition, they were asked if they knew that they were rating two different individuals and if they remembered if they had rated two individuals who had the same major or two individuals who had different majors.

Results and Discussion

A between (same or different)–within [Persons A or B and trials (first or second meeting)] repeated measures analysis of variance was performed. Since all three measures were nearly identical, the total of all three measures was used in the analysis. (See Table 1 for the means of each of the experimental conditions.) The analysis revealed a significant Group (same or different) × Person (Strangers A or B) × Trials interaction, $F(1, 34) = 55.30$, $p < .001$; a Person × Trial interaction, $F(1, 34) = 83.99$, $p < .001$; and a Group × Person interaction, $F(1, 34) = 5.93$, $p < .05$. In addition, consistent with the Byrne–Clore (1970) framework, a person main effect, $F(1, 34) = 87.32$, $p < .001$, and a trials main effect were obtained, $F(1, 34) = 106.43$, $p < .001$. All subjects that were asked stated that they had read all of the instructions and the attitudes of the strangers. In addition, the subjects were aware of whom they had rated (e.g., two math majors).

Results Supporting Contrast and Generalization

To interpret the significant interactions a Tukey (a) test was performed on all logical comparisons of paired means. (See Table 1 for a summary of the comparisons.) Within the different-group condition, Stranger A was liked less after the second meeting than after the first meeting ($p < .05$), whereas within the same-group condition, Stranger A was liked more after the second meeting than after the first meeting ($p < .05$). These results support the contrast–generalization interpretation, since within the different-group condition, contrast should predominate, whereas in the same-group condition, generalization should predominate. In the different-group condition, Stranger A's 50% rate of agreement should be contrasted to Stranger B's 100% rate, resulting in a reduced level of attraction for Stranger A. Within the same-group condition, Stranger B's 100% agreement rate should generalize and positively affect the value of Stranger A's 50% rate.

Although the comparisons did not reach significance, after the second meeting Stranger B was liked somewhat more in the different-group condition than in the same-group condition, whereas after the second meeting Stranger A was liked somewhat more in the same-group condition than in the different-

Table 1
Attraction Score Means for Each Condition

Person A		Person B	
Trial 1	Trial 2	Trial 1	Trial 2
Different-group condition			
11.50	13.61	13.55	7.66
Same-group condition			
13.33	11.55	12.61	10.00

Note. The higher the score, the less the attraction. Range of scores is 3–21. The following is a summary of the significant logical comparisons among paired means using the Tukey (a) test with the level of significance set at .05. A pooled error term was employed as suggested by Winer (1971). For comparisons between Trials 1 and 2 for Person A or Person B, the critical difference was 1.43. In the different-group condition, Person A was liked significantly less after the second than after the first meeting. In the same-group condition, Person A was liked significantly more after the second than after the first meeting. In both the same and different conditions, Person B was liked significantly more after the second than after the first meeting. None of the comparisons between the different- and same-group conditions reached significance. The critical differences for these comparisons was 4.75. For comparisons between Persons A and B within either the different- or same-group conditions, the critical difference was 1.26. During the first meeting, Person A was liked significantly more than Person B in the different-group condition. During the second meeting, Person B was liked significantly more than Person A in both the same- and different-group conditions.

group condition. This result is consistent with a contrast–generalization interpretation. In the different-group condition, contrast should predominate, whereas in the same-group condition, generalization should predominate.

Stranger B was liked more after the second meeting than after the first meeting in both the same- and different-group conditions ($p < .05$). In both the same-group and different-group conditions, Stranger B's attraction should have increased from the first to the second meeting because of Stranger B's increase in agreement rate (50%–100%). However, the increase in attraction from the first to the second meeting was greater for B in the different-group than in the same-group condition. (In the different-group condition the difference was 5.89, whereas in the same-group condition the difference was 2.61.) This should have been the case, since in the same-group condition, generalization should predominate, whereas in the different-group condition, contrast should predominate. In the same-group condition, Stranger B's attraction might have been lessened due to B's association with A (A's rate is lower than B's); in the different-group condition, Stranger B's attraction should have increased because of A's low rate of agreement. Finally, after the second meeting, there was less of a difference between the attractiveness of Strangers A and B in the same-group compared to the different-group condition. An analysis of variance on difference scores revealed that after the second meeting, there was less of a difference between the attractiveness of Strangers A and B in the same-group condition (28 difference) compared to the different-group condition (107 difference), $F(1, 34) = 67.9$, $p < .001$. This should have been the case, since differences should be accentuated in the contrast condition and minimized in the generalization condition.

The above results support the notion that the attractiveness of a particular individual should depend not only on the reinforcers and punishers obtained from that particular individual but also on the reinforcers and punishers obtained from other individuals. In addition, the magnitude of the reinforcers and punishers should depend on whether there is or is not a strong commonality between the two individuals. If there is an association, then the affective response associated with each of the individuals should generalize—There should be a tendency to average the reinforcers obtained from each individual. However, if the two individuals are perceived as distinct entities, a contrast effect should be pronounced.

Consensual Validation

As one can see from Table 1, after the first meeting, subjects liked Stranger A more than Stranger B in the different-group condition ($p < .05$). Although no a priori predictions were made concerning subjects' first ratings of Strangers A and B (both agreed with the subject on 50% of the topics), these results are quite interesting and explicable within Byrne & Clore's (1967) theoretical framework if one assumes that a 50% agreement rate is below that which the subjects are accustomed to receiving. Byrne and Clore (1967) argued that a possible reason why increasing rates of agreement result in increments in attraction is that attitudinal agreement supplies an individual with consensual validation. Hence, there should be greater invalidation the greater the number of groups disagreeing with the subject. Due to the above, subjects in the different-group condition should like Stranger B less than Stranger A because Stranger B's general disagreement results in a greater degree of consensual invalidation. Further, since the weight of consensual invalidation is greater for B than for A, there should be a greater probability of discounting the first stranger's general disagreement.

In sum, the results of the present experiment support the notion that both contrast and generalization must be considered when investigating the effects of attraction. In addition, these results suggest that the probability of obtaining a generalization effect is increased as the perceived commonality between two individuals is increased. Further, the probability of obtaining a contrast effect is increased as the perceived commonality between two individuals is decreased. The pres-

ent results indicate that certain personality types should enhance the contrast effect—Individuals who concentrate on individual differences should demonstrate a pronounced contrast effect when compared to individuals who concentrate on the similarities among people.

References

Byrne, D. Attitudes and attraction. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 4). New York: Academic Press, 1969.

Byrne, D. *The attraction paradigm*. New York: Academic Press, 1971.

Byrne, D., & Clore, G. L. Predicting interpersonal attraction toward strangers presented in three different stimulus modes. *Psychonomic Science,* 1966, *4,* 239–240.

Byrne, D., & Clore, G. L. Effectance arousal and attraction. *Journal of Personality and Social Psychology Monograph,* 1967, *6*(4, Whole No. 638).

Byrne, D., & Clore, G. L. A reinforcement model of evaluative responses. *Personality: An International Journal,* 1970, *1,* 103–128.

Byrne, D., & McGraw, D. Interpersonal attraction towards Negroes. *Human Relations,* 1964, *17,* 201–213.

Byrne, D., & Nelson, D. Attraction as a linear function of proportion of positive reinforcements. *Journal of Personality and Social Psychology,* 1965, *1,* 659–663.

Clore, G. L., & Baldridge, B. Interpersonal attraction: The role of agreement and topic interest. *Journal of Personality and Social Psychology,* 1968, *9,* 340–346.

Griffitt, W. Context effects in response to affective stimuli. *Personality: An International Journal,* 1971, *2,* 23–33.

Helson, H. *Adaptation-level theory: An experimental and systematic approach to behavior*. New York: Harper & Row, 1964.

Mascaro, G. F., & Graves, W. Contrast effects of background factors on the similarity–attraction relationship. *Journal of Personality and Social Psychology,* 1973, *25,* 346–351.

Schachter, S. Deviation, rejection and communication. *Journal of Abnormal and Social Psychology,* 1951, *46,* 190–207.

Winer, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1971.

Wyant, K. W., Lippert, J. A., Wyant, F. W., & Moring, D. G. The duration and generalization of attraction. *Journal of Research in Personality,* 1977, *11,* 347–355.

Received June 9, 1978 ■

Notice to Authors of the *Journal of Personality and Social Psychology*

Papers submitted to the *Journal of Personality and Social Psychology* after April 1, 1979 will be considered under new policies approved by the Publications and Communications Board. Effective with the 1980 volume of the journal, the following policies will apply:

The *Journal of Personality and Social Psychology* publishes original papers in all areas of personality and social psychology. It emphasizes empirical reports but may include specialized theoretical, methodological, and review papers. Beginning in 1980, the journal will be divided into 3 independently edited sections: ATTITUDES AND SOCIAL COGNITION, INTERPERSONAL RELATIONS AND GROUP PROCESSES, and PERSONALITY PROCESSES AND INDIVIDUAL DIFFERENCES. Authors should submit manuscripts to the appropriate section editor according to the following definitions:

- ATTITUDES AND SOCIAL COGNITION includes papers on attitudes dealing with such topics as the formation or change of beliefs and attitudes, measurement of attitudes, and the relation between attitudes and behavior. Papers on social cognition deal with the formation and utilization of knowledge about the social world and embrace such topics as social and person perception, attributional processes, and information processing.
- INTERPERSONAL RELATIONS AND SOCIAL INTERACTION focuses on the interaction between two or more people. Included in this section are papers on group structure; helping and aggression; group decision making and task performance; leadership; cooperation and competition; ingroup–outgroup relations, etc.
- PERSONALITY PROCESSES AND INDIVIDUAL DIFFERENCES encourages papers on all aspects of personality psychology as traditionally defined. This includes personality assessment, measurement, structure, and dynamics. All methodological approaches will be considered.

Section editors reserve the right to redirect papers among themselves as appropriate unless an author specifically requests otherwise. Rejection by one section editor is considered rejection by all, therefore a manuscript rejected by one section editor should not be submitted to another.

Authors should prepare manuscripts according to the *Publication Manual of the American Psychological Association* (2nd edition) and the "Guidelines for Nonsexist Language in APA Journals" (Publication Manual Change Sheet 2, *American Psychologist,* June 1977, pp. 487–494). Instructions on tables, figures, references, metrics, and typing (all copy must be double-spaced) appear in the Manual. Articles not prepared according to the Manual will not be reviewed.

Manuscripts should be submitted in quadruplicate (an original and 3 copies) to the appropriate section editor. Authors should keep a copy of the manuscript to guard against loss.

Blind review is optional and must be specifically requested when a manuscript is first submitted. In those cases, authors' names and affiliations must appear only on a separate title page, included with each copy of the manuscript. Footnotes containing the identity of authors or their affiliations must be on a separate sheet.

Mailing addresses for the section editors are as follows:

ATTITUDES AND SOCIAL COGNITION

April 1 through June 15:

Melvin Manis
c/o Anita DeVivo
Suite 700
1400 North Uhle Street
Arlington, Virginia 22201

after June 15:

Melvin Manis
Research Center for Group Dynamics
Institute for Social Research
University of Michigan
Ann Arbor, Michigan 48104

INTERPERSONAL RELATIONS AND GROUP PROCESSES

Ivan D. Steiner
Department of Psychology
University of Massachusetts
Amherst, Massachusetts 01003

PERSONALITY PROCESSES AND INDIVIDUAL DIFFERENCES

Robert Hogan
Johns Hopkins University
Department of Psychology
33rd and Charles Streets
Baltimore, Maryland 21218

Schizophrenia

Towards a New Synthesis

Edited by J. K. WING

This book provides a comprehensive and readable account of modern knowledge about schizophrenia. The authors are all recognized experts, outstanding in their various fields, and all have attempted to weigh the firm empirical findings from other approaches with the current state of research in their own specialty, rather than make extravagant theoretical or therapeutic claims. Progress in biochemistry has been made through investigations into the mode of action of certain drugs, seen to be effective: genetics, first approached through the hazardous means of twins studies, has now produced incontrovertible evidence for an inherited component in schizophrenia. But perhaps the greatest recent advances—many of them made in the UK and brought together here for the first time—concern the way in which social and psychological factors precipitate onset and relapse and influence the course of schizophrenic disorders. Implications for treatment, rehabilitation, home management and social work are discussed in full.

1978, 304 pp., $16.50 ISBN: 0-8089-1140-6

The Social Skills Basis of Psychopathology

Alternatives to Abnormal Psychology and Psychiatry

By E. LAKIN PHILLIPS, Ph.D.

CONTENTS: What Are Social Skills? Diagnosis, Psychometric Evaluations, and Social Skills: A Rapprochement. Types of Depressive Reactions on the Minnesota Multiphasic Personality Inventory and Their Relationship to Social Skills. On Presenting Complaints—Continuing the Discussion of Psychometric Screening and Social Skills. What Is Psychopathology? Training as a Therapeutic Modality: Implications for Social Skills. The "Inner" Problem: Feelings, Locus of Control, Biofeedback and Crisis Intervention. Social Skills and Psychotherapy. Law and Psychotherapy in the Light of Social Skills. Toward a Social Curriculum.

1978, 304 pp., $16.50 ISBN: 0-8089-1126-0

Send payment with order and save postage and handling charge.

Prices are subject to change without notice.

Grune & Stratton

A Subsidiary of
Harcourt Brace Jovanovich, Publishers
111 FIFTH AVENUE, NEW YORK, N.Y. 10003
24-28 OVAL ROAD, LONDON NW1 7DX

Choice.

And the Creation of Self.

New

from Harper & Row, a humanistic perspective on psychological growth.

FRANK BARRON

The Shaping of Personality

CONFLICT, CHOICE, AND GROWTH

In this personal, humanistic introduction to the study of personality, Barron focuses on four areas: Human Design—those characteristics and potentialities which are uniquely human; Human Development—the growth of personality from conception to mature individuality; Conflict—the psychic product of man's dualistic nature; Human Selves—the study of the person as a whole, distinct from either psychoanalytic or behaviorist viewpoints. Borrowing from gestalt psychology, pragmatism, and traditional clinical approaches, Barron sees *choice* as the decisive factor in human growth.

Available April, 1979. 384 pps. $15.95

To order your copy, please send your check or money order for $15.95 to Suite 5D, College Department. Instructors entitled to 10% discount. Add 94¢ handling and postage per copy plus sales tax where applicable.

Harper & Row

10 E. 53d Street, New York, NY 10022

Prices subject to change without notice. Prices quoted by Harper & Row are *suggested* list prices only and in no way reflect the prices at which the book may be sold by suppliers other than Harper & Row.

Look UP to Dorsey in '79

Understanding Social Psychology, Revised Edition

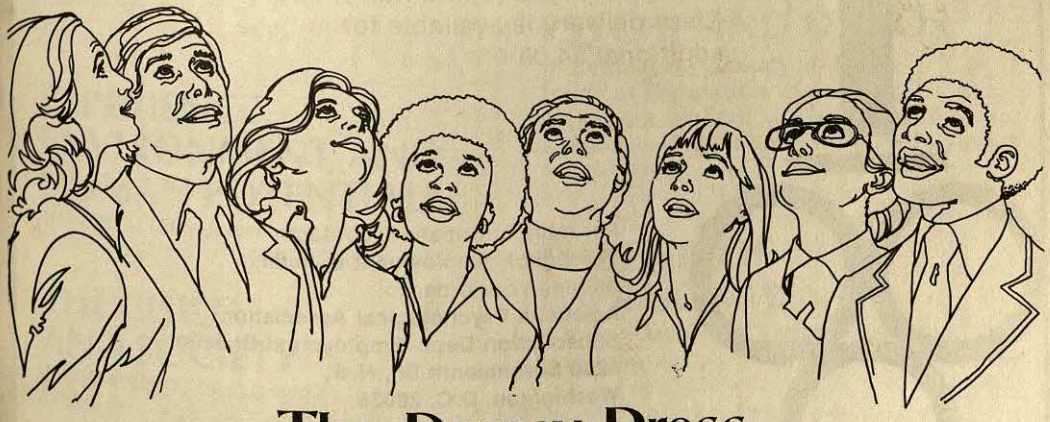
By **Stephen Worchel,** University of Virginia and
Joel Cooper, Princeton University

This text presents both traditional and contemporary experimental research and theory with the emphasis on the active areas of research. Experimental research is placed in the context of "critical incidents" from the real world or literature to help the student understand these events. The text begins with the reaction of individuals in the formation of attitudes, change of attitudes and the attribution of characteristics to self and others and then moves toward a consideration of the individual in a more macroscopic environment. **Teacher's Manual** and **Student Workbook.** *Just Published*

Personality: Strategies and Issues, Third Edition

By **Robert M. Liebert,** State University of New York at Stony Brook
and **Michael D. Spiegler,** Providence College

This text provides an accurate, respected survey of psychological theory and research about human personality. An even-handed view is taken toward the various approaches to the subject while at the same time presenting the strengths and weaknesses of each viewpoint. All the various theories of personality are seen as growing out of four basic strategies for studying personality including: theory, research, personality assessment, and personality change. **Instructor's Manual.** *Recently published*



The Dorsey Press
Homewood, Illinois 60430

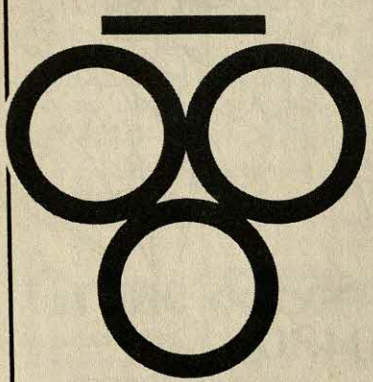
Examination copies for adoption consideration available on request. Please indicate course title and text presently used.

APA EMPLOYMENT BULLETIN

Possibly your key to greater career opportunities.

Each month 125 to 150 position openings are listed by geographical area and employment field for psychologists with a minimum of a Master's degree. Members may also publish availability notices. Coded identification numbers assure confidentiality. Ideal for employers, students, and Master's and Ph.D graduates.

A one year subscription is as little as $9.00 for members, $16.00 for non-members. If you're in a hurry, First Class delivery is available for an additional $4.00.



For more information, contact:
The Editor, Employment Bulletin,
or send your order to:
American Psychological Association
Subscription Dept.-Employment Bulletin
1200 Seventeenth St., N.W.
Washington, D.C. 20036

OXFORD UNIVERSITY PRESS

Publishers of Fine Books for Five Centuries

SOCIAL PSYCHOLOGY
A Contemporary Approach
LOUIS A. PENNER, University of South Florida
1978 400 pp.; 92 illus. $12.00
Complimentary Instructor's Manual

NONVERBAL COMMUNICATION
Readings with Commentary
Second Edition
Edited by **SHIRLEY WEITZ,** Graduate Faculty, The New School for Social Research
February 1979 448 pp.; 83 illus. cloth $13.95 paper $8.95

SEX ROLES
Biological, Psychological and Social Foundations
SHIRLEY WEITZ
1977 256 pp. cloth $10.95 paper $5.95

IMPLICIT PSYCHOLOGY
An Introduction to Social Cognition
DANIEL M. WEGNER, Trinity University, and **ROBIN R. VALLACHER,** Illinois Institute of Technology
1977 336 pp.; 31 illus. cloth $11.95 paper $6.00

EXPERIENCE, ENVIRONMENT, AND HUMAN POTENTIALS
HERBERT L. LEFF, University of Vermont
1978 544 pp. cloth $15.00 paper $8.00

THE CHILD AND SOCIETY
DAVID ELKIND, Tufts University
1978 352 pp. cloth $10.00 paper $3.95

PRINCIPLES AND METHODS OF SOCIAL PSYCHOLOGY
Third Edition
EDWIN P. HOLLANDER, State University of New York, Buffalo
1976 684 pp.; 60 photos, 86 figs. $13.95

MASTERS OF SOCIAL PSYCHOLOGY
Freud, Mead, Lewin, and Skinner
JAMES A. SCHELLENBERG, Indiana State University, Terre Haute
1978 136 pp. cloth $9.95 paper $2.50

THE MMPI
A Practical Guide
JOHN R. GRAHAM, Kent State University
1977 250 pp.; 15 illus. $10.95

AGGRESSION AND CRIMES OF VIOLENCE
JEFFREY H. GOLDSTEIN, Temple University
1975 (paper, 1976) 208 pp. cloth $9.95 paper $3.00

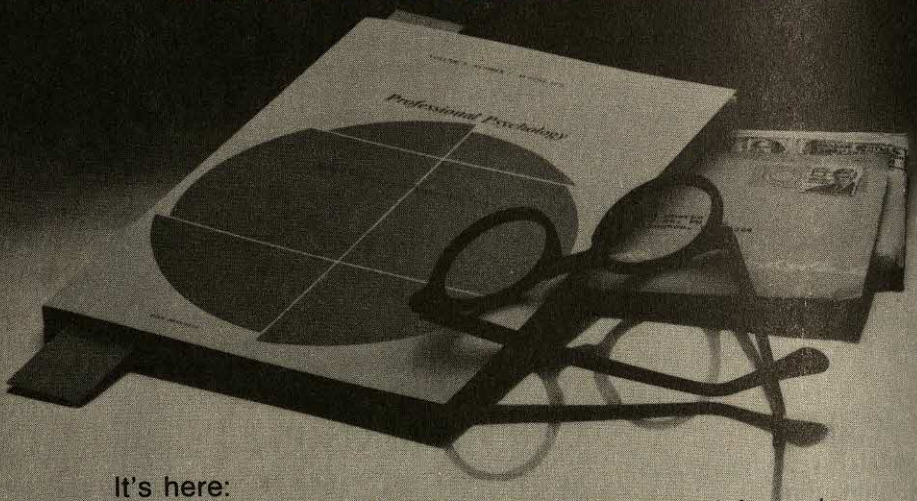
PSYCHE AND DEMOS
Individual Psychology and the Issues of Population
WARREN B. MILLER, American Institutes for Research; and **R. KENNETH GODWIN,** Oregon State University
1977 336 pp.; 14 line drawings cloth $12.00 paper $5.00

THE END OF IMPRISONMENT
ROBERT SOMMER, University of California, Davis
1976 (paper, 1977) 224 pp. cloth $9.95 paper $3.50

Prices and publication dates are subject to change.

Oxford University Press 200 Madison Avenue, New York, New York 10016

NOW: INFORMATION ABOUT THE LAW AND PSYCHOLOGY.



It's here:
Law and Professional Psychology. This special August 1978 issue of **Professional Psychology** brings you thirteen original articles full of fresh thought and constructive recommendations on the issues that vitally affect your day-to-day practice. You'll find precautions you may take to protect yourself as a professional and a plan for improving investigations of malpractice claims. Expert witness testimony, jury selection, civil commitment, confidentiality and privilege, minor's consent to treatment, and the use of psychological devices are critically examined.

Send your order today. And soon you'll receive this special issue.

Single copies of the issue are available for $5 each. All orders $25 or less must be prepaid. Make and send checks payable to:

American Psychological Association
Order Department
1200 17th Street, N.W.
Washington, D.C. 20036

Enclosed is $_____for__copies of **Professional Psychology's** special issue—Law and Professional Psychology—at $5.00 each.

NAME _____

ADDRESS _____

CITY _____

STATE _____ ZIP CODE _____

PP-8

(Continued from front cover)

Sex Differences in the Allocation of Pay 433
Charlene M. Callahan-Levy and Lawrence A. Messé

Self-Esteem, Self-Consciousness, and Task Performance: Replications, Extensions, and Possible Explanations 447
Joel Brockner

Effects of Contrast and Generalization on the Attitude Similarity – Attraction Relationship 462
John J. Seta, Lenny Martin, and George Capehart

Ad Hoc Reviewers 423

Notice to Authors of the *Journal of Personality and Social Psychology* 468

Student-involving introductory texts

NEW

EXPERIENCING SOCIAL PSYCHOLOGY: Readings and Projects

Ayala Pines and **Christina Maslach,** both of the University of California, Berkeley

This integrated reader/project manual features research articles and popular readings on current topics, classical problems, and new content areas such as social pathology, ethical dilemmas of researchers, and the problems of "burn-out" among helping professionals. The introduction shows students how to read such material critically, while thoroughly class-tested projects make this material concrete by involving students in a variety of independent activities, including data collection, theoretical critiques, and group exercises.

Random House/1979/320 pages paperbound/Order Codes: 31281, 32168 (I.M.)

PERSONALITY: An Introduction

John Lamberth, Herbert Rappaport, and **Margaret Rappaport,** all of Temple University

Each chapter of this comprehensive text was written by the author with the greatest expertise and interest in the area, making the whole text completely authoritative. Every major aspect of personality study — assessment techniques, development, research, and applications — is fully covered and viewed in the light of all theoretical approaches, giving you the most comprehensive, balanced, up-to-date presentation possible.

Alfred A. Knopf/1978/544 pages hardbound/Order Codes: 31190, 32167 (I.M.)

For more information, write to:

Random House/Alfred A. Knopf

College Review Desk J-59, 400 Hahn Road, Westminster, Md. 21157

JOURNAL OF

Personality AND Social Psychology

VOLUME 37 NUMBER 4 · APRIL 1979

Cognitive Processes in Implicit Personality Trait Inferences 471
Ebbe B. Ebbesen and Robert B. Allen

The time taken to decide whether traits co-occur indicates the nature of the semantic memory processes underlying implicit personality inferences.

Effects of Exogenous Changes in Heart Rate on Facilitation of Thought and Resistance to Persuasion 489
John T. Cacioppo

Increasing the heart rate improved cognitive functioning, even though subjects were unaware of the acceleration.

Encoding of Personal Information: Self–Other Differences 499
N. A. Kulper and T. B. Rogers

Self-referent information is processed more efficiently than other-referent information.

Making Trust Easier and Harder Through Two Forms of Sequential Interaction 515
Philip Brickman, Lawrence J. Becker, and Sidney Castle

Trust remains a dilemma in continuing interaction even when people are not required to act simultaneously.

Probability of Acceptance in Dating Choice 522
James Shanteau and Geraldine F. Nagy

Consequences of Responsiveness in Dyadic Interaction: Effects of Probability of Response and Proportion of Content-Related Responses on Interpersonal Attraction 534
Deborah Davis and William T. Perkowitz

Behavioral Change in a Constant Environment: Shift to More Difficult Tasks With Constant Probability of Success 551
Julius Kuhl and Virginia Blankenship

Effects of Paced Respiration and Expectations on Physiological and Psychological Responses to Threat 564
Kevin D. McCaul, Sheldon Solomon, and David S. Holmes

Interpersonal Attraction in Aversive Environments: A Problem for the Classical Conditioning Paradigm? 572
Douglas T. Kenrick and Gregory A. Johnson

Blood Donation and the Foot-in-the-Door Technique: A Limiting Case 580
Robert D. Foss and Carolyn B. Dempsey

(Continued on inside back cover)

Published monthly by the
AMERICAN PSYCHOLOGICAL ASSOCIATION
ISSN 0022-3514

Editor

Anthony G. Greenwald, *Ohio State University*

Topic Editors

Irwin Altman, *University of Utah*
Robert A. Baron, *Purdue University*
Daryl Bem, *Stanford University*
Eugene Burnstein, *University of Michigan*
Robert B. Cialdini, *Arizona State University*
Reid Hastie, *Harvard University*
Douglas N. Jackson, *University of Western Ontario*
Leslie Z. McArthur, *Brandeis University*

Consulting Editors

Icek Ajzen, *University of Massachusetts—Amherst*
Robert S. Baron, *University of Iowa*
Gordon Bear, *Ramapo College of New Jersey*
Ellen Berscheid, *University of Minnesota*
Michael Birnbaum, *University of Illinois at Urbana-Champaign*
Philip Brickman, *Institute for Social Research*
J. Merrill Carlsmith, *Stanford University*
John S. Carroll, *Loyola University of Chicago*
Jerry L. Cohen, *University of Illinois at Urbana-Champaign*
Sheldon Cohen, *University of Oregon*
Rick Crandall, *Texas Christian University*
William D. Crano, *Michigan State University*
Robyn M. Dawes, *University of Oregon*
Kay K. Deaux, *Purdue University*
Marshall Dermer, *University of Wisconsin—Milwaukee*
Edward Diener, *University of Illinois at Urbana-Champaign*
Anthony Doob, *University of Toronto, Toronto, Canada*
Susan V. Eisen, *McLean Hospital, Belmont, Massachusetts*
Baruch Fischhoff, *Decision Research, Eugene, Oregon*
Irene H. Frieze, *University of Pittsburgh*
David Funder, *Stanford University*
David L. Hamilton, *University of California, Santa Barbara*
Ranald D. Hansen, *Oakland University*
Richard J. Harris, *University of New Mexico*
Edward Helmes, *University of Western Ontario, London, Canada*
Donald S. Hiroto, *VA Hospital, Brentwood, Los Angeles, CA*
Edward E. Jones, *Princeton University*
Martin F. Kaplan, *Northern Illinois University*
Douglas T. Kenrick, *Montana State University*
Norbert L. Kerr, *University of California, San Diego*
Samuel Komorita, *University of Illinois at Urbana-Champaign*
Vladimir J. Konečni, *University of California, San Diego*
Robert M. Krauss, *Columbia University*
Ellen J. Langer, *Harvard University*
Mark R. Lepper, *Stanford University*
Melvin Manis, *University of Michigan (Ann Arbor)*
Hazel Markus, *University of Michigan (Ann Arbor)*
Margaret Matlin, *SUNY College at Geneseo*
Dale T. Miller, *University of British Columbia, Vancouver*
Thomas M. Ostrom, *Ohio State University*
Paul B. Paulus, *University of Texas at Arlington*
Richard Petty, *University of Missouri—Columbia*
Harry T. Reis, *University of Rochester*
Judith Rodin, *Yale University*
Seymour Rosenberg, *Livingston College, Rutgers*
Ralph L. Rosnow, *Temple University*
Lee Ross, *Stanford University*
Michael Ross, *University of Waterloo, Waterloo, Canada*
Diane Ruble, *Princeton University*
Michael J. Saks, *Nat'l Center for State Courts, Williamsburg, VA*
Michael F. Scheier, *Carnegie-Mellon University*
Barry R. Schlenker, *University of Florida*
Philip Shaver, *New York University*
Steven J. Sherman, *Indiana University*
Jerome E. Singer, *University of the Health Services, Bethesda, Maryland*
Eliot R. Smith, *University of California, Riverside*
Mark Snyder, *University of Minnesota*
Melvin L. Snyder, *University of Texas at Austin*
Lawrence Stricker, *Educational Testing Service*
Shelley E. Taylor, *Harvard University*
Abraham Tesser, *University of Georgia*
Bernard Weiner, *University of California, Los Angeles*
Gary L. Wells, *University of Alberta, Edmonton, Canada*
Stephen Worchel, *University of Virginia*
Robert S. Wyer, *University of Illinois at Urbana-Champaign*
Mark Zanna, *University of Waterloo, Waterloo, Canada*
Miron Zuckerman, *University of Rochester*

Assistant to the Editor
Holly Wong

APA Journal Staff

Anita DeVivo, *Executive Editor*
Ann I. Mahoney, *Manager, Journal Production*
Barbara R. Richman, *Production Supervisor*
Marjorie M. Henry, *Production Editor*
Robert J. Hayward, *Advertising Representative*
Juanita Brodie, *Subscription Manager*

The Journal of Personality and Social Psychology publishes original reports of research, methodology, theory, criticism, and literature review in the fields of personality and social psychology of normal processes and populations. Submitted manuscripts should delineate an important question and provide a clear answer to it. In the case of research reports, this will often require demonstration that the reported results are generalizable beyond a single research setting. Consult the editorial in the January 1976 issue (Vol. 33, pp. 1–7) for a detailed statement of policies and instructions to authors.

Manuscripts: Submit manuscripts in quadruplicate, prepared according to the *Publication Manual of the American Psychological Association* (2nd ed.). Copies should be highly legible and will not be returned to authors unless they are original typewritten copies. Mail manuscripts according to instructions on pp. 635.

Required supplementary information: Manuscripts cannot be reviewed for publication unless the author provides supplementary information in the following categories: (a) a cover letter that indicates the relation (or the lack of relation) of the submitted manuscript to other papers by the same author(s) that are currently submitted, in press, or recently published; (b) one copy of each such related paper; (c) one copy of a statement of the procedures of the research reported in the submitted manuscript in sufficient detail to serve as a guide for conduct of an exact replication; and (d) one copy of the summary tables (e.g., correlation tables, analysis of variance summaries) for the *major* analyses reported. If it appears that the supplementary materials in categories (c) or (d) might be bulky, the author should exercise judgment and restaint in selecting what to send. It is not necessary to send supplementary material that repeats verbatim manuscript content.

After a manuscript is accepted, authors will be expected to provide assurance that their data are available in unanalyzed form for reanalyses by other investigators and that their research procedures were consistent with the American Psychological Association's published principles of research ethics, except insofar as may be detailed otherwise in the manuscript.

Copyright and Permission: All rights reserved. Written permission must be obtained from the American Psychological Association for copying or reprinting text of more than 500 words, tables, or figures. Permission is normally granted contingent upon like permission of the author, inclusion of the APA copyright notice on the first page of reproduced material, and payment of a fee of $10 per page, table, or figure. Abstracting is permitted with credit to the source. Permissions and fees are waived for authors who wish to reproduce their own material for personal use; fees only are waived for authors who wish to use their own material commercially. Permission and fees are waived for the photocopying of isolated articles for nonprofit classroom or library reserve use by instructors and educational institutions. Libraries are permitted to photocopy beyond the limits of U.S. copyright law: (1) those post-1977 articles with a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through the Copyright Clearance Center, P.O. Box 765, Schenectady, NY 12301; (2) pre-1978 articles, provided that the per-copy fee stated in the Publishers' Fee List is paid through the Copyright Clearance Center, P.O. Box 765, Schenectady, NY 12301. Address requests for reprint permission to the Permissions Office, APA, 1200 Seventeenth Street, N.W., Washington, D.C. 20036.

Subscriptions: Subscriptions are available on a calendar year basis only (January through December). Nonmember rates for 1979: $60 domestic, $64 foreign, $6 single issue. APA member rate: $20. Write to Subscription Section, APA.

Change of Address: Send change of address notice and a recent mailing label to the attention of the Subscription Section, APA, 30 days prior to the actual change of address. APA will not replace undelivered copies resulting from address changes; journals will be forwarded only if subscribers notify the local post office in writing that they will guarantee second-class forwarding postage.

Reprints: Authors may order reprints of their articles from the printer when they receive proofs.

Back Issues and Back Volumes: For information regarding back issues or back volumes write to Order Department, APA.

Microform Editions: For information regarding microform editions write to any of the following: Johnson Associates, Inc., P.O. Box 1017, Greenwich, Connecticut 06830; University Microfilms, Ann Arbor, Michigan 48106; or Princeton Microfilms, Princeton, New Jersey 08540.

Published monthly in one volume per year by the American Psychological Association, Inc., 1200 Seventeenth Street, N.W., Washington, D.C. 20036. Printed in the U.S.A. Second-class postage paid at Arlington, Va., and at additional mailing offices.

JOURNAL OF
Personality
AND
Social Psychology

Volume 37 April 1979 Number 4

Cognitive Processes in Implicit Personality Trait Inferences

Ebbe B. Ebbesen and Robert B. Allen
University of California, San Diego

The cognitive processes by which people infer whether individuals who possess one trait also possess another were examined in two reaction time experiments. Subjects took less time to affirm and more time to deny that two traits co-occurred, the greater the *semantic* similarity of the traits. As the amount of recalled evidence required to affirm that two traits co-occurred was increased by altering the nature of the co-occurrence statements, true response times increased and false response times decreased. Although it was not possible to determine whether the stored semantic "features" of a trait are locations on meaning dimensions, specific behaviors, known people characterized by the trait, or something else, the results nevertheless strongly suggest that implicit personality inferences result from a two-stage process in which the second and more detailed memory search stage is entered only if the similarity of the semantic features of the traits falls in between two task-established decision criteria.

People are willing to infer that an individual has one trait or characteristic from information about the other traits or characteristics possessed by that individual (Asch, 1946; Schneider, 1973). Inferences about the nature of someone's personality are made with relative confidence even in the absence of direct (or behavioral) indicators of many of the inferred traits (Asch, 1946; D'Andrade, 1974; Fiske, 1978; Mischel, 1968; Shweder, 1975). This tendency to infer some traits from knowledge about other traits has been termed "implicit personality theory" (Bruner & Tagiuri, 1954; Schneider, 1973). Several issues within this general topic area have received considerable research attention. Some studies have attempted to characterize the *structure* of the interrelationships among trait terms (see Rosenberg & Sedlak, 1972, for a review). Factor analysis models (e.g., Mulaik, 1964; Passini & Norman, 1966), multidimensional scaling models (e.g., Rosenberg & Sedlak, 1972), and hierarchical clustering models (Johnson, 1967) have all been explored and generally provide

This research was supported by Grant MH 26069 from the National Institute of Mental Health to the first author. Analyses of portions of the data were supported by Bell Laboratories, Murray Hill, New Jersey. Portions of the results described in this article were reported at the meeting of the American Psychological Association, San Francisco, California, August 1977. The assistance of Marie-Claire Valtz in running the first experiment is greatly appreciated.

Requests for reprints should be sent to Ebbe B. Ebbesen, Department of Psychology, C-009, University of California, San Diego, La Jolla, California 92093.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3704-471$00.75

Diary No. 68
Date 2/11/1979
Librarian

interpretable representations. Other studies have been concerned with the way groups of traits *combine* to affect the amount of another trait (e.g., likeableness) that is inferred or perceived (Anderson, 1974). Finally, some have examined the role of implicit personality theory in other phenomena, such as memory for a list of trait terms or behavior descriptions (e.g., Cantor & Mischel, 1977).

Although these various approaches to the study of implicit personality theory have yielded a wealth of evidence about the type, strength, and structure of the relationships among people's trait conceptions, little is known about the nature of the *inference processes* that allow these relationships to emerge in different tasks. There is, however, a growing body of literature within cognitive psychology that focuses on such issues—but for information sets other than trait terms. For example, in a classic study, Collins and Quillian (1969) used a binary forced-choice (true/false) decision task to study the nature of semantic memory. Subjects were asked to make true/false decisions about the assertion "A(n) *X* is a(n) *Y*," where *X* and *Y* were object and category names such as *robin* and *bird*, or *bus* and *vehicle*. They suggested that the time that subjects took to answer such questions could be used as evidence for models of the structure of semantic information, of the retrieval of that information from memory, and of the inference or decision processes that governed how that information was used in the construction of answers to such questions.

Researchers interested in implicit personality theory have required that subjects answer questions with similar formats to those used by Collins and Quillian. For example, in one method, subjects are asked to indicate the likelihood that a person who has Trait *X* also has Trait *Y* (Bruner, Shapiro, & Tagiuri, 1958; Hays, 1958; Lay & Jackson 1969). We can reformulate such questions so that they conform to the binary forced-choice paradigm used by Collins and Quillian by asking people to verify whether "a(n) *X* person is a(n) *Y* person," where *X* and *Y* are trait terms. Although these questions are not identical to the noun questions typically used in cognitive research, we can still examine the length of time it takes subjects to indicate whether a given class of objects (people, in our case) possess particular attributes (traits, in our case). The decision time data can then be used in the same way Collins and Quillian (1969) and others (e.g., Smith, Shoben, & Rips, 1974) have used them.

Although the major predictions of such decision time experiments depend on the specific nature of the search and decision processes that are proposed, it is usually the case that search processes are constructed to fit a particular theory about the structure of semantic memory. Two different structural representations that have been proposed for object categories seem most reasonable in the case of personality traits. In one view, a trait can be characterized in terms of the set of "people I have known" who possess that trait. The meaning of a trait stems from the set of person examples associated with that trait. People know what the trait term *generous* means because they can retrieve from memory one or more examples who seem to possess the trait of generosity. In other words, a set of remembered examples defines the trait term.

A second view characterizes the meaning of a trait term as a list of more elementary features. Some of these features might be locations on underlying dimensions (see Rosenberg & Sedlak, 1972) such as socially good–bad and/or dominant–submissive. Features might also be behaviors typically linked to that trait. For example, donating money to charity could be a feature of the trait of generosity. Here, the meaning of a trait term is derived from the list of basic features rather than from the sample of known people it activates in memory. These two models of the structure of personality trait terms suggest, on an intuitive basis, somewhat different inference processes for the type of binary forced-choice tasks outlined above. Let us consider these next.

Exemplar Scanning Model

If personality traits are represented in memory not as a list of more basic features

but rather as sets of persons who have the relevant characteristic (see Walker, 1975), then, when asked to verify trait co-occurrence statements (e.g., "A witty person is also generous"), people might retrieve one or more of the known persons who possess the first trait (witty) and examine this sample to determine whether the second trait (generous) is an attribute of the people in the sample. We can further assume, as Walker has done for object categories, that a "true" response will be made when a sufficient number of the person examples in the sample possess the second trait and that a "false" response will be emitted when an insufficient number of the exemplars in the sample possess the second trait. It should be emphasized that there is no need to assume that a subject considers the entire set of known people who possess the first trait. We need merely assume that a subject creates a sample, probably consisting of the most available and/or typical people in the entire set (Kahneman & Tversky, 1972; Walker, 1975). In this model, the length of time a person takes to reach a decision should increase as the number of examples that are examined increases.

Feature Comparison Model

Regardless of the intuitive plausibility of the person-exemplar model, beginning with Smith et al. (1974), several reaction time experiments, using the Collins and Quillian (1969) paradigm, have supported a somewhat different view. Specifically, people might verify assertions about object categories on the basis of the overall similarity between lists of features that the two concepts (X and Y) activate in memory. This model assumes that when the similarity of the two feature lists is above some task-established cut point, C_T, the participant responds "true." However, when the similarity of the two feature lists is below another cut point, C_F, a "false" response is made. When the similarity of the feature lists falls between these two cut points, as might happen for assertions such as "A robin is an animal" or "A dolphin is a fish," a second stage of processing is hypothesized, and therefore the time to reach a decision regarding such assertions should be longer. While Smith et al. (1974) argued that this additional processing consisted of an examination of the similarity of only the *defining* (as opposed to the most *typical*) features of the two concepts, it is sufficient for our purposes to argue that the additional processing requires that more features be examined than in the initial stage (or that the same ones be reexamined) without specifying the exact nature of the second stage. The most straightforward decision time prediction of this model is that the time to make "false" decisions should *increase* with the similarity of the two feature lists, whereas the time to make "true" decisions should *decrease*.

Effects of Quantification in Each Model

Quantification refers to the fact that we can alter an assertion by replacing "a person" with "all people" or "some people." Thus, rather than asking participants to verify the sentence "A witty person is also intelligent," we could ask them to verify the *universal* assertion "All witty people are also intelligent" or the *particular* assertion "Some witty people are also intelligent." The effects of these alterations on decision times can offer insights into the decision and search processes used in generating responses to our statements (see Meyer, 1970; Rips, 1975).

Exemplar scanning. If a participant does retrieve from memory a sample of person exemplars with the X trait and examines them to determine whether they also possess the Y trait, decision times should increase as the number of exemplars that are examined increases. Consider universal assertions. Here, participants are attempting to decide whether all Xs are Ys. It seems reasonable to expect that the participants would respond "true" only if all of the X exemplars in the retrieved sample were found to possess the Y trait. However, once the first X exemplar that did not possess the Y trait was encountered, this should provide sufficient evidence to stop the examination and respond "false." The likelihood of finding an X exemplar without the Y trait in the sample will depend on the co-occurrence of the X and Y traits in the set of known people. If every

known X exemplar possesses the Y trait (complete co-occurrence), long, "true" decision times should be found. If, on the other hand, none of the X exemplars possesses the Y traits (no co-occurrence), then the first retrieved example will always provide sufficient evidence for fast, "false" decision times. When some, but not all, of the known Xs are Ys (partial co-occurrence), then the length of time should depend on the size of the sample retrieved and the proportion of X exemplars with the Y trait that are in the retrieved sample. The greater the number of X exemplars that are examined before a negative instance is encountered, the longer the time to reach a "false" decision.[1]

It should be noted, however, that if the sample size is constant, and greater than one, all of the latter "false" decision times should be shorter than "true" decision times (because fewer examples are examined) and longer (on the average) than the "false" times for the no-co-occurrence case (since only one example needs to be examined in the latter instance). Thus, for universal assertions, participants should take the longest time to verify statements in which there is complete co-occurrence, the shortest time to negate statements in which the X and Y traits do not co-occur, and an intermediate amount of time to negate statements in which the X and Y traits co-occur for some but not all of the exemplars.

Applying similar reasoning to particular assertions predicts that participants should emit fast, "true" responses when there is complete co-occurrence, because the first example should provide sufficient evidence that at least one X exemplar has the Y trait.[2] On the other hand, long, "false" response times should be observed when there is no co-occurrence. The entire sample should be examined before the participants are sure that not even one X exemplar has the Y trait. Moderately long "false" times should result when there is partial co-occurrence, because, on the average, more than one but less than the entire sample of exemplars should be examined before a positive case is found.

Feature comparison. Rips (1975) has outlined how the feature comparison model might be affected by quantifying assertions with *all* and *some*. His argument consisted of several major points. First, he noted that the similarity of the X and Y feature lists will generally be confounded with co-occurrence relationships. The feature lists of X and Y should be most similar when there is complete co-occurrence and least similar when there is none. Second, he argued that quantification of assertions would affect the placement of C_T (but not C_F) on the feature similarity dimension. C_T should be set high (making it harder to say "true") for universally quantified sentences and moderately low (making it easier to say "true") for sentences quantified by *some*. Note that if the position of C_F does not change, then the number of sentences requiring second-stage processing should be smaller for particular than for universal assertions. The distance between C_T and C_F should be smaller, and therefore the probability of any given X–Y similarity estimate falling between them should be reduced. Finally, to make exact predictions, we must know how the similarities of the three types of trait pairs are distributed. Rips (1975) assumed that for object categories, the average similarity of Xs and Ys in which only some of the X exemplars could be classified as Y types (e.g., *mother* and *doctor*) would be closer to the average similarity of X–Y pairs in which all of the X exemplars could be classified as Y types (e.g., *mother*

---

[1] Because we are proposing that the decision is based on an examination of a *sample* rather than on all of the known persons, decision errors can be easily accounted for by this model. For example, a subject could respond "true" to a universal assertion even though only some of the known persons who possess the X trait also possess the Y trait. We need merely assume that the retrieved *sample* consisted of X exemplars, all of whom possessed the Y trait.

[2] To make this prediction we have assumed that subjects interpret the quantifier *some* to mean "at least one." It should be noted, however, that similar predictions to those made in the text can be derived if we assume that the term *some* is interpreted to mean "a few." In this case, we would assume that a "true" response would be emitted only if "a few" (say two or three) exemplars were found to possess the relevant trait.

and *woman*) than to the average similarity of *X–Y* pairs in which none of the *X* exemplars could be classified as *Y* types (e.g., *mother* and *chair*). Extending Rips's reasoning to the co-occurrence of trait terms, Figure 1 shows three distributions of feature list similarities for the three types of co-occurrence relationships whose spacing conforms to Rips's assumption.[3] Also shown is the postulated effect on C_T of changing the quantifier from *all* to *some*.

The decision time predictions of this model are based on the proportion of trait pairs that fall between C_F and C_T and therefore require second-stage processing in any given condition. One prediction is that decision times should be longer for universal than particular assertions, since, as can be seen in Figure 1, a smaller number of trait pairs falls between C_F and C_T in the latter than in the former case. A second prediction is that "false" response times should be positively related and "true" response times should be negatively related to the similarity of trait pairs, even for those trait pairs in the same co-occurrence relationship.[4] This prediction follows from the fact that as the similarity increases, there is a higher likelihood that a "false" response and a lower likelihood that a "true" response has resulted from second-stage processing. Predictions for particular types of trait pairs (e.g., those in a specific co-occurrence relationship) are easily derived by noting the relative areas of each curve that fall between C_F and C_T in Figure 1.

The two models outlined above can be contrasted in terms of the predictions they make regarding the effects on decision time of the co-occurrence relationship between the *X* and *Y* traits and the use of the quantifiers *all* and *some*. Table 1 presents the predictions of the two models in terms of relative decision times. As can be seen, a major difference between the two models is their prediction for the partial co-occurrence trait pairs. The exemplar scanning model predicts that the times for these statements will fall midway between those for the other two types, whereas the feature comparison model predicts that these trait pairs should yield the longest decision times. Additional specific predictions can be found by further study of Table 1.

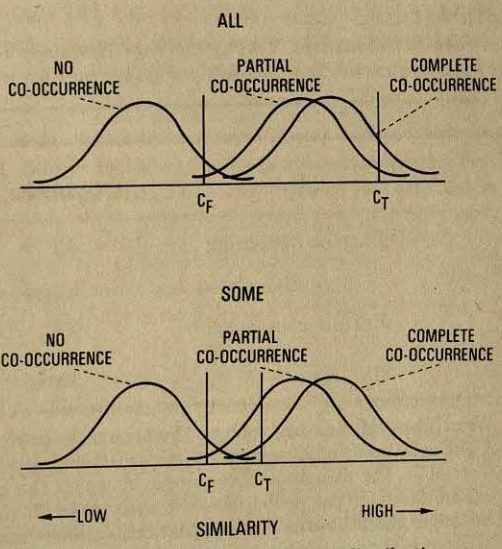


Figure 1. Spacing of similarity distributions assumed by the feature comparison model for the three types of trait pairs: pairs that are not perceived to co-occur, pairs that are perceived to co-occur in some but not all people (partial), and pairs that are perceived to co-occur in all people (complete). (Also shown is the effect on the true criterion, C_T, of changing the quantifier from *all* to *some* in the co-occurrence assertions: C_T moves down, making it easier to say "true.")

Defining Trait Sentences as Complete, Partial, or No Co-Occurrence

Unlike the noun sentences typical in cognitive research, there is no a priori method of

[3] It is important to note that many of the predictions of the feature comparison model depend directly on the assumed spacing of the similarity distributions for different types of trait pairs. Therefore, it is useful to have independent evidence about the validity of these assumptions. Without such evidence, it is possible to account for a much wider range of outcomes merely by postulating different similarity distributions.

[4] To the extent that our semantic similarity and co-occurrence estimates are highly correlated, this result could be explained by the exemplar scanning model as well. We need merely assume that the proportion of positive exemplars in a sample increases monotonically with semantic similarity. The number of exemplars that would have to be scanned before a "true" response could be emitted would then decrease as semantic similarity increased.

Table 1
Order Relationships Among Average Response Times Predicted by Exemplar Scanning and by Feature Comparison Models for Different Quantifiers and Extent of Trait Co-Occurrence

Model	Quantifier	Trait co-occurrence: Complete		Partial		None
Exemplar scanning	All	True	>	False[a]	>	False
		>				<
	Some	True	<	True[a]	<	False
Feature comparison[b]	All	True	<	False	>	False
		>		>		=
	Some	True	<	True	>	False

Note. True/false entries refer to the type of response being made in the condition.
[a] The relationship between these two outcomes depends on the proportion of X exemplars that possess the Y trait in the sample and the order in which the exemplars are examined. Assuming a random order of examination, if the proportion is greater than .5, then the false times to universal (all) assertions should be longer than the true times to particular (some) assertions. If less than .5, then the reverse should be found.
[b] The predictions of this model are based on the similarity distributions depicted in Figure 1.

deciding into which type of relationship a given trait pair, such as *witty* and *intelligent*, falls. While almost everyone would agree that the set of known robin exemplars is a subset of the set of known birds, the relationship between the set of known witty persons and the set of known intelligent persons is less clear. We can, however, determine from subjects' responses to sentences quantified by *all* and *some* what the *majority* of people believe about a given sentence. If the majority of subjects respond "true" to both the *all* and *some* versions of a sentence, then (for the majority) all of the X exemplars may be considered to possess the Y trait. When the majority of people respond "false" to the *all* version and "true" to the *some* version, then only some of the X exemplars possess the Y trait. Finally, when the majority respond "false" to both versions, the X and Y traits will not be considered to co-occur.

Similarity of Trait Feature Lists

The similarity between feature lists for the different traits was assessed by requiring that subjects rate the similarity in meaning of all relevant trait pairs. We assumed that these ratings would provide monotonic indicators of the overall similarity of each pair of feature lists (Rips, 1975; Smith et al., 1974).

Experiment 1

Experiment 1 was designed to examine the contrasting predictions of the two models. Ratings of the similarity of trait pairs and decision times to verify universal, particular, and "a person" assertions about these trait pairs were recorded.

Method

Subjects

Nine male and nine female college students each participated in three 1-hour sessions in partial fulfillment of an introductory psychology course requirement. All of the participants were native speakers of English.

Apparatus and Software

The decision time data were collected using a PDP-12 laboratory computer. Responses were made to stimuli, displayed on the video screen, on two telegraph keys, one labeled "true," the other "false." A user-oriented computer program was used to control and randomize stimulus pairs and to record the response latency and which of the two keys had been pressed. Decision times were accurate within 1 msec.

Procedure

Participants were told that the study dealt with their reactions to various personality characteristics of people and that they would have to participate in

three separate sessions, each lasting from 45 minutes to 1 hour and separated by 3 to 5 days.

Semantic similarity and co-occurrence ratings. All of the participants completed two questionnaires. One was filled out in the first session and one in the third, counterbalanced for order over participants. Both questionnaires presented all possible pairs of 12 different trait adjectives (see Figure 2). The semantic similarity questionnaire asked the participants to indicate "how similar in meaning" the two terms were on a 20-cm unmarked scale anchored by "not similar in meaning" and "identical in meaning." The other, co-occurrence, questionnaire asked the participants to indicate, "How likely is a(n) _____ person to also be _____?" Unmarked 20-cm scales were again used, but in this case they were anchored by "not likely" and "extremely likely." The participants were told to complete the semantic similarity scales on the basis of each term's "dictionary definition" but were not given a particular instruction regarding the co-occurrence ratings. Participants were run individually in an isolated cubicle and took as much time as they needed to complete each questionnaire.

Forced-choice decision task. The procedures for the three decision tasks were virtually identical. One was conducted in the first session after the first questionnaire was completed, and two were conducted in the second session. Order of task was counterbalanced across the participants. Upon arrival at the computer laboratory, participants were seated in front of the PDP-12 and video stimulus display. They were told that pairs of trait terms would appear on the video screen in blank spaces in a cardboard mask that was placed on the video screen. In a particular session, one of the three following sentences was written on the mask: "A(n) _____ person is also a(n) _____ person";[5] "Some _____ people are also _____"; and "All _____ people are also _____." The subjects were told that when a trait pair appeared on the screen, they were to indicate whether the *entire sentence* was true or false by pressing the right (left for half of the subjects) telegraph key if they thought it was true and the left (right) key if false. They were told to read the entire sentence each time that a new pair of trait terms was presented. They were also told to think about each sentence carefully before responding, but as soon as they were sure of the answer to press one of the keys. After demonstrating to the subjects how to position their hands near the telegraph keys, the experimenter began a short program that presented five practice trait pairs (not in the actual experimental list). The experimenter observed the participants while they completed these practice trials, reinstructing them if necessary. After these trials were completed, and following a 2- to 3-minute delay, the main program was initiated. The experimenter left the room, instructing the participant that the word *finished* would appear in both blank spaces on the video display when the task was completed.

The main program presented each word in the trait pair virtually simultaneously. Measurement of the response latency began the moment that the two words appeared on the screen and ended either when one of the two keys was pressed or when a 15-sec maximum had been reached. There was a 4-sec intertrial interval, during which time the display was blank, between the presentations of the stimulus pairs. No "ready" signal was used in this study. Each participant was presented with a given pair of trait terms only once. The two terms in a given pair were presented in only one order, which was held constant over participants.

Results

The data were analyzed in an attempt to examine several major issues. Of initial concern was whether the semantic similarity ratings (uncommon in implicit personality research but common in cognitive research) would yield a multidimensional representation similar to that obtained from the co-occurrence ratings (which are more common in studies of implicit personality). The second analysis studied the effect of the universal and particular quantifiers on decision times for the three types of trait pairs: complete, partial, and no co-occurrence. In the third analysis, the relationship between semantic similarity and "true" and "false" decision times was examined. The fourth analysis examined the ability of the semantic similarity ratings to explain the effects of the co-occurrence relationships on decision times.

Multidimensional Structure of Trait Terms

The structure inherent in the semantic similarity ratings was compared to that obtained with the more common co-occurrence rating method. The average (over participants) semantic similarity and the average co-occurrence rating between each pair of the 12 trait terms used in the study were computed. The two half-matrices (the trait terms were paired in only one order) were then separately analyzed using KYST2 (see Kruskal, Young, & Seery, Note 1), a monotone multidimensional scaling program. The en-

[5] For reasons of clarity, the results for this decision task will not be presented. It should be noted, however, that they were completely consistent with the results from Experiment 2.

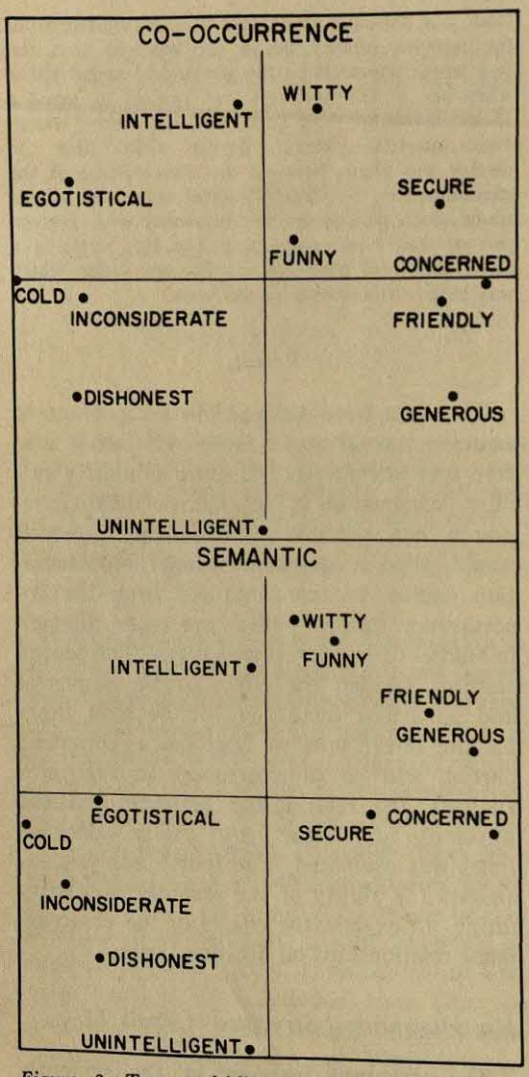


Figure 2. Two multidimensional scaling solutions for the 12 trait terms used in Experiment 1. (The top panel shows the solution for co-occurrence ratings, rotated to its principle components. The bottom panel shows the solution for semantic similarity ratings of the same traits, rotated to minimize the difference between the two solutions. The high correspondence between the two solutions is of most interest.)

tries in each lower half-matrix of the two types of ratings were analyzed as direct distance estimates. Two-dimensional solutions seemed to provide reasonable representations of both data sets. The one-dimensional stress (Formula 1) values were .2,976 for the co-occurrence ratings and .2,302 for the semantic similarity ratings. A two-dimensional fit reduced these values to .1,245 and .0,817, respectively. It seemed inappropriate to fit a three-dimensional solution to the data, given the small stimulus set.

The top panel in Figure 2 contains the two-dimensional solution for the co-occurrence ratings, rotated to its principal components. The bottom panel contains the results for the semantic similarities rotated to minimize the difference between the two solutions. The two solutions yielded remarkably similar results. As in Rosenberg and Sedlak's (1972) work, the two most important dimensions in this trait set seem to be socially good–bad and intellectually good–bad. In short, not surprisingly, the participants in our study seemed to have an implicit personality theory similar to that found for other subject samples in other laboratories. More important, however, the obvious congruence of the semantic similarity and co-occurrence structures implies that processes involved in implicit personality may be related to more basic semantic memory processes, as we have argued.

Complete, Partial, and No Co-occurrence

Of major concern in this experiment were the effects on decision time of universal and particular assertions involving trait pairs that were defined as being in a complete, partial, or no co-occurrence relationship. In order to examine the predictions outlined in Table 1, the 66 sentences quantified by *all* were divided into two groups according to the majority response (true vs. false) given to each. The same was done for the corresponding 66 sentences quantified by *some*. The mean decision time (over participants) for those participants whose response agreed with the majority was then computed for *each* universal assertion and for *each* particular assertion. The 66 trait pairs and their associated decision times were divided into three groups: those for which the majority response was "true" to both the universal and particular versions (complete co-occurrence), those for which it was "false" and "true," respectively (partial co-occurrence), and those for which it was "false" for both versions (no co-occurrence). There were no trait pairs in which the

Table 2
Mean Decision Time (in sec) to Respond to the Complete, Partial, and No Co-Occurrence Trait Pairs When Quantified by All *and* Some

	Trait co-occurrence					
	Complete (n = 16)		Partial (n = 33)		No (n = 17)	
Quantifier	Time	Response	Time	Response	Time	Response
All	3.719	True	3.540	False	3.134	False
Some	3.314	True	3.811	True	3.730	False

Note. n = the number of trait pairs in each co-occurrence category.

majority response was "true" for the universal, but "false" for the particular, version.

Table 2 presents the mean (over trait pairs) decision times for the *all* and *some* quantifiers applied to the three types of trait pairs. A 2 (quantifier) × 3 (type of trait pair) analysis of variance was performed.[6] Contrary to the feature comparison model, the main effect of the type of trait pair was not significant, $F(2, 63) = 1.14$. Also contrary to the feature comparison model, the time to verify sentences quantified by *some* was significantly longer (rather than shorter) than the time to verify those quantified by *all*, $F(1, 126) = 5.57$, $p < .05$. The interaction between quantifier and trait pair type was highly significant, however, $F(2, 126) = 18.33$, $p < .001$. For universal assertions, subjects negated the no co-occurrence pairs much faster than they affirmed the complete co-occurrence pairs. In contrast, for the particular assertions, subjects negated the no and partial co-occurrence trait pairs much more slowly than they affirmed the complete co-occurrence pairs. This pattern of mean reaction times seems to conform more closely to that predicted by the exemplar scanning model than to that predicted by the feature comparison model (compare Tables 1 and 2).

Decision Times and Semantic Similarity

Recall that the feature comparison model argues that "true" response times should decrease and "false" response times should increase as the semantic similarity of the X and Y trait terms increases. To examine these predictions, the decision times were divided according to the participant's "true" and "false" responses, and by quantifier. Each set was then correlated with its associated semantic similarity ratings, separately for each participant. This analysis yielded one "true" correlation and one "false" correlation per quantifier for each participant. Fisher's z transformations of these correlations served as the raw data for the analyses below.

Table 3 presents the mean "true" and "false" correlations for the two quantifiers. Only the mean "true" correlations were significantly different from zero for the sentences quantified with *some*, $t(17) = 4.62$, $p < .01$, whereas only the mean "false" correlations were significant for the sentences quantified with *all*, $t(17) = 2.80$, $p < .05$. The correlations between rated similarity and decision time were examined for each subject. These were significant ($p < .05$) and in the predicted direction for two "true" and six "false" cases when *all* was used, and for eight "true" and three "false" cases when *some* was used. Thus, the average t test results seem to

[6] Actually, two different analyses of variance were computed. In one, the fact that there were more intersection pairs than the other types was treated as representative of the entire population of trait terms, and a least squares solution was therefore employed. In the second analysis, the unequal ns were assumed to be artifactual, and an unweighted-means analysis was used. Since the conclusions were identical, only the results of the second analysis are presented in the text. In addition, it should be noted that the error terms in both of these analyses were based on variability over sentences rather than subjects. The logic of using sentences to generate the error term has been outlined by Clark (1973).

Table 3
Mean of Fisher z Transformed Correlations Between Decision Times and Semantic Similarity Ratings for "True" and "False" Responses to Sentences Quantified With All *and* Some

Response	Quantifier	
	All	Some
True	−.06	−.26**
False	.15*	.08

* $p < .05$ (two-tailed). ** $p < .01$.

represent the proportion of subjects for whom the semantic similarity – decision time correlation was as predicted.

A 2 (true/false) × 2 (quantifier) mixed analysis of variance on these transformed correlations indicated that the "true" response correlations were significantly different from the "false" response correlations, $F(1, 17) = 9.82$, $p < .01$. The effect for type of statement was not significant, $F(1, 17) = 1.48$, $p > .05$, and the interaction between true/false and quantifier was also not significant, $F(1, 17) = 2.85$, $p > .10$. Separate t tests on the difference between the mean "true" and "false" correlations for each quantifier supported this interaction result; the difference was significant in both cases ($p < .05$). The finding that "false" response times were more positively related to semantic similarity than "true" response times across both quantifiers offers relatively strong support for the feature comparison model.

It appears that the results of the co-occurrence analysis (Table 2) supported the exemplar scanning model, whereas the results from the semantic similarities (Table 3) supported the feature comparison model. One plausible explanation for this outcome is that one or more of our assumptions about the nature of the models may require modification. In particular, recall that the predictions of the feature comparison model described in Table 1 were based on an extension of a suggestion made by Rips (1975), namely, that the distributions of semantic similarities for complete, partial, and no co-occurrence trait pairs would be spaced as depicted in Figure 1. To examine the validity of this assumption, we computed the mean similarity rating for each trait pair (only for those participants whose "true"/"false" response agreed with the majority). The average (over sentences) similarities for the three types of sentences were then derived, and they are presented in Table 4. An unweighted-means 2 (quantifier) × 3 (type of trait pair) analysis of variance of these data indicated that there was a highly significant difference among the three types of trait pairs, $F(2, 63) = 350.66$, $p < .001$. The trait pairs defined as being in a complete co-occurrence relationship were indeed seen as much more similar than the no co-occurrence pairs. Partial co-occurrence trait pairs were midway between these two, but, contrary to expectation, they were closer to the no co-occurrence than to the complete co-occurrence trait pairs. No other effects were found to be significant.

As we will discuss below, these similarity results suggest an explanation for the pattern in the previously presented mean decision times that is based on a revised feature comparison model. With this explanation in mind, we examined the ability of the semantic similarity estimates to predict variations in response times *within* each of the three types of trait pairs. The feature comparison model makes the strong prediction that "true" response times should be positively related and "false" times negatively related to similarity, even within these trait pair types. Table 5 presents the correlation (over sentences) between the mean similarities for each trait pair type. As can be seen, the correlations were significantly different from zero in two of the six cases: the true times to the complete and partial co-occurrence pairs that were quanti-

Table 4
Mean Semantic Similarity of Complete, Partial, and No Co-Occurrence Trait Pairs Quantified by All *and* Some

Quantifier	Trait co-occurrence		
	Complete	Partial	No
All	14.19	6.36	3.48
Some	13.52	8.77	2.62

Note. Similarity was rated on a 20-point scale. Higher numbers indicate greater similarity.

fied by *some*. The sign of each of the correlations was in the direction predicted by the feature comparison model, however.

Discussion

Taken together, the results from this experiment offered some support for both models proposed in the introduction.

Mean Decision Times

With one exception, the pattern of mean decision times for the three types of trait pairs seemed to support the exemplar scanning model fairly well. It took less time for subjects to affirm that some X were Y when all of the X exemplars possessed the Y trait than when only some did. Similarly, it took less time to say that all Xs were not Ys when none of the X exemplars possessed the Y trait than when some of the X exemplars did. On the other hand, the time to verify that some Xs were Ys when only some X exemplars possessed the Y trait was not midway between the complete and no co-occurrence trait pairs, as predicted (Tables 1 and 2); instead, it was the longest time.[7]

Decision Time and Semantic Similarity

Although the mean decision times seemed to support the exemplar scanning model, some aspects of the correlations between semantic similarity and decision time supported predictions of the feature comparison model. As predicted, semantic similarity ratings correlated more positively with "true" decision times than with "false" decision times (Table 3). However, the fact that the correlations for "false" responses to particular assertions and "true" responses to universal assertions were not significantly different from zero is somewhat inconsistent with the feature comparison model. The correlations within each type of trait pair also provided only weak support for this model.

The Feature Comparison Model Revisited

It is possible to offer a consistent explanation for the entire pattern of results by revising our assumptions about the effects of quantifiers on the cut points, C_F and C_T, in the feature comparison model and by considering some of the methodological inadequacies of this experiment. Take the latter point first. In this study, each participant produced only one decision time and one semantic similarity rating per trait pair. If the error variability in the decision times were reasonably large, it could account for the failure to find consistently significant correlations for individuals between semantic similarity ratings and decision times (Table 3). Furthermore, the fact that the standard deviation of the mean semantic similarity ratings (over trait pairs within a given type) was very small (min. $\hat{\sigma} = .826$, max. $\hat{\sigma} = 1.974$, on a 20-point scale) could explain the failure to find significant correlations within each type of trait pair (Table 5).

Table 5
Correlation Between Mean Decision Times and Mean Semantic Similarity Ratings for Each Type of Trait Pair and for Each Quantifier

Quantifier	Trait co-occurrence: Complete ($n = 16$)	Partial ($n = 33$)	No ($n = 17$)
All	−.24	.04	.01
Some	−.67*	−.41*	.06

Note. n = the number of trait pairs used to compute each correlation.
* $p < .05$.

Our original predictions for the feature comparison model were based on Rips's (1975) assumption that the semantic similarity distribution for the partial co-occurrence trait pairs would be more similar to

[7] It should be noted that the average semantic similarity results can be used to explain this outcome. If a large majority of the X exemplars did not possess the Y trait in the partial co-occurrence trait pairs, then the participants would be expected to take a long time to determine that at least one exemplar possessed the Y trait. On the average, most of the exemplars in the sample would have to be examined before an answer was given. Thus, the decision time should be almost as long as the time to assert that not even one exemplar possessed the relevant trait.

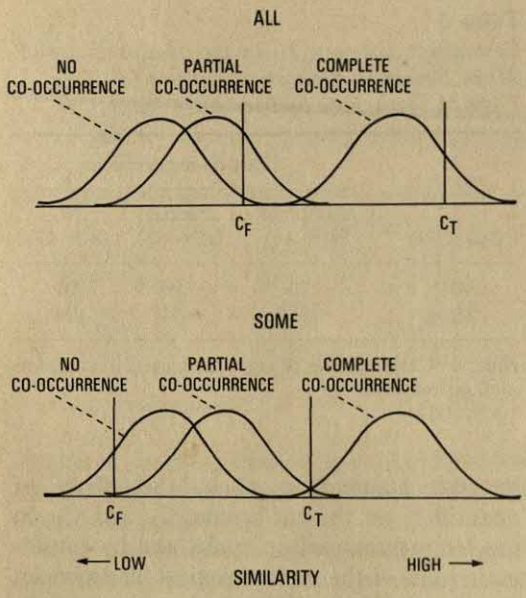


Figure 3. Spacing of similarity distributions obtained for the three types of trait pairs (no, partial, and complete co-occurrence) in Experiment 1. (Also shown is the effect on the "true," C_T, *and* the "false," C_F, criteria of changing the quantifier from *all* to *some*, as assumed by the revised feature comparison model.)

the complete than to the no co-occurrence similarity distributions (see Figure 1) and that changing the quantifiers would only affect the position of C_T. The mean semantic similarity results (Table 4) suggest that for our trait terms, the first of these assumptions was not satisfied. The average similarity of the *X–Y* terms in the partial co-occurrence pairs was closer to that for the no co-occurrence than the complete co-occurrence pairs. With this finding in mind, it is possible to explain the *exact* pattern of means obtained for the two types of sentences (Table 2) by assuming that the change in quantifier affected the placement of *both* C_F and C_T. That is, changing from universal to particular assertions may lower both C_F and C_T rather than just C_T. The effect of such a change in assumptions can be seen in Figure 3. Three distributions of semantic similarities, one for each type of trait pair, are presented so that the differences in their means correspond, roughly, to the differences in semantic similarity ratings reported in Table 5.

The predictions for this new version of the feature comparison model can be derived by noting the area of each distribution that falls between the cut points. The area is a direct index of the proportion of trait pairs (within the trait-pair type) that requires extra processing time. Note that in the *all* case, most of the complete co-occurrence pairs are between the two cut points, thereby producing long, "true" decision times. A moderate number of the partial co-occurrence pairs falls between the cut points, resulting in moderate times for "false" responses to these pairs. Since most of the no co-occurrence pairs are below C_F, very fast "false" responses would be expected. On the other hand, in the *some* case, most of the complete co-occurrence pairs are above the new C_T cut point, and fast "true" responses would be expected. Almost all of the partial co-occurrence pairs are between C_F and C_T, resulting in very long (rather than moderate, as predicted earlier), "true" responses. Finally, the large proportion of no co-occurrence pairs falling within C_F and C_T should produce longer "false" response times for these pairs. As can be seen, this describes the exact pattern of decision time means that was obtained (Table 2).[8]

Given the relatively better fit of this revised feature comparison model, it is important to ask why Rips (1975) did not find evidence for a shift in C_F and we did. One possibility is that the mean of the similarity distribution for the no co-occurrence pairs used by Rips and others (e.g., Meyer, 1970) may have been so low and different from the other distributions that a change in C_F went undetected. The psychological distance be-

[8] If the similarity distributions and cut points are as depicted in Figure 3, we would expect that the mean "true" response times, over all types of trait pairs, should be longer when *all* as opposed to *some* is used as the quantifier. However, the reverse should obtain for the mean "false" response times. This prediction follows from the fact that more fast (single-stage) "true" responses (those above C_T) would be expected with *some* than with *all* as the quantifier. On the other hand, many more long (second-stage) "false" responses would be expected for *some* than for *all* sentences. This pattern of predicted effects was obtained. For universal assertions, the mean "true" response time was 3.85 sec, and the mean "false" time was 3.65 sec, but for particular assertions, the means were 3.61 and 4.16 sec, respectively. The interaction was significant, $F(1, 17) = 6.64$, $p < .05$.

tween an "intelligent" and a "dishonest" person may be much smaller than that between a "woman" and a "chair." One implication of this point is that future research should include evidence about the average similarities of different types of pairs. Unless such evidence is presented, almost any result will be interpretable within the context of the feature comparison model.

The revision in the feature comparison model greatly increases the similarity between what initially appeared to be very different views of implicit personality trait inferences. One of the two crucial aspects that initially differentiated the two models does so no longer, namely, the fact that the criterion for a "false" response varied with the quantifier in the exemplar scanning model and not in the feature comparison model. That is, in the former model we assumed that a change from a universal to a particular quantifier altered the evidence the participant needed for a "false" as well as a "true" response. In the universal case, one negative example was sufficient for a "false" response, but this was not so in the particular case. By allowing C_F to vary, as well as C_T, the same effect is now achieved in the feature comparison model: moderate feature list similarities are sufficient for a "false" response when *all* is used but not when *some* is used.

The commonalities between the two models can be emphasized by noting that both are sensitive to the psychological distance between trait pairs (feature list overlap vs. degree of co-occurrence of attributes in exemplars), and both assume that a different evidentiary criterion is used for "true" than for "false" responses. In the feature comparison model, this last point is made explicit by the dual cut points; however, the exemplar scanning model also postulated a dual criterion (e.g., for particular assertions, the first positive example was sufficient for a "true" response, but a number of negative examples were required for a "false" response). Thus, the two models, while making quite different assumptions about the way in which traits are represented in memory, now predict virtually identical mean reaction time results as a function of alterations in decision criteria because of common processing assumptions.

Experiment 2

It seemed necessary to postulate the existence of a dual-criterion decision process (whether based on features or exemplars) in order to explain the results from Experiment 1. This conclusion greatly limits the class of models that can be used to understand implicit personality inferences of the type being examined here. On the other hand, the evidence for this strong conclusion was not as consistent as we would have liked. For this reason, a second experiment was designed in which more reliable estimates of reaction times and of psychological distances were obtained. The participants were presented with each trait pair six times, three times in one order and three times in the other, both when making binary forced-choice decisions and when rating the similarity of the trait pairs.

Another purpose of this second experiment was to provide evidence relevant to one of the few remaining differential predictions of the two models. Specifically, the exemplar scanning model predicts that the relationship between psychological similarity and "true" and "false" decision times should, in general, vary with the evidentiary criteria used to reach a decision. While "false" times should increase and "true" times should decrease as the similarity increases, regardless of the decision criteria, the absolute value of the *slopes* of these functions should change as the criteria change. When a smaller number of exemplars is required for a false than for a true response (e.g., as in the *all* assertions), the slope for the "false" times should be steeper than for the "true" times. The reverse is expected when a smaller number of exemplars is required for a "true" than for a "false" response. This prediction follows from the fact that the length of a self-terminating search of a sample can extend over a wider range when termination depends on the discovery of a few instances in the sample than when it depends on the discovery of many instances in a sample.[9]

[9] Although this prediction can be tested against the data from Experiment 1, the failure to find significant correlations between similarity and decision time in some conditions prevented a reasonable test from being made.

It should be noted that the revised feature comparison model is quite capable of explaining the pattern described above (were it to be found), provided the distributions of semantic similarities have appropriate forms. Nevertheless, a failure to find the predicted slope changes can be taken as direct evidence against the current version of the exemplar scanning model. Thus, if the predicted changes in slope are found, neither model will receive selective support; but if the predicted changes are not found, the exemplar scanning model, as proposed, can be rejected. Naturally, this prediction need only be examined after other aspects of the results, which are predicted by both models, are found to conform to the common predictions of both models.

As can be seen, an essential aspect of any experiment that attempts to provide evidence relevant to either of the models is that the evidentiary criteria for "true" and for "false" responses be varied in some manner. Rather than repeat the same quantification design used in Experiment 1, we chose to broaden the empirical domain of both models by varying these criteria in a less common manner than is typical of semantic memory research with object categories. Specifically, one half of the participants were asked to verify co-occurrence assertions (in which "a person" was used as a general quantifier), and the other half were asked to verify the similarity of the *dictionary definition* of the trait terms. We assumed that the participants would use a more stringent "true" criterion and a less stringent "false" criterion in the latter than in the former condition. In addition, half of the participants in each of the conditions mentioned above rated the likelihood of the co-occurrence of trait pairs, and half rated their similarity in meaning. This resulted in a 2 (co-occurrence vs. semantic decisions) × 2 (co-occurrence vs. semantic ratings) factorial design.

Method

Subjects

Forty-eight male and female undergraduate college students, who were all native speakers of English, served as participants in exchange for introductory psychology course credit. Six males and six females were randomly assigned to each condition.

Apparatus and Software

The decision times and ratings were collected using a PDP-8A laboratory computer equipped with a video teletype. A user-oriented, experimental-control system, FOEP (Allen, 1978), presented the questions, randomized the trait terms, and recorded the decision times and ratings for later analysis. For both the forced choices and the ratings, the entire question was written onto the video screen at high speed (960 characters/sec).

Procedure

At the beginning of the session, the participants were told that the study dealt with their judgments of various personality characteristics of people. The participants, who were run individually, were then randomly assigned to one of four experimental conditions in which they were asked to make either semantic similarity or co-occurrence ratings and either semantic similarity or co-occurrence binary forced-choice decisions. To provide more stable estimates of the data points, there were three replications for both the ratings and forced-choice decisions. The order of question presentation was randomized for each replication. The order of the ratings and forced-choice judgment tasks was counterbalanced across subjects, and all of the replications for the two tasks were blocked together. The data collection was completed in a 1-hour session.

Semantic similarity and co-occurrence ratings. The question "How similar is the meaning of the trait term _____ to the meaning of the trait term _____?" was used to obtain semantic similarity estimates for all of the trait pairs. The question "How likely is a person who is _____ also to be _____?" was used to obtain co-occurrence estimates. The blanks were completed with all possible combinations of pairs of 10 trait terms. These terms were similar to those used in Experiment 1. For the semantic similarity ratings, the scale was anchored by "not very similar" and "very similar." For the co-occurrence ratings, the scale was anchored by "not very likely" and "very likely." The participants responded by typing a number between 1 and 20, using the digits on the teletype keyboard. For half of the participants, the positive end of the scale was assigned a value of 20; for the other half of the participants, the positive end of the scale was assigned a value of 1.

Forced-choice decision task. The question "Is the meaning of the trait term _____ similar to the meaning of the trait term _____?" was used for the forced-choice semantic decisions, and the question "Is a _____ person also likely to be _____?" was used for the co-occurrence version.[10] The par-

[10] It should be noted that this is a differently quantified statement than those presented in Experiment 1. It seems reasonable that the quantifier *a* would be processed much like the quantifier *a few*. Therefore, relatively little evidence should be required for a "true" response.

ticipants were asked to respond to the questions by pressing one of two teletype keys with their right and left index fingers. The pairing of response key and response direction was balanced across participants. Unlike Experiment 1, the participants in this study were instructed to respond as rapidly but as accurately as possible. It was hoped that this instruction would reduce the variance in decision times. We therefore expected that the average times would be considerably shorter in this study than in the first one.

Each decision trial started with the presentation of the word *ready* on the screen for ½ sec. One-half second after the termination of the ready signal, a question was written on the screen, and the decision time clock was started. When the subject responded, or if no response had been made after 15 sec, the question was erased, and the next trial was started 1 sec later. There was a pause of approximately 1 minute between replications.

Results

Mean Decision Times

As in Experiment 1, the data were analyzed in several different ways. Of initial interest were the effects of question type on the mean "true" and "false" response times. Table 6 presents the average of each subject's mean "true" and mean "false" response times in each condition of the 2 × 2 design. As can be seen, in a manner consistent with both models, people negated semantic assertions faster than they affirmed them; but the reverse was true for the co-occurrence assertions. A 2 (true/false) × 2 (semantic/co-occurrence rating) × 2 (semantic/co-occurrence decision) analysis of variance of the mean decision times indicated that this interaction was highly significant, $F(1, 44) = 29.91$, $p < .001$. There was also a significant tendency for "false" responses to be emitted faster than "true" ones, $F(1, 44) = 4.96$, $p < .05$. No other results were found to be significant. Both models explain the findings mentioned above by assuming that more confirming evidence is needed for a "true" response and less for a "false" response when a semantic rather than a co-occurrence question is asked. Apparently, our attempt to manipulate the true and the false evidentiary criteria was successful.

Table 6
Mean Decision Times (in sec) for Semantic and Co-Occurrence Questions

	Type of response			
	Semantic		Co-occurrence	
Type of rating	True	False	True	False
Semantic	2.085	1.766	2.171	2.254
Co-occurrence	2.201	1.676	1.839	2.103

Table 7
Means of Fisher z Transformed Correlations Between Similarity Ratings and "True" and "False" Decision Times (in log sec)

	Type of response			
	Semantic		Co-occurrence	
Type of rating	True	False	True	False
Semantic	−.26	.35	−.21	.29
Co-occurrence	−.33	.18	−.20	.22

Decision Times and Similarity

One of the most important predictions of both models, which yielded ambiguous results in Experiment 1, concerns the relationship between psychological similarity and "true" versus "false" decision times. In the present experiment, both models predict a negative relationship for "true" response times and a positive one for "false" response times. Table 7 presents the mean correlation for each condition in the 2 × 2 design. Correlations were computed for each subject by taking the mean decision time for the three replications for each trait pair and treating the different orders of trait terms in a pair as unique observations. The "true" and "false" means were then separately correlated with the associated mean similarity ratings for corresponding orders. Not only was the pattern of correlations exactly as predicted, but in addition, with the more sensitive procedure used in Experiment 2, all of the average correlations were significantly different from zero, minimum $t(11) = -2.51$, $p < .05$.

A 2 (true/false) × 2 (semantic/co-occurrence rating) × 2 (semantic/co-occurrence decision) analysis of variance of the results indicated that the true/false difference was highly significant, $F(1, 44) = 252.11$, $p < .0001$. No other effects were found to be sig-

nificant. In short, as predicted by both models, "false" response times increased and "true" response times decreased as the similarity of the pairs increased, whether the similarities of the pairs were measured by semantic overlap or trait co-occurrence, and whether the decision times were produced from verification of a semantic or a co-occurrence assertion.

The previous analyses had established the necessary conditions for an examination of the effects of criteria alterations on the slopes of the relationship between reaction times and similarity. These slopes were computed separately for each subject in the same fashion as were the correlations. A $2 \times 2 \times 2$ analysis of variance of these data indicated that the interaction between true/false and type of question predicted by the exemplar scanning model was not significant ($F < 1$). No other effects were significant. Apparently, the mean absolute values of the slopes were the same in all conditions.

General Discussion

The results from both experiments can be easily summarized. First, when an assertion about a pair of traits was altered (either by changing the quantifier or by changing the nature of the assertion), such that greater psychological similarity was required for a positive response, the time to reach a positive decision increased, and the time to reach a negative decision decreased (Tables 2 and 6). Second, as the psychological similarity between trait terms (whether assessed by definitional or co-occurrence ratings) increased, the latency of positive decisions about the two traits decreased, and the latency of negative decisions increased (Tables 3 and 7). Third, the strength of the relationships between psychological similarity and "true" and "false" latencies did not vary as the evidentiary criteria varied. Taken together, these results seem more consistent with the revised feature comparison model than with the exemplar scanning model.

The former model explains the findings summarized above by postulating a two-stage decision process in which some decisions are made after a first stage, whereas others are made after a second stage. The more time-consuming second-stage processing is required when the similarity of the two feature lists falls in between a "true" and a "false" cut point. The placement of *both* cut points depends on the exact nature of the assertion. The relationship between similarity and "true" and "false" response times follows from the fact that an increase in similarity increases the probability that "false" responses will result from second-stage processing and decreases that probability for "true" responses. The effects of quantifier and type of assertion are explained in terms of shifting cut points.

Given the intuitive plausibility of an exemplar representation of traits, it is important to ask whether the exemplar scanning model can be revised in a manner that retains this representation of traits but which does not predict the slope interaction. One possibility is to alter our assumptions about decision processes. Assume that subjects take more than one exemplar from memory and that their "true"/"false" decision is based on whether the overall *proportion* of X exemplars with the Y trait is greater than a "true" criterion or less than a (different) "false" criterion. Suppose further that subjects will continue to draw samples until the overall proportion is above the "true" or below the "false" cut point. Finally, assume that decision times are an increasing function of the number of samples that are drawn. This model does not predict the slope changes. It also makes decision time predictions that are identical to the feature comparison model. In fact, it should be obvious that the two new models assume virtually identical decision processes and differ primarily in the way in which trait terms are represented in memory, and therefore in the dimension that underlies the true/false criterial decision.

While the exemplar scanning model can perhaps be salvaged with the modifications suggested above, it is important to realize that the data have constrained that model so that the decision processes in this new model have much in common with those of the feature comparison model. Although it appears that we cannot decide from the data pre-

sented here whether traits are represented in memory as feature lists, exemplars, or even both, consistent with the original goal of this research, we can conclude that (a) whatever the mental representation of traits may be, the similarity of those representations seems to play a major role in implicit personality inferences; (b) decisions about people's personality characteristics seem to be governed by a multistage process in which some answers are based on a less extensive (or detailed) search of memory than other answers (the search is not exhaustive in all cases); (c) a different criterion seems to be used for "true" than for "false" responses; and (d) *both* evidentiary criteria seem to change as the nature of the assertion changes. Thus, models that argue that all decisions are made after an exhaustive scan or processing of the relevant evidence cannot explain the current results. Such models would, in general, not be able to explain the effects of quantification on decision times found in these experiments. Similarly, models that assume that exhaustive scans or processing occur as a default option when sufficient evidence for a "true" response is not found (e.g., Collins & Quillian, 1969) are incapable of explaining the changes in mean "false" response times observed in both experiments.

On the other hand, almost any view of the representation of trait terms (e.g., prototype, schema, script, network, feature list, set of exemplars, visual image, etc.) could be postulated and combined with a multistage processing model similar to that described here, so as to explain the results from these experiments. This fact suggests the need for caution when claiming that social information takes a particular form in memory. As we noted in the introduction, most models of cognitive–social phenomena include assumptions about the structure of information in memory, the mechanisms by which that information is retrieved from memory, and the decision process that operates on that information. It can happen, therefore, that very different assumptions about one or more of these aspects of semantic decisions can yield identical predictions, provided other aspects of the model take an appropriate form (see Anderson, 1978, and Townsend, 1974, for similar views). For this reason, we believe caution must be exercised when concluding that a given set of results validates a particular structural view or a particular retrieval process.

It is of interest to note that a process model that provides a useful explanation of implicit personality inferences also provides a useful explanation of decisions about entirely different kinds of semantic information. Apparently, the same model can be used to understand how people decide whether a mother is a chair as whether an intelligent person is also honest. The same model also seems to apply to other social stereotypes besides traits. For example, Cohen (1977) reported results of a reaction time experiment in which people decided whether particular attributes were characteristic of given occupations. "False" reaction times increased and "true" reaction times decreased as the rated typicality of the attribute to the occupation increased. It even appears that decisions about the relevance of particular attributes to oneself are governed by a dual-criterion multistage decision process. Markus (1977) found that reaction times for "like me" decisions about trait attributes decreased as the rated self-applicability of the trait terms increased. In short, there is reason to believe that questions about semantic information, whether that information is about traits, objects, occupations, oneself, or something else, seem to be answered by a multistage process in which the similarity of the elements in the question is compared to task-established evidentiary criteria.

It is both somewhat surprising and important that the present dual-criterion multistage model appears to have such wide generality. As mentioned above, it seems likely that similar, if not identical, processes may be found for a variety of social psychological inference processes besides those involved in implicit personality inferences.

Reference Note

1. Kruskal, J. B., Young, W., & Seery, J. B. *How to use KYST, a very flexible program to do multidimensional scaling and unfolding.* Bell Laboratories technical memorandum, 1973.

References

Allen, R. B. FOEP, a Fortran-based experimental package. *Behavior Research Methods and Instrumentation*, 1978, *10*, 63–64.

Anderson, J. R. Arguments concerning representations for mental imagery. *Psychological Review*, 1978, *85*, 249–277.

Anderson, N. H. Cognitive algebra: Integration theory applied to social attribution. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 7). New York: Academic Press, 1974.

Asch, S. E. Forming impressions of personality. *Journal of Abnormal and Social Psychology*, 1946, *41*, 258–290.

Bruner, J. S., Shapiro, D., & Tagiuri, R. The meaning of traits in isolation and in combination. In R. Tagiuri & R. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford, Calif.: Stanford University Press, 1958.

Bruner, J. S., & Tagiuri, R. The perception of people. In G. Lindzey (Ed.), *Handbook of social psychology* (Vol. 2). Cambridge, Mass.: Addison-Wesley, 1954.

Cantor, N., & Mischel, W. Traits as prototypes: Effects on recognition memory. *Journal of Personality and Social Psychology*, 1977, *35*, 38–48.

Clark, H. H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 335–359.

Cohen, C. E. Cognitive basis of stereotyping: An information processing approach to social perception (Doctoral dissertation, University of California, San Diego, 1976). *Dissertation Abstracts International*, 1977, *38*, 412B–413B. (University Microfilms No. 77-13,681)

Collins, A. M., & Quillian, M. R. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 1969, *8*, 240–247.

D'Andrade, R. G. Memory and the assessment of behavior. In H. M. Blalock, Jr. (Ed.), *Measurement in the social sciences*. Chicago: Aldine, 1974.

Fiske, D. W. *Strategies for personality research*. San Francisco: Jossey-Bass, 1978.

Hays, W. L. An approach to the study of trait implication and trait similarity. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford, Calif.: Stanford University Press, 1958.

Johnson, S. C. Hierarchical clustering schemes. *Psychometrika*, 1967, *32*, 241–254.

Kahneman, D., & Tversky, A. Subjective probability: A judgment of representativeness. *Cognitive Psychology*, 1972, *3*, 430–454.

Lay, C. H., & Jackson, D. N. Analysis of the generality of trait-inferential relationships. *Journal of Personality and Social Psychology*, 1969, *12*, 12–21.

Markus, H. Self-schemata and processing information about the self. *Journal of Personality and Social Psychology*, 1977, *35*, 63–78.

Meyer, D. E. On the representation and retrieval of stored semantic information. *Cognitive Psychology*, 1970, *1*, 242–300.

Mischel, W. *Personality and assessment*. New York: Wiley, 1968.

Mulaik, S. A. Are personality factors raters' conceptual factors? *Journal of Consulting Psychology*, 1964, *28*, 506–511.

Passini, F. T., & Norman, W. T. A universal conception of personality structure? *Journal of Personality and Social Psychology*, 1966, *4*, 44–49.

Rips, L. J. Quantification and semantic memory. *Cognitive Psychology*, 1975, *7*, 307–340.

Rosenberg, S., & Sedlak, A. Structural representations of implicit personality theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6). New York: Academic Press, 1972.

Schneider, D. J. Implicit personality theory: A review. *Psychological Bulletin*, 1973, *79*, 294–309.

Shweder, R. A. How relevant is an individual difference theory of personality? *Journal of Personality*, 1975, *43*, 455–485.

Smith, E. E., Shoben, E. J., & Rips, L. J. Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, 1974, *81*, 214–241.

Townsend, J. T. Issues and models concerning the processing of a finite number of inputs. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition*. New York: Wiley, 1974.

Walker, J. H. Real-world variability, reasonableness judgments, and memory representations for concepts. *Journal of Verbal Learning and Verbal Behavior*, 1975, *14*, 241–252.

Received August 26, 1977 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 4, 489-498

Effects of Exogenous Changes in Heart Rate on Facilitation of Thought and Resistance to Persuasion

John T. Cacioppo
University of Notre Dame

Two experiments were conducted to examine the effects of an accelerated heart rate on information processing and resistance to persuasion. Experiment 1 addressed the effects on cognitive performance of manipulating heart rate exogenously for brief periods of time. Fourteen subjects wearing implanted demand-type cardiac pacemakers performed reading comprehension and sentence generation tasks while their heart rate was either accelerated or not accelerated. Results revealed that performance was better when heart rate was accelerated than when it was not accelerated. Experiment 2 addressed the effects on counterargumentation and resistance to persuasion of manipulating heart rate using the cardiac-pacing technique employed in Experiment 1. Subjects read highly involving counterattitudinal communications while their heart rate was either ostensibly or actually accelerated. Accelerated heart rate resulted in the generation of more total thoughts and counterarguments than did basal heart rate; resistance to persuasion was related significantly to the number of counterarguments generated. The methodology used provides a means by which social psychologists can study the effects on social processes of actual but unperceived changes in physiological processes.

Much of our social behavior is affected by the state of our physiological systems. The important work on this issue by social psychologists has been concerned with the effects on cognition and behavior of *perceived* changes in the functioning of the autonomic nervous system (ANS; e.g., Detweiler & Zanna, 1976; Schachter, 1964; Valins, 1966). However, little has been done with respect to the effects of *actual* but *unperceived* changes in ANS functioning. Methodological rather than theoretical problems seem responsible for this empirical deficit (cf. J. Lacey, 1967; J. Lacey & B. Lacey, 1974). Recently, a technique has been found (Cacioppo, 1977) that may alleviate the latter shortcoming. The purpose of this article is to describe this procedure as well as to demonstrate that heart rate influences cognitive elaboration. This will be done in two ways: (a) by showing that performance on an intellectual task requiring cognitive elaboration is enhanced by increased heart rate (Experiment 1) and (b) by showing that when resistance to persuasion depends on cognitive elaboration, greater resistance (counterarguing) occurs with high heart rates (Experiment 2).

Cardiac Activity and Information Processing

Angell and Thompson (1899) were perhaps the first to conclude that performing complex cognitive tasks led to an accelerated heart rate. Somewhat later, Darrow (1929a) dis-

This article is based on a doctoral dissertation submitted to the Department of Psychology of the Ohio State University. I am especially grateful to Anthony G. Greenwald, Richard E. Petty, Curt A. Sandman, Timothy C. Brock, and D. Chris Anderson for their comments on an earlier draft of this manuscript, to Charles V. Meckstroth, M.D., for providing subjects and for medical supervision during the research, and to Douglas Lundy for serving as the experimenter in Experiment 2.

Parts of the article were presented at the meeting of the American Psychological Association, San Francisco, August 1977, and at the meeting of the American Association for the Advancement of Science, Washington, D.C., February 1978.

Requests for reprints should be sent to John T. Cacioppo, Department of Psychology, University of Notre Dame, Notre Dame, Indiana 46556.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3704-489$00.75

tinguished between stimuli that led to "excitation of sensory end organs causing immediate physiological effects but calling for no extensive association of ideas" from those that led to "excitation in which the stimulation of sensory end organs is but incidental to the initiation of associative processes" (p. 185). He concluded from a review of the literature that (a) momentary sensory stimulation was accompanied generally by a decelerated heart rate and (b) "ideational stimuli" were associated with an accelerated heart rate. This review and an experimental study by Darrow (1929b) were followed by a lull of almost three decades in this area of research (J. Lacey, 1959; J. Lacey & B. Lacey, 1958).

Contemporary investigations of cardiovascular psychophysiology have included (a) monitoring heart rate during the anticipation and performance of a wide variety of cognitive (e.g., sentence generation) and sensory (e.g., viewing flashing lights) tasks (see J. Lacey, 1959; J. Lacey, Kagan, B. Lacey, & Moss, 1963; Obrist, 1963); (b) the measurement of the differences in heart rate change during performance by individuals who differ dispositionally in their mode of processing task information (e.g., Blatt, 1961; Kagan & Rosman, 1964); (c) the manipulation of attributes of the task while monitoring heart rate (e.g., Cacioppo & Sandman, 1978; Campos & Johnson, 1967; B. Lacey & J. Lacey, 1974; Tursky, Schwartz, & Crider, 1970); (d) monitoring heart rate *and* somatic activity during the anticipation and performance of tasks, often under varying levels of motivation (e.g., Elliott, 1974; Obrist et al., 1974); and (e) monitoring sensory thresholds, reaction time, or cognitive and attitudinal responses following endogenous changes in heart rate (e.g., Cacioppo, Sandman, & Walker, 1978; Sandman, McCanne, Kaiser, & Diamond, 1977; Surwillo, 1971).

Methodological Considerations

Although heart rate has been found generally to covary with the cognitive complexity, or difficulty, of a task (e.g., Cacioppo, 1977), the biological basis of this association has been and continues to be debated. The Laceys (J. Lacey, 1967; J. Lacey & B. Lacey, 1958) have formulated a neurophysiological account of the variations of heart rate observed during these tasks. They have speculated that, ceteris paribus, an accelerated heart rate is associated with and facilitates cognitive elaboration, whereas a decelerated heart rate is associated with and facilitates sensory reception. Obrist and his colleagues (Obrist, Gaebelein, Shanks, Langer, & Botticelli, 1976; Obrist, Webb, Sutterer, & Howard, 1970), however, have contended that the observed changes in heart rate are accompanied by changes in somatic activity, both of which are mediated by a common central nervous system (CNS) integrating mechanism. Unfortunately, most of the research with humans on which these writers base their arguments is correlational rather than experimental.

Several design strategies might be adopted to provide information about the effects of changes in heart rate on information processing. For instance, heart rate might be varied exogenously without the individual's knowledge, to determine if changes in heart rate are sufficient for altering information processing; [1] or physiological activity except for heart rate could be varied systematically to determine if changes in heart rate are necessary for altering information processing.

The first strategy was adopted in the present research. In Experiment 1, cardiac-pacing techniques were employed to accelerate heart rate during the performance of two cognitive tasks (reading comprehension and sentence generation). These procedures involve only minimal risks; nevertheless, the research was conducted in a clinical setting under the supervision of a cardiologist. Unfortunately,

[1] Exogenous manipulations of heart rate are more appropriate for investigating causal relationships between changes in heart rate and cognitive/sensory behavior because (a) exogenous manipulations of heart rate can be accomplished without the subject knowing when and in what direction heart rate was varied, thus circumventing the confounding effects of the subject's knowledge or belief of heart rate changes (cf. Valins, 1966); and (b) if behavioral effects are observed to result from an exogenous change in heart rate, then causal direction of the observed association is established.

this setting contained no resources for the measurement of any physiological process besides heart rate, but previous research has demonstrated that paced changes in heart rate (within the range of rates employed in the present research—i.e., 72–88 beats per minute) are unaccompanied by significant changes in normal visceral, somatic, or cardiac-output responses (e.g., Gill, Jakobi, Morton, & Wechsler, 1975; Karlof, Bevegard, & Ovenfors, 1973).

The selection of heart rates for investigation and the length of the experimental sessions in this research also were restricted. The pacemaker worn by each subject had been implanted for at least 3 months and was necessary to maintain heart rate in most instances at 72 beats per minute (bpm).[2] Except for surgical procedures, paced heart rate could be changed only by the use of a magnet, which, when placed properly, immediately accelerated and maintained heart rate at 88 bpm. Thus, although it would have been desirable, the effects of a decelerated heart rate on performance could not be studied in this research. And time limitations restricted the type and number of tasks on which observations could be made.

Complex cognitive tasks were chosen for the present study because of the theory and research that suggest that an accelerated heart rate should facilitate performance on these tasks (e.g., J. Lacey, 1967; J. Lacey et al., 1963). The sentence generation task has been used in previous research and was found to be associated with an accelerated heart rate (J. Lacey et al., 1963). A task similar to reading comprehension was employed by Spence, Lugo, and Youdin (1972): Subjects were instructed to attend to, and were later asked to recall, sentences reflecting a certain theme while they listened to a 17-minute passage from a clinical interview. Spence et al. found that subjects displayed a decelerated heart rate (presumably denoting sensory intake) followed by an accelerated heart rate (presumably denoting cognitive elaboration or encoding) during the presentation of thematic sentences that were subsequently recalled. This waveform was distinct from those displayed in the absence of thematic sentences and in the presence of unrecalled thematic sentences. Since the present procedure allowed the subject to take as long as he or she wished to read the passages (hence, allowing subjects to control sensory intake by their allocation of reading time to items), we expected that an accelerated heart rate would facilitate the cognitive elaboration, or encoding, of the stimuli and thereby improve reading comprehension.

Experiment 1

Method

Subjects. Eighteen healthy and articulate outpatients at a large midwestern university hospital volunteered to participate in the experiment. Each subject had worn a permanently implanted demand-type cardiac pacemaker for at least 3 months prior to participation in the experiment. All subjects were examined by a cardiologist and, immediately afterwards, were given a pretest by the experimenter. Four subjects were not included in analyses: (a) One could not read; (b) one was extremely nervous and refused to follow instructions once the experiment had begun; (c) one was reading a passage when her daughter entered the testing room and began speaking with her; and (d) one was replaced because the placement of the magnet to increase heart rate could not be determined by the experimenter. The remaining 14 subjects completed the pretest satisfactorily: Each displayed (a) a basal heart rate of 72 bpm when a capped magnet was placed over a reed of the pacemaker, (b) an accelerated heart rate of 88 bpm when an uncapped magnet was placed over a reed of the pacemaker, and (c) an inability to identify the intervals during which heart rate was accelerated (as determined by questions asked of the subject during the pretest).

---

[2] Cardiac Pacemakers, Inc. (CPI) was the manufacturer of the pacemakers. The demand-type pacemaker is characterized by its pacing of a person's heart rate at a constant rate when natural pacing produces a rate below the preset level of the cardiac pacemaker. To our knowledge, this is the only pacemaker that provides a nonsurgical means of altering heart rate simply, exogenously, and without the subject's awareness. The preset level of the CPI pacemakers used in the present study was 72 bpm. The subjects employed in Experiments 1 and 2 required cardiac pacing when seated to attain a heart rate of 72 bpm.

Importantly, subjects in the present experiments were unable to report accurately when and how frequently their pacemaker was pulsing their heart. This finding is consistent with that of Nowlin, Eisdorfer, Whalen, and Troyer (1971); they reported subjects were unaware of paced changes in heart rate that exceeded 110 bpm.

Subjects were told to discontinue their participation at any time they wished. Four subjects left after their completion of the first series of tests (i.e., the reading comprehension test); reasons for leaving included feeling fatigued, feeling nervous, and being late for another appointment.

Materials and apparatus. Two types of cognitive tests were administered. The instructions, example passage and questions, and four test passages and questions were selected from a college-level reading comprehension test for use in the first portion of the experiment. In the second portion of the experiment, a sentence generation task was employed. The task involved spending 90 sec generating sentences that (a) consisted of exactly five words, (b) made sense, (c) had good grammatical structure, (d) consisted of at least three words not contained in a previous sentence, and (e) began with the same letter. (The letter for the practice trial was *O;* the first test-trial letter was *E;* and the letter used for the second test trial was *A*.) An Apollo stopwatch was used to time subjects during the tasks, and a CPI horseshoe magnet was used to manipulate heart rate.

Design. A 2×2 factorial design was used. Heart rate (basal vs. accelerated) served as a within-subjects factor, and the order in which heart rate was varied served as a between-subjects factor. The tasks were administered in the same sequence to all subjects. However, half of the subjects (a) read the first and third passages of the reading comprehension test and (b) performed the first test trial of the sentence generation task with a basal heart rate of 72 bpm; these subjects (c) read the second and fourth passages of the reading comprehension test and (d) performed the second test trial of the sentence generation task with an accelerated heart rate of 88 bpm. The heart rate of the remaining subjects was varied in the opposite sequence during the tasks. Subjects were assigned randomly to the order condition. (Passages served as an additional within-subjects factor for the analyses of the reading comprehension measures.)

Procedure. Subjects were tested individually while sitting calmly at a table in a quiet room. The experimental tasks were described, and an example passage and questions for the reading comprehension test were administered. Once subjects understood what they were to do and were relaxed with the procedure, the reading comprehension test was administered.

Either a capped (basal heart rate) or uncapped (accelerated heart rate) magnet was placed over a reed of the subject's pacemaker while he or she read each practice and test passage.[3] Immediately following the subject's response that he or she had completed reading the passage, the magnet was removed, the next page of the booklet was presented, and the subject answered the multiple-choice questions about the passage just read. Subjects were allowed to take as long as they wished to read each passage and answer the subsequent questions, but subjects were not allowed to return to a passage once they had reported having read it. Four passages and multiple-choice tests constituted the reading comprehension test.

Next, the nature of and instructions for the sentence generation task were explained. Subjects practiced the procedure until the requirements of the task were understood. Each subject's heart rate was monitored, and either a capped or uncapped magnet was placed over the pacemaker during the 90 sec in which sentences were generated. At the onset of the trial, the experimenter announced the letter with which each sentence was to begin, and during the trial, subjects listed the sentences on ruled paper as they generated them.

Dependent variables. The percentage of the questions for each passage that was answered correctly served as the measure of reading comprehension; the time that was spent reading each passage was also recorded and analyzed.

The dependent measure of the sentence generation task was calculated in the following manner: (a) Each sentence that was generated in accordance with all of the instructions listed was assigned a value of 1.0; (b) sentences with errors were assigned a value of $(1.0 - .2X)$, where X denoted the number of errors (0 was the least a sentence could be assigned).[4] The total points accumulated served as the measure of cognitive performance.

Results and Discussion

Fourteen subjects completed the reading comprehension task, and 10 of the 14 subjects completed the sentence generation task. A multivariate analysis of variance was conducted for the cognitive performance of the 10 subjects who completed the experiment; the dependent measures were the percentage of correct answers in the reading comprehension test and the number of points accumulated in the sentence generation task. The analyses indicated that accelerated heart rate facilitated performance, $F(3, 6) = 5.63$, $p < .05$.

Univariate analyses of variance revealed

---

[3] Since pacemakers are sometimes implanted in other than the upright position, location of the reed was determined by monitoring heart rate while placing the magnet on the skin over the upper and lower portions of the implanted pacemaker. When placement of the magnet resulted in an accelerated heart rate of 88 bpm, the placement was marked with a felt pen. Heart rate was also monitored during the pretest and test intervals. In all instances, heart rate did not vary from the paced levels.

[4] The person who scored the sentences was blind to the experimental conditions to which the subject had been assigned.

that (a) reading comprehension was better during accelerated heart rate ($M = 48.68\%$) than during basal heart rate ($M = 39.33\%$), $F(1, 12) = 5.24$, $p < .025$, one-tailed comparison; (b) performance of the sentence task was improved marginally by accelerated heart rate ($M = 3.88$) compared to basal heart rate ($M = 3.27$), $F(1, 8) = 2.87$, $p < .08$, one-tailed comparison; and (c) reading time was not affected by heart rate or order ($Fs < 1$).

A significant Heart Rate × Order interaction, $F(1, 12) = 51.68$, $p < .01$, for reading time signified that the second and fourth passages of the reading comprehension test ($M = 78.1$ sec) took longer to read, on the average, than did the first and third passages ($M = 45.4$ sec). Since this interaction was not significant for the percentage of correct answers on the reading comprehension test ($F < 1$), these results indicate that reading time did not affect the role of heart rate on cognitive performance in the present study.

Together, these results indicate that heart rate accelerated peripherally is sufficient to facilitate performance of a difficult cognitive task when the change in heart rate is brief (e.g., 30–90 sec) and ongoing physiological activity reflects a normal state.[5]

Experiment 2

Previously, multiple sessions of discriminative operant training have been employed to modify heart rate while subjects maintained a relatively constant level of somatic and respiratory activity (Cacioppo et al., 1978). The results revealed that individuals generated more counterarguments and were more resistant to persuasion when heart rate was accelerated than when it was decelerated. It appeared that brief accelerations of heart rate when unaccompanied by large increases in somatic activity produced resistance to persuasion by stimulating the cognitive elaboration of and responding to the counterattitudinal communication. Thus, it was expected that the brief increase in heart rate produced by using cardiac-pacing techniques would facilitate thought production (e.g., counterargumentation) and, ultimately, resistance to persuasion. The purpose of Experiment 2 was to provide a direct test of this hypothesis.

Method

Subjects. Twenty-four healthy and articulate outpatients at a large midwestern university hospital volunteered to participate in the experiment. Experiment 2 was conducted in the same setting and in a similar manner to Experiment 1. As before, subjects were examined first by their cardiologist and by an experimenter who was unaware of the experimental hypotheses. Each subject displayed (a) a basal heart rate of 72 bpm when a capped magnet was placed over a reed of the pacemaker, (b) an accelerated heart rate of 88 bpm when an uncapped magnet was placed over the reed of the pacemaker, and (c) an unawareness of the intervals in which heart rate was manipulated. Two subjects were deleted from analyses because the experimenter failed to follow the experimental procedure correctly.

Materials and apparatus. Subjects were given the first half of the reading comprehension test employed in Experiment 1 during a preliminary examination and instruction period. These stimuli constituted the materials used in the preliminary task. Materials for the experimental task were developed in pilot testing with four elderly subjects; communications were developed that were highly involving and counterattitudinal. The advocacies selected were that (a) all Social Security and Medicare programs be eliminated and (b) the drinking and voting age in Ohio be lowered to 13 years of age.

CPI horseshoe magnet again was used to close the reed in the pacemaker and thus raise heart rate, and an Apollo stopwatch was used to time reading intervals.

Design. A 2 × 2 factorial design was used. Heart rate (basal vs. accelerated) served as the within-subjects factor, and the order in which heart rate was varied (basal–accelerated vs. accelerated–basal)

---

[5] To assess if it was necessary to vary heart rate within a constant pattern of peripheral nervous system activity to affect cognitive performance, a portion of the experiment was replicated using college undergraduates who had no pacemakers. They either stood or lay down while performing a reading comprehension task; these postural variations caused not only heart rate to differ but also respiration, muscle tension, and blood pressure. In this situation neither reading time nor reading comprehension was affected by the heart rate of the subject during the performance of the task. One should be cautious in comparing these results with the previous experimental results because of the noncomparability of the subject samples. Nevertheless, these studies together suggest that increased heart rate facilitates cognitive performance, particularly if the increase is relatively independent of major changes in physiological activity.

served as the between-subjects factor. Again, the stimuli were presented to all subjects in the same sequence.

Procedure. The procedure was similar to that used in Experiment 1 except that subjects engaged in a longer preliminary task to all adaptation to the experimental procedures prior to the administration of the experimental task. Subjects were tested individually while sitting at a table in a quiet room. Subjects were told that they would first read two passages and answer questions about the content of each passage; afterwards, they would hear two advocacies, and their opinions about each would be solicited.

The preliminary task was then begun. Two passages of the reading comprehension test were administered. During the performance of the preliminary task, (a) heart rate while a capped and uncapped magnet was placed over a reed of the pacemaker was again determined; (b) task requirements were clarified; and (c) all subjects were assured that the changes in their heart rate were not hazardous to their health. (Greater precautions were taken in the present experiment than in Experiment 1 because of the involving nature of the experimental task.)

Subjects then performed the experimental task. Subjects were informed that two proposals pertinent to them were under consideration and that messages had been prepared to explain the proposals. Subjects read each advocacy at their own speed. After reading each, subjects were given 2.5 min. to list everything about which they had thought while reading (thoughts were reported aloud by speaking into a tape recorder) and were asked to rate their agreement with the advocacy.[6]

Dependent variables. The tape-recorded listings of thoughts were transcribed by a judge who was unaware of the experimental hypotheses. Transcribed as a "cognitive response" was any phrase or sentence expressing a single thought or idea; although grammar was not a criterion, a subject's pauses and voice inflections were used in a few instances to determine whether a phrase or a group of phrases was a single thought or idea. The total number of thoughts and ideas listed served as a measure of the amount of cognitive responding.

A second judge, unaware of the experimental conditions to which a subject was assigned, scored the transcribed cognitive responses. Each response was scored as favorable, unfavorable (i.e., counterargument), or neutral/irrelevant toward the advocacy.[7] A third judge, who also was unaware of the experimental conditions to which subjects were assigned, spot-checked the scoring of cognitive responses. The second and third judges agreed completely on the scoring of the responses. (This was due largely to the uniform nature of the responses generated—see discussion below).

The subjects were asked, "How much do you agree with the proposal that you just read about?" after listing their responses to each advocacy. Subjects responded to the question using a 15-point scale in which 1 was labeled "disagree completely" and 15 was labeled "agree completely."

Although there were procedural differences among the subjects during the preliminary task of this experiment, the percentage of questions answered correctly and the reading time were also monitored and analyzed.

Results

Twenty-two subjects completed the preliminary and experimental tasks. Multivariate analyses of variance were performed for the set of eight dependent measures; heart rate and order served as the criteria. The analyses revealed that heart rate affected cognitive and affective response, $F(8, 13) = 3.17$, $p < .05$. Univariate analyses of variance were then performed. The means are summarized in Table 1.

Preliminary task. Neither heart rate nor order affected reading comprehension or reading time during the preliminary task. Heart Rate × Order interactions were found for both reading comprehension, $F(1, 20) = 7.39$, $p < .05$, and reading time, $F(1, 20) = 11.65$, $p < .01$. These effects denoted only that the first passage was read more quickly and comprehended less completely than the second passage (see Table 1). This result is due possibly to the subjects' apprehension

---

[6] Subjects were asked to verbally report everything about which they thought during their reading of the advocacy, because pilot testing indicated subjects had noticeable difficulty in writing. This difficulty was also noted during Experiment 1, and may possibly account for the attenuated effects of heart rate on the performance of the sentence generation task observed in Experiment 1.

[7] The procedures employed for scoring were adapted from Petty and Cacioppo (1977) and Cacioppo and Petty (1979). Statements directed against the advocated position that mentioned specific unfavorable consequences, statements of alternative methods, challenges to the validity of arguments in the message, and statements of affect opposing the advocated position were counted as counterarguments. Statements in favor of the advocated position that mentioned specific favorable consequences, statements eliminating alternatives, statements that supported the validity of the message, and statements of positive affect regarding the advocacy we scored as favorable thoughts. All other statements were classified as neutral/irrelevant thoughts.

Table 1
Summary of Cognitive Measures for Preliminary and Experimental Tasks, Experiment 2

Measure	Heart rate			
	Basal (72 beats per minute)		Accelerated (88 beats per minute)	
	Order 1 (Message 2)	Order 2 (Message 1)	Order 1 (Message 1)	Order 2 (Message 2)
Preliminary task				
Reading comprehension	41.73	26.36	19.73	38.91
Reading time	96.73	47.54	47.36	92.91
Experimental task				
Total thoughts	3.82	3.00	4.82	3.55
Counterarguments	3.73	2.91	4.73	3.27
Favorable thoughts	.09	0	0	0
Neutral/irrelevant thoughts	0	.09	.09	.27
Attitude	1.67	3.82	1.91	3.36
Reading time	145.00	121.64	134.82	138.64

Note. Entries for reading comprehension are in terms of percentage of correct answers, whereas entries for reading time are in seconds. Higher means on the attitude measure indicate more agreement with the speaker's position, where 1 indicates "disagree completely" and 15 indicates "agree completely." Remaining entries designate the mean frequency observed for each type of cognitive response measured.

about the experiment and confusion about what they were to do when they started.

Experimental task. It was hypothesized that an accelerated heart rate would facilitate cognitive responding. The analyses supported the hypothesis: More total thoughts, $F(1, 20) = 20.64$, $p < .001$, and counterarguments, $F(1, 20) = 12.36$, $p < .01$, were produced when heart rate was accelerated than when it was not accelerated (see Table 1).

It was also hypothesized that heart rate would affect resistance to persuasion. The analysis of variance provided no support for this hypothesis: No treatment or interaction of treatments affected subjects' rating of agreement with the advocacies. However, an inspection of the raw data revealed that 59% of the subjects' responses to the attitude measure were 1 ("disagree completely"); thus, the attitude measure was probably insensitive to treatment effects. In other words, subjects were able to counterargue and reject completely the advocacies, whether their heart rate was accelerated (thus, facilitating cognitive responding) or not accelerated. Accordingly, a total of one favorable thought and five neutral/irrelevant thoughts were generated during the entire experiment. Within-cell correlations between the cognitive response and attitude measures were calculated and indicated that agreement with the advocacy was related (a) negatively with the number of counterarguments generated ($r = -.46$, $p < .05$); (b) positively with the number of neutral/irrelevant thoughts generated ($r = .42$, $p < .05$); and (c) negatively with the number of total thoughts generated ($r = -.44$, $p < .05$). Counterargumentation and total thoughts were correlated positively and highly ($r = .99$, $p < .001$), denoting that most of the thoughts generated were unfavorable toward the advocacies (see Table 1). Finally, counterargumentation and the number of neutral/irrelevant thoughts were related negatively ($r = -.35$, $p < .05$).

The positive association between the number of neutral/irrelevant thoughts generated and susceptibility to persuasion was unexpected. The production of neutral/irrelevant thoughts may indicate that a person was momentarily or partially distracted from responding cognitively to the advocacy while reading. Previous research has demonstrated that distraction during the presentation of a counterattitudinal communication increases susceptibility to persuasion by inhibiting the production of counterarguments (cf. Petty, Wells, & Brock, 1976). Thus, the generation

Gill, C., Jakobi, W., Morton, T., & Wechsler, A. *The cardiovascular system as it relates to heart pacing: Medtronics Currents* (Vol. 1, 2nd ed.). Minneapolis, Minn.: Medtronics, 1975.

Harris, V. A., Katkin, E. S., Lick, J. R., & Habberfield, T. Paced respiration as a technique for the modification of autonomic response to stress. *Psychophysiology*, 1976, *13*, 386–391.

Kagan, J., & Rosman, B. L. Cardiac and respiratory correlates of attention and an analytic attitude. *Journal of Experimental Child Psychology*, 1964, *1*, 50–63.

Kaiser, D. N., & Sandman, C. A. Physiological patterns accompanying complex problem solving during warning and nonwarning conditions. *Journal of Comparative and Physiological Psychology*, 1975, *89*, 357–363.

Karlof, I., Bevegard, S., & Ovenfors, C. O. Adaptation of the left ventricle to sudden changes in heart rate in patients with artificial pacemakers. *Cardiovascular Research*, 1973, *7*, 322–330.

Lacey, B. C., & Lacey, J. I. Studies of heart rate and other bodily processes in sensorimotor behavior. In P. A. Obrist, A. H. Black, J. Brener, & L. V. DiCara (Eds.), *Cardiovascular psychophysiology*. Chicago: Aldine, 1974.

Lacey, J. I. Psychophysiological approaches to the evaluation of psychotherapeutic process and outcome. In E. A. Rubinstein & M. B. Parloff (Eds.), *Research in psychotherapy*. Washington, D.C.: American Psychological Association, 1959.

Lacey, J. I. Somatic response patterning and stress: Some revisions of activation theory. In M. H. Appley & R. Trumbull (Eds.), *Psychological stress: Issues in research*. New York: Appleton-Century-Crofts, 1967.

Lacey, J. I., Kagan, J., Lacey, B., & Moss, H. A. The visceral level: Situational determinants and behavioral correlates of autonomic response patterns. In P. H. Knapp (Ed.), *Expression of the emotions in man*. New York: International Universities Press, 1963.

Lacey, J. I., & Lacey, B. C. Verification and extension of the principle of autonomic response stereotypy. *American Journal of Psychology*, 1958, *71*, 50–73.

Lacey, J. I., & Lacey, B. C. Some autonomic–central nervous system interrelationships. In P. Black (Ed.), *Physiological correlates of emotion*. New York: Academic Press, 1970.

Lacey, J. I., & Lacey, B. C. On heart rate responses and behavior: A reply to Elliot. *Journal of Personality and Social Psychology*, 1974, *30*, 1–18.

Nowlin, J. B., Eisdorfer, C., Whalen, R., & Troyer, W. G. The effect of exogenous changes in heart rate and rhythm upon reaction time performance. *Psychophysiology*, 1971, *7*, 186–193.

Obrist, P. A. Cardiovascular differentiation of sensory stimuli. *Psychosomatic Medicine*, 1963, *25*, 450–459.

Obrist, P. A., Gaebelein, C. J., Shanks, E. M., Langer, A. W., & Botticelli, L. J. Cardiovascular behavioral interactions. In D. I. Mostofsky (Ed.), *Behavior control and modification of physiological activity*. Englewood Cliffs, N.J.: Prentice-Hall, 1976.

Obrist, P. A., Howard, J. L., Lawler, J. E., Galosy, R. A., Meyers, K. A., & Gaebelein, C. J. The cardiac–somatic interaction. In P. A. Obrist, A. H. Black, J. Brener, & L. V. DiCara (Eds.), *Cardiovascular psychophysiology*. Chicago: Aldine, 1974.

Obrist, P. A., Webb, R. A., Sutterer, J. R., & Howard, J. L. The cardiac–somatic relationship: Some reformulations. *Psychophysiology*, 1970, *6*, 569–587.

Petty, R. E., & Cacioppo, J. T. Forewarning, cognitive responding, and resistance to persuasion. *Journal of Personality and Social Psychology*, 1977, *35*, 645–655.

Petty, R. E., Ostrom, T. M., & Brock, T. C. (Eds.). *Cognitive response in persuasion*. Hillsdale, N.J.: Erlbaum, in press.

Petty, R. E., Wells, G. L., & Brock, T. C. Distraction can enhance or reduce yielding to propaganda: Thought disruption versus effort justification. *Journal of Personality and Social Psychology*, 1976, *34*, 874–884.

Sandman, C. A., McCanne, T. R., Kaiser, D. N., & Diamond, B. Heart rate and cardiac phase influences on visual perception. *Journal of Comparative and Physiological Psychology*, 1977, *91*, 189–202.

Schachter, S. The interaction of cognitive and physiological determinants of emotional state. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 1). New York: Academic Press, 1964.

Spence, D. P., Lugo, M., & Youdin, R. Cardiac change as a function of attention to and awareness of continuous verbal text. *Science*, 1972, *176*, 1344–1346.

Surwillo, W. W. Human reaction time and endogenous heart rate changes in normal subjects. *Psychophysiology*, 1971, *8*, 680–682.

Tursky, B., Schwartz, G., & Crider, A. Differential patterns of heart rate and skin resistance during a digit-transformation task. *Journal of Experimental Psychology*, 1970, *83*, 451–457.

Valins, S. Cognitive effects of false heart rate feedback. *Journal of Personality and Social Psychology*, 1966, *4*, 400–408.

Received March 1, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 4, 499-514

Encoding of Personal Information: Self–Other Differences

N. A. Kuiper
University of Calgary, Calgary, Canada, and
University of Western Ontario, London, Canada

T. B. Rogers
University of Calgary, Calgary, Canada

Self–other differences in processing personal information were investigated by having subjects make self-referent (describes you?) or other-referent (describes experimenter?) ratings of personal adjectives. Results from five studies indicated that self-ratings were consistently judged as easier to make, and subjects always placed more confidence in these judgments. An analysis of rating times showed that only adjectives with long rating times were recalled for the unknown-other-referent task (Experiments 2 and 3). In contrast, the recalled words for the self-referent task had very short rating times. This difference is explained via a "two-process" interpretation. Unknown-other-referent processing involves a relatively inefficient rehearsal or effort strategy, whereas self-referent processing involves the self as a highly organized and efficient schema. Finally, the effects of familiarity on other-referent processing were examined. A model of other processing is formulated to account for the observed changes in processing information about a familiar other (Experiments 4 and 5).

The present article is concerned with how information about the self and others is processed. The focal concern is upon the kinds of memory traces produced by judgments about the self and judgments about others. Our major goal is to delineate processing differences between these two kinds of judgments.

Some recent research and theory in personality and social psychology has been concerned with how people process personal information (Cantor & Mischel, 1977; Markus, 1977; Rogers, Kuiper, & Kirker, 1977; Ross, 1977). These researchers have begun to map out some of the biases and facilitations that emerge when people process input data about themselves and others. Some of the findings include the following (a) Adjectives processed using self-reference produce extremely deep and elaborate memory encodings (Rogers et al., 1977); (b) subjects can produce more behavioral exemplars for traits that they themselves possess (Markus, 1977); (c) there are several kinds of biases in recognition memory that are related to the personality of either a target person (Cantor & Mischel, 1977) or the subject (Rogers, Rogers, & Kuiper, in press); and (d) resistance to incorrect personal data is related to self-perception (Markus, 1977). Each of these findings sheds some light on the kinds of transformations that occur to personal data as they are absorbed into the human cognitive system.

The Role of Cognitive Prototypes or Schemata

The theoretical notion of prototype or schema has been advanced to account for the findings mentioned above. For example, Markus (1977) postulates that a person's organization, summation, and explanation of personal data involve the use of "self-schemata," which are "cognitive generalizations

The research reported herein was supported by a grant from the Canada Council. We thank Scott Kirker, Gloria Miller, and Ed Hansen for comments on earlier drafts and Barb Blakemore, Laurie Williams and Scott Kirker for their help in running subjects and data analysis.

Requests for reprints should be sent to T. B. Rogers, Department of Psychology, University of Calgary, Calgary, Alberta, Canada T2N 1N4.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3704-499$00.75

about the self, derived from past experience, that organize and guide the processing of self-related information" (p. 63). These schemata or cognitive structures are thought to interact with incoming data and produce some of the biasing and facilitating effects outlined earlier. For example, Cantor and Mischel (1977) showed subjects a series of statements describing an introvert (as well as some other statements). The subjects were then shown a list of items and asked to indicate whether each item had been in the first series of statements. These subjects showed a consistent bias to identify novel, but highly introvert-related, items as having been in the first series, when they had not. These data indicate that the subjects had abstracted the concept of introvert during the initial list presentation and used this abstraction as a comparator or reference during the identification task. When novel items representing similarity to this abstraction were presented, the subjects tended to misidentify them. The abstraction represents a prototype or schema, and its biasing effect on information processing is clearly reflected in the Cantor and Mischel data (see also Rogers et al., in press). The notion of schema or prototype has found support in most of the work concerned with processing personal data (e.g., Markus, 1977; Rogers et al., 1977).

Schemata or prototypes have been postulated to underlie processing of other kinds of information as well. For example, Posner and Keele (1970) demonstrate biases, not unlike those shown by Cantor and Mischel (1977), in a pattern recognition situation. Memory for connected discourse (Bartlett, 1932), faces (Reed, 1972), and embedded sentences (Bransford & Franks, 1971) has been interpreted using schemata or prototypes as the major theoretical construct.

Self–Other Differences in Processing Personal Information

The present article is concerned with the kinds of structures that are involved in processing information about other people. It is assumed that some kind of cognitive structure—possibly a schema or prototype—is involved when we process information about others. The critical question is whether these "other structures" are any different from the "self-schema" documented in the previously outlined research. In an effort to explore this question, the present article involves a comparison between memory encodings generated from tasks involving judgments of self and those from tasks involving judgments of others. If it is found that the self-schema and the other schema generate similar memory encodings, the contention that all personal information (both self- and other-referent) is processed using a common schema system will be supported. If, on the other hand, self- and other information produces different memory encodings, we will have evidence supporting the self as a distinctive cognitive schema. Further, it will be possible to explore the kinds of differences between self- and other information processing, with an eye to formulating a model of how we process information about other people.

The self–other comparison used in the present experiments also speaks to an important methodological point embedded in the Rogers et al. (1977) experiments. These experiments documented the relative strength of self-reference as an encoding device, thereby supporting the analyses indicating that the self is a schema. In these studies, subjects first made a series of ratings of personal adjectives. For example, subjects had to make synonymity judgments for some of the words. Different words were rated as to whether they described the subject. These rating tasks produced memory traces. The strength of the memory trace was assessed using an incidental recall task, after the ratings were completed. Subjects were instructed to remember as many as they could of the words they had rated. The results of these experiments revealed a clear recall superiority for the self-reference rating task. This finding was interpreted as indicating that the self produces strong and elaborate memory traces and in turn was offered as support for the self as a schema.

There is one important weakness in this argument. Self-reference was compared (and found superior) to rating tasks involving

udgments of synonymity, semantic specificity, rhyming quality, and type size. None of these comparison tasks involved making judgments about people. *It is possible that the superior recall for the self-reference task is due to the involvement of a person in the task and that any rating task involving a person would produce superior recall.* This "generalized person hypothesis" suggests that any rating task with an explicit person referent (self or anyone else) enhances incidental recall in this paradigm. If this hypothesis were correct, both the conclusion that the self produces an elaborate memory trace and the formulation of the self as a schema would be questioned.

The present article presents a series of experiments intended to explore the processing of self- and other-referent information using the incidental recall paradigm from Rogers et al. (1977), which derives from Craik and Tulving (1975). The first two experiments explored self–other differences in this paradigm; the later experiments were designed to elaborate the obtained results. The goal of this experimental series was to clarify the nature of self- and other-referent processing of personal data, in an attempt to better understand the cognitive structures involved in processing information about people.

The Present Experiments

The first two experiments were simple attempts to determine if there are incidental recall differences following self- or other-referent rating tasks. However, before the data are presented, it is instructive to consider what kinds of processes may be involved when a person is required to rate whether a given adjective (e.g., jolly) describes another person. These kinds of rating tasks are not foreign to the personality domain, since they are used in the assessment literature (e.g., peer ratings) and person perception research (Schneider, 1973). Our cognitive position persuades us to explore the kinds of processes and structures that may be involved in this task.

The cognitive literature gives a starting point for considering processes involved in rating other people. Several researchers have explored memory traces involved in making various ratings about people, using a photo as the stimulus. For example, subjects may rate either physical qualities of the person (e.g., size of nose) or psychological qualities (e.g., honesty). In a subsequent recognition task, subjects are better able to identify the faces given the psychological (deeper) ratings (Bower & Karlin, 1974). These data indicate that ratings of others that involve psychological qualities tend to create elaborate memory traces, compared to physical (shallow) tasks. The studies are not overly instructive as to what the underlying mechanisms might be during the rating task. The one hint is Bower and Karlin's suggestion that prototypes mediate the judgments of others.

Another area that relates to judgments about others is person perception (e.g., Schneider, 1973; Wiggins, 1973). One finding that consistently emerges in this domain is that the lay perceiver finds it easy to go from behavioral data to inferences about dispositions (Hastorf, Schneider, & Polefka, 1970). In rating another person, the rater is apt to use factors such as appearance, behavior, social relationships, typical contexts, personal origins, and internal properties (Fiske & Cox, in press). The mechanisms involved in making these inferential leaps are usually thought to involve certain classes of conceptual structures that prescribe certain behavior–trait relationships. These structures, sometimes referred to as implicit personality theories, facilitate generation of dispositional inferences.

These two domains of research (rating faces and implicit personality theory) give us some ideas as to how people may make ratings of other people. Presumably, the cognitive structures associated with the person's implicit personality theory form an integral part of this rating process. This being the case, rating another person with respect to psychological characteristics should provide a relatively elaborate (or deep) memory trace. The research involving faces supports this contention. In toto, the data suggest that trait or "deep" ratings of others should enhance incidental recall. The critical ques-

tion is whether this enhancement is comparable to that documented for self-referent judgments in the Rogers et al. (1977) experiments.

Since both the self- and other-referent tasks involve cognitive structures (self and implicit personality theories, respectively), it may be possible that they share some functions as well. It is most likely that both cognitive structures allow us to simplify and reduce the incoming data to a manageable size. Implicit personality theories are thought to allow us to "simplify the complex world of other people" (Hastorf et al., 1970, p. 16), whereas the self-schema represents the "abstracted essence of a person's perception of him or herself," which evolves to "help the person keep track of the vast amounts of self-relevant information encountered over a lifetime" (Rogers et al., 1977, p. 677). Thus, we have two cognitive structures that appear to possess very similar functions. This situation suggests that differentiation between self- and other-referent tasks may be difficult.

The difficulty of separating self- from other-referent tasks prompted the inclusion of several additional measures in the following experiments. Since differences may be subtle, measures of difficulty and confidence in making the various ratings have been included. It is hoped that these measures will provide information about the subject's perception of these tasks and offer some hints as to the nature of self- and other information processing.

Experiment 1: Self–Other Recall Differences Using the Group Method

This first experiment introduced the basic methodology of the incidental recall paradigm and offered initial data speaking to self–other referent task differences. It was followed immediately by a second experiment using a somewhat different methodology.

Method

In all of the following experiments two basic stages were involved. First, subjects made a series of ratings of a set of adjectives. The ratings were varied during this stage and were communicated to the subject by way of a cue question preceding the presentation of the adjective. Following the ratings, the second stage of the experiments involved a free-recall period in which subjects attempted to remember the words they rated.

Materials. The 48 adjectives (40 target adjectives plus 4 recency and 4 primacy buffers) represented a broad set of personality characteristics, chosen from the scale descriptions of the Personality Research Form (Jackson, 1967). The details of selection are presented in Rogers et al. (1977, p. 680).

The rating tasks used are detailed in Table 1. Four task orders were generated to ensure that each adjective was rated under each task. Within each list one quarter of the words were rated under each task. Task order was randomized in blocks of four, such that each rating task was performed once in a block.

Procedure. Data were collected during the first session of an undergraduate psychology class. With the exception of substituting "describes him" for the phonemic cue question, the procedure was identical to that in Experiment 2 of Rogers et al. (1977).

Experimenter. The target for the other-referent ratings was a male psychology graduate student. The experimental session was his first encounter with these subjects, who were students in his statistics lab.

Subjects. Twelve (11 female) undergraduate psychology students served as subjects. Their mean age was 20.2 years.

Results

Recall performance. The recall protocols were scored under the following conditions: (a) The first and last four adjectives presented were not included, since memory for these is probably due to primacy and recency effects; (b) grammatical transformations of the words were scored as wrong; and (c) a proportion correct score was employed to insure that differential numbers of Yes and No ratings were not affecting the recall scores (see Rogers et al., 1977, pp. 683–684, for details and rationale of this proportionality score).

The mean proportions correct as a function of rating task and rating (Yes or No) are presented in Table 1. As can be seen, the self-referent task produced superior recall. This result was verified in an analysis of variance of the data making up these means. The main effect of rating task was significant, $F(3, 33) = 2.98$, $p < .05$, as was the difference between the self- and other-referent tasks ($p < .05$). The post hoc test also revealed that other-referent recall was not

Table 1
Rating Tasks, Mean Adjusted Recall, and Confidence and Difficulty Ratings for Experiment 1

Rating task	Question	Definition	Mean adjusted recall			Mean difficulty rating	Mean confidence rating
			Yes rating	No rating	*M*		
Structural	Long?	Rate whether you feel the word is long or short.	.16[a]	.19	.18	3.08[b]	3.83[c]
Semantic	Specific?	Rate whether you feel the word has a specific meaning or relates to a specific situation.	.16	.13	.15	4.25	3.83
Other-referent	Describes him?	Rate whether you feel the word describes the experimenter.	.09	.27	.18	5.00	2.75
Self-referent	Describes you?	Rate whether you feel the word describes you.	.42	.21	.31	2.00	6.00
		M	.20	.20			

[a] Adjusted recall values can range from 0 to 1.00. A value of 1.00 indicates that all adjectives receiving a particular rating (either yes or no) for a particular task were recalled. A value of 0 indicates that none of these adjectives were recalled.
[b] Difficulty ratings were made on a 7-point scale with endpoints of (1) "not at all difficult" and (7) "extremely difficult."
[c] Confidence ratings were made on a 7-point scale with endpoints of (1) "not at all confident" and (7) "extremely confident."

superior to the semantic task. For the structural, semantic, and self-referent tasks, the pattern of present results replicates exactly the findings of Rogers et al. (1977).

Clustering. One possible explanation of the recall findings is that subjects simply wrote down every adjective they felt was self-descriptive during the recall phase. If they did this, their self-referent recall would appear to be excellent relative to the other tasks. This hypothesis predicts two things. First, there should be a large number of intrusions in the recall protocols. This was not found ($M = .25$ intrusions per subject). Second, if subjects were merely listing the features stored in their self-concepts, all words rated under the self-referent task would be emitted together, or in a definite cluster. The measure of clustering outlined by Dalrymple-Alford (1970) was calculated for the recall protocols to assess this possibility. This measure permits a comparison between the number of repetitions observed in the output and the number expected if the recalled words were emitted randomly. The Dalrymple-Alford measure also adjusts for lengths of protocol and the number of each kind of element observed. Significant deviations between the observed and expected number of repetitions would support the hypothesis that the superior self-referent recall is due to a simple listing of the features comprising the self.

The series analyzed for each subject was the task under which each successive item in the recall protocol was rated. While there was a tendency for there to be more repetitions observed than expected ($M = 3.17$ and 2.61, respectively), these figures never approached statistical significance, $\chi^2(11) = 1.13$, p (sign test) $= .61$, suggesting that output order was close to that expected by chance. This difference remained nonsignificant even when several variations of the Dalrymple-Alford measure were used. To save space, it will be mentioned here that the same results were obtained for all recall data reported in this article, suggesting that there is not a meaningful amount of clustering in the recall data of these experiments. As suggested in Rogers et al. (1977), the incidental recall data appear to reflect more encoding than retrieval processes.

Difficulty and confidence ratings. The means of the postexperimental ratings of the encoding tasks are presented in Table 1. One-way analyses of variance of these ratings revealed significant effects for both the difficulty, $F(3, 33) = 9.04$, $p < .001$, and confidence, $F(3, 33) = 16.22$, $p < .001$ ratings. Post hoc tests revealed that the subjects found the other-referent task significantly more difficult than the self-referent rating ($p < .01$). They also placed a significantly greater degree of confidence in the

self-referent, compared to the other-referent, task ($p < .01$).

Experiment 2: Self–Other Recall Differences Using the Individual Method

Method

Experiment 2 was conducted to extend the generality of the self–other difference found in the first study by incorporating three methodological changes. First, a different other–target was used to ascertain if the target was an important aspect of the ratings. Second, the semantic rating task was changed to a synonymity rating. Third, subjects were tested individually, which permitted assessment of the time required to make the ratings (RT). The procedures outlined in Rogers et al. (1977, Experiment 1) were used. The only change was that the rhythmic task was replaced with an other-referent task.

Materials. The adjectives from Experiment 1 were used. The necessary synonym list for the new semantic task was taken from Experiment 1 of Rogers et al. (1977).

Procedure. All of the stimuli were presented on a television monitor driven by a PDP8/1 computer, which also recorded the ratings (yes or no) and rating times (in msec). Each of the 48 trials consisted of (a) a 3-sec presentation of the cue question, (b) a ½-sec blank interval, (c) presentation of the adjective, (d) a response by the subject that terminated the visual display, and (e) a 2-sec intertrial interval. Following the 48 trials the subject was allowed 3 min to recall, in any order, as many of the words as he or she could.

Experimenter. The rating target for the other-referent condition was a female psychology graduate student, who was unknown to the subjects. The only contact between her and a given subject was a telephone call asking for the subject's participation and the 5-min period before the experiment began, which was filled with greetings and instructions.

Subjects. The 24 volunteers (14 females) who took part were students in introductory psychology courses. Their mean age was 18.5 years, and each was paid $2 for his or her participation. They were randomly assigned to one of the eight lists.

Results

Recall performance. The mean proportion of items recalled as a function of rating task and rating is presented in the upper panel of Table 2. As can be seen, the self-referent task produced superior recall, followed by the other-referent encoding task. As in Rogers et al. (1977), the structural task produced the poorest recall. An analysis of variance of these recall data revealed a main effect of rating task, $F(3, 69) = 15.67$, $p < .001$, replicating Experiment 1. However, a fine-grained analysis of this main effect revealed that the self- and other-referent recall difference did not reach statistical significance.

Response time. RT was defined as the time from the onset of the adjective on the monitor until the subject pressed the button revealing a Yes or No response. These RTs were calculated separately for each rating task for each rating (yes or no); the means are presented in the second panel of Table 2. A two-way analysis of variance of the data involved in these means revealed a single meaningful main effect of rating task, $F(3, 69) = 27.05$, $p < .001$.

The focal concern of this study is the memory trace laid down by the rating tasks. When considering RTs for *all* ratings, we are dealing with nonrecalled words as well as recalled ones. Since nonrecalled words generated nonfunctional traces, the overall RTs do not speak directly to the issue of useful memory traces. Because of this, mean RT was analyzed as a function of whether a word was recalled. For the two person tasks, the mean RT for recalled and nonrecalled words was calculated, and the means are entered in the third panel of Table 2.[1] The critical observation of these data is the 848-msec difference between the two rating tasks for recalled words. Analysis of variance of the 2×2 matrices ($n = 23$) revealed a main effect of rating task, $F(1, 22) = 12.02$, $p < .01$, and an interaction, $F(1, 22) = 4.42$, $p < .05$. Post hoc tests revealed that for the other-referent task, the recalled words had meaningfully higher RTs than the nonrecalled words. Further, considering only the recalled words, the self-referent task produced significantly shorter RTs than the other-referent task. All other comparisons were nonsignificant.

Confidence and difficulty ratings. As can

[1] It was not possible to fill sufficient cells for this analysis if the rating (yes or no) variable was retained, hence it was not included. One subject was eliminated from this analysis because of an empty cell.

be seen from the lower panel of Table 2, the confidence and difficulty ratings replicate those shown in Table 1. Analysis of variance of the confidence rating means revealed a significant effect of rating task, $F(3, 69) = 25.29$, $p < .001$, with post hoc tests revealing significance for the self/other-referent comparison ($p < .01$). For the difficulty ratings the self/other comparison of the significant main effect of rating task, $F(3, 69) = 54.73$, $p < .001$, revealed that the other-referent task was rated as significantly ($p < .01$) more difficult to perform than the self-referent judgment.

Discussion

The results of these two experiments indicate that (a) more adjectives rated under the self-reference task were remembered compared to the remaining encoding tasks; (b) clustering analyses revealed no unusual patterning in the recall protocols; (c) RT for encoding interacted with rating task and recall of words; and (d) difficulty and confidence ratings indicated that self-reference judgments were easier and were made more confidently than the other-referent judgments. These results, found using two different methods, paint an interesting picture of the differences between the self- and other-referent rating tasks.

First, the recall data suggest that the "general person hypothesis" is not tenable. In Experiment 1 an encoding task involving another person did not produce recall equivalent to the self-reference task. This finding suggests that it is not simply the involvement of a person per se in the encoding rating that facilitated recall. Rather, the superior self-reference recall documented here and in Rogers et al. (1977) is the result of the self being active during the encoding task. Experiment 2 showed a similar pattern of results, but fine-grained analysis revealed that the self- and other-referent recall were

Table 2
Summary of Results for Experiment 2

Measure	Rating task				*M*
	Structural	Semantic	Other-referent	Self-referent	
Mean proportion correct recall					
Yes rating	.09[b]	.20	.26	.35	.23
No rating	.07	.12	.25	.30	.19
M	.08	.16	.26	.33	
Overall mean RT[a] (in msec)					
Yes rating	1,280	2,639	2,855	2,536	2,327
No rating	1,449	2,632	2,837	2,674	2,398
M	1,364	2,635	2,846	2,605	
Mean RT[a] for selected subjects ($n = 23$)					
Recalled words			3,223	2,375	2,799
Nonrecalled words			2,714	2,558	2,636
M			2,968	2,466	
Mean difficulty rating	1.33[c]	3.20	5.66	2.83	
Mean confidence rating	6.45[d]	4.83	2.79	5.37	

[a] RT = response time.
[b] Adjusted recall values can range from 0 to 1.00. A value of 1.00 indicates that all adjectives receiving a particular rating (either yes or no) for a particular task were recalled. A value of 0 indicates that none of these adjectives were recalled.
[c] Difficulty ratings were made on a 7-point scale with endpoints of (1) "not at all difficult" and (7) "Extremely difficult."
[d] Confidence ratings were made on a 7-point scale with endpoints of (1) "not at all confident" and (7) "extremely confident."

not significantly different. This result could be due to any number of factors relating to the different methods employed (e.g., the group method is experimenter paced, whereas the individual method is self-paced). This failure to find self–other differences could be damaging for our self theory, were it not for the consistent pattern of results emerging from the other dependent measures, particularly the RT and rating data.

The major difference between the self- and other-referent tasks was that words recalled under the other-referent task were those upon which the subject expended a large amount of effort, as witnessed by the inflated RTs ($M = 3{,}223$ msec). When effort was not expended during the other-referent rating, the word tended to not be recalled ($M = 2{,}714$ msec). The opposite pattern of results was obtained for the self-rating task, with recalled words showing faster RTs ($M = 2{,}375$ msec) compared to those that were not recalled ($M = 2{,}558$ msec). Thus, recall under the other-referent condition is "effort tied," whereas self-reference recall is not. The finding that subjects found the self-referent task easier and placed more confidence in their self-ratings compared to the other-referent task underscores the differences between these two encoding tasks. Overall, these results indicate that the self- and other-referent tasks involve two different types of cognitive processes.

The other-referent task can be characterized as a rehearsal or frequency type of process. Here, if the subject works hard during encoding (e.g., rehearses or repeats the adjective), he or she is apt to recall the item. The absence of this effort during encoding predicts poor recall. This type of process is reminiscent of verbal learning frequency theory (e.g., Klatzky, 1975) in that there is a simple linear relationship between RT and recall. The effort during encoding could be sheer rote repetition (though this is unlikely) or related to trying to "pigeonhole" or form an impression of the target person. This process may involve some kind of a schema or prototype, albeit a somewhat weak one. Subjects clearly indicated they were not confident of their other-referent ratings and found them difficult to make, which reinforces this rehearsal-type interpretation.

The data suggest a very different process for the self-referent ratings. The recalled words do not show inflated RTs during the ratings. Rather, there is even a hint that the words that were easiest to rate are those that were recalled. This suggests that words that were easily integrated into the self-schema were those that were recalled. The finding that subjects found the self-referent task easy and had confidence in their ratings supports this interpretation. It appears that the self-referent task is based on an efficient schema. Presumably, the adjective is compared to the self-concept as it is rated (see Rogers, 1974), and the appropriate response is generated. This involvement of the self produces a rich and functional memory trace, which facilitates recall. The present data do not permit precise inferences about how the self-ratings are actually performed. However, the results indicate that these ratings involve the use of a highly organized and efficient self-schema and that the process involved is very different from the other-referent rating.

These data, then, suggest a "two-process" interpretation of self–other differences in processing information. Data about others are processed using a frequency of rehearsal process, whereas self-referent data involve the use of a highly organized cognitive schema.

Experiment 3: A Closer Look at the Process Differences

Despite the relatively strong support for the two-process interpretation of self- and other-referent tasks provided by the first two experiments, several weaknesses remain. First, it is a post hoc explanation and thereby demands independent testing. Second, it was not possible to analyze the RT data of Experiment 2 for the effects of rating (yes or no). Third, lack of sufficient observations forced elimination of one subject in the previous experiment, which attenuates generalizability. Experiment 3 was designed to overcome these problems by (a) providing a direct test of the two-process interpretation, (b) ensuring a sufficient number of ratings

to permit assessment of yes/no variations, and (c) increasing the number of ratings per task to eliminate any possibility of empty cells in the subject data matrices. In order to fulfill (b) and (c) without making the overall task unmanageable, the structural and semantic tasks were eliminated, thereby doubling the number of self- and other-referent ratings to 20 per task. A further change was the use of yet a third target for the other-referent task.

Predictions in this study relate to the two-process interpretation presented above. If the other-referent task involves frequency or rehearsal processes *and* the self-referent rating is based on an efficient schema, RTs should be greater for the other-referent task. Furthermore, this effect should be stronger for recalled words, predicting an interaction between rating task and recall (recalled and nonrecalled words). It is even possible that the effect would emerge as a three-way interaction when the rating (yes or no) variable is added.

Method

All aspects of this study were identical to Experiment 2 except that only self- and other-referent tasks were used. This meant that only two stimulus orders were necessary to counterbalance adjectives across rating tasks.

Experimenter. A female psychology graduate student ran the subjects and served as the target for the other-referent ratings. She was not the same experimenter used in Experiment 2.

Subjects. Twelve undergraduate volunteers (six female) served as subjects. Each was paid $1.50 for participating. Their mean age was 20.3 years.

Results and Discussion

The adjusted recall as a function of rating task and rating was calculated. The other-referent recall ($M = .23$) was less than the self-referent recall ($M = .28$), replicating the previous experiments. However, as with Experiment 2, this difference was not significant, $F(1, 11) = 1.01$. Furthermore, the remaining terms in the adjusted recall analysis of variance were also nonsignificant.

The mean times required to make the self- and other-referent ratings were calculated for each subject. A t test on these data revealed that other-referent ratings required significantly more time ($M = 3{,}784$ msec) than did the self-referent ($M = 3{,}191$ msec) ratings, $t(11) = 2.98$, $p < .01$. This pattern of results also replicates Experiment 2 and offers preliminary support for the two-process interpretation.

Table 3
Mean Reaction Time (in msec) as a Function of Rating Task, Recall, and Rating for Experiment 3

Rating	Rating task: Other-referent	Rating task: Self-referent	*M*
Recalled words			
Yes	5,133	2,936	4,035
No	4,054	5,124	4,589
M	4,594	4,030	4,312
Nonrecalled words			
Yes	3,461	3,119	3,290
No	3,843	3,488	3,666
M	3,652	3,304	3,478

Note. $n = 12$.

The important analysis in this experiment involves RTs as a function of rating task, recall, and rating. Such a matrix was calculated for each subject; the overall means based on all subjects are presented in Table 3. The noteworthy aspect of these means is the evident "crossover" of RTs in relation to rating and rating task *for the recalled words only*. This crossover emerges statistically as a significant three-way interaction, $F(1, 11) = 6.77$, $p < .025$, in the analysis of variance conducted on these data. Other significant terms in this analysis were (a) the main effect of recall, $F(1, 11) = 21.34$, $p < .001$, indicating that RTs were higher for recalled words, and (b) the Rating Task × Rating interaction, $F(1, 11) = 8.50$, $p < .025$, which is clearly qualified by the triple interaction. The main effect of rating task was not significant.[2]

[2] This finding may appear incongruous with the earlier "overall RT" analysis, but it is not. The categorization procedure used to derive the eight mean RTs per subject weighs the raw RTs unequally. The overall RT analysis assigned an equal weight to each raw observation.

These data indicate that the Yes-rated RTs neatly replicate the pattern of results observed in Experiment 2 (Table 2). These data are a second verification of the two-process interpretation, wherein it appears that the self- and other-referent tasks involve schema-based and rehearsal-based processes, respectively.

The words receiving No ratings paint an almost opposite picture to the Yes-rated adjective. The longest RT in the No condition was for recalled self-referent words (5,124 msec), which was the shortest RT for Yes words (2,936 msec). However, post hoc tests failed to reveal any significant differences within the No words, which suggests that the Yes-rated words carried the useful information for differentiating other- and self-referent processing. This result reinforces the conclusions in Rogers et al. (1977), where emphasis was placed on the Yes-rated words in their schema interpretation.

Experiment 4: Familiarity of the Other-Target

The three studies reported above were concerned with an unfamiliar other. Care was taken to ensure a minimum of contact between the subjects and the other-referent target. This was done to minimize possible confounds in the designs associated with how well the subjects knew the other-target. However, it is possible that familiarity is an important aspect of other-referent ratings, and as such deserves systematic investigation. Experiments 4 and 5 considered this possibility, using the same paradigm as the previous studies.

There are some hints in the literature that familiarity does affect processing of information about others. Koltuv (1962) observed a simpler multivariate structure for judgments of an unfamiliar other, compared to known targets. This led Hastorf et al. (1970) to suggest the following:

> Implicit theories of personality in the form of assumptions about which traits are related to one another operate most strongly when the perceiver faces an ambiguous person, one he does not know well. On the other hand, when the perceiver is rating people he knows something about, such stereotyped inferences are modified to fit actual characteristics of the other more closely. (p. 44)

There is also evidence to suggest that degree of acquaintance has a powerful effect on the validity of trait inferences (Norman & Goldberg, 1966). Willingness to ascribe a trait to a person has been found to decrease with familiarity (Nisbett, Caputo, Legant, & Marecek, 1973). Each of these studies suggests that familiarity with the other is a salient attribute of how we process personal information.

The ratings of an unfamiliar other in the first three studies of the present article were probably based on a stereotype or expectancy of the psychology graduate student who served as the target (Jones & Gerard, 1967). During this time the subjects were trying to develop a cognitive structure for the target, but the lack of available information forced recourse to a stereotype (Secord & Backman, 1974). This suggests that effort in making the judgment is likely to be an important variable, which was reflected in the RT analyses. However, as familiarity increases, two critical elements change. First, there is more information available, and second, there is more time to formulate a cognitive organization or structure for that person. These two factors should lead to a reasonably accurate and organized cognitive structure as familiarity increases (e.g., Hastorf et al., 1970). Other-referent decisions should then be based on this new organization and produce reasonably elaborate memory encodings. Furthermore, during the recall phase, this organization can become part of the retrieval environment and thereby facilitate recall. All of this suggests that other-referent ratings for a familiar person should produce excellent recall in the incidental recall paradigm.

Method

Experiment 4 was an exact replication of Experiment 1, involving the same subjects and the same experimenter/other-target. The only difference was that the experiment was run 11 weeks after Experiment 1. During this period the experimenter became more familiar to the subjects, since he was their lab instructor in the course. One feature added after the recall period was the collection of several familiarity ratings.

Results and Discussion

The familiarity ratings indicated that the subjects felt they knew the experimenter "moderately well" and a little less than a "casual acquaintance." Almost all subjects indicated that their interaction with the experimenter had been limited to the classroom and the instructor's office.

The adjusted recall figures, the means of which are presented in Table 4, were subjected to an analysis of variance with rating tasks and ratings as factors. The only significant effect was the main effect of rating task, $F(3, 33) = 4.23$, $p < .025$. Post hoc tests failed to reveal a self-referent–other-referent difference, but recall for both person tasks was superior to that of the semantic task.

The difficulty ratings showed a pattern of significance similar to that of Experiment 1, $F(3, 33) = 7.09$, $p < .001$, with the self-referent judgments being rated as significantly easier than the other-referent judgments ($p < .05$). The confidence ratings also replicated those of Experiment 1, $F(3, 33) = 14.96$, $p < .001$, with the self-referent judgments receiving significantly higher confidence ratings ($p < .05$) than all of the remaining tasks.

These data confirm the predictions deriving from a view that suggests that familiarity results in the development of an accurate organization. The enhanced recall under the other-referent rating ($M = .18$ in Experiment 1 and .40 in Experiment 4) appears to be due, in part, to the increased familiarity with the other-target, derived from 11 weeks of experience with him. The self-referent recall is roughly the same for the two experiments (Ms $= .31$ and .32, respectively). Hence, the effect of familiarity is to dramatically enhance incidental recall for words rated under an other-referent task. This result suggests the development of a cognitive structure for the now-familiar other.

Table 4
Mean Adjusted Recall and Confidence and Difficulty Ratings for Experiments 4 and 5

Rating	Rating task: Structural	Semantic	Other-Referent	Self-Referent	*M*
		Experiment 4 ($n = 12$)			
Mean adjusted recall					
Yes	.10[a]	.24	.30	.35	.24
No	.22	.29	.50	.27	.32
M	.16	.26	.40	.32	
Mean difficulty	3.66[b]	4.66	3.83	2.16	
Mean confidence	3.42[c]	3.00	3.16	6.00	
		Experiment 5 ($n = 15$)			
Mean adjusted recall					
Yes	.09	.20	.37	.39	.26
No	.17	.07	.28	.29	.20
M	.13	.13	.32	.34	
Mean difficulty	3.20	4.53	4.60	1.60	
Mean confidence	3.66	3.53	3.06	6.13	

[a] Adjusted recall values can range from 0 to 1.00. A value of 1.00 indicates that all adjectives receiving a particular rating (either yes or no) for a particular task were recalled. A value of 0 indicates that none of these adjectives were recalled.

[b] Difficulty ratings were made on a 7-point scale with endpoints of (1) "not at all difficult" and (7) "extremely difficult."

[c] Confidence ratings were made on a 7-point scale with endpoints of (1) "not at all confident" and (7) "extremely confident."

Experiment 5: Intentional Versus Incidental Recall

The repeated experimental design employed in Experiments 1 and 4 brings one obvious problem with it. Since subjects had been tested previously in the same situation, possibly the observed effects of Experiment 4 were due to a sensitizing effect (see Campbell & Stanley, 1967, p. 18). The second study (Experiment 4) represented an intentional learning situation, whereas Experiment 1 was an incidental paradigm. An examination of the cognitive literature (Postman, 1964) indicates that an intentional/incidental manipulation may, under certain circumstances, affect subsequent recall levels. To rectify this, a second group of subjects was run. These subjects were in the same course as the previous subjects, but in a different lab section. The same experimenter was the lab instructor for this section. When tested at the same time of year as the subjects in Experiment 4, these subjects had an equivalent degree of familiarity with the experimenter—but had not been sensitized with a pretest. If the results of Experiment 4 are replicated, the sensitizing effect will be ruled out as an alternate explanation.

Method

The same procedures and materials employed in Experiment 4 were used in this study. Subjects were 15 students (3 male) in an undergraduate psychology course, with a mean age of 19.8 years. They were run during a lab session.

Results and Discussion

The familiarity ratings of this group were the same as those found for Experiment 4. The data from this experiment are summarized in Table 4. The analyses of these data revealed the same pattern as observed with Experiment 4: (a) a main effect of rating task, $F(3, 42) = 9.89$, $p < .001$; (b) equivalent levels of recall for both the self- and other-referent tasks; (c) significant recall differences between the two person tasks and the semantic judgment; and (d) significantly lower confidence ($p < .01$) and higher difficulty ($p < .01$) ratings for the other-referent task compared to the self-reference condition. The major impact of these findings is that they permit rejection of a sensitizing effect as an explanation of the data in Experiment 4.

Taken in combination, the results from Experiments 4 and 5 suggest that familiarity with a person tends to enhance the probability of generating an elaborate and functional memory trace in the incidental recall paradigm. It appears that familiarity aids in the development of a cognitive structure that can be used to process information about that person. Presumably, this structure becomes more elaborate and embellished with increased familiarity.

General Discussion

The data from the five experiments construct a consistent argument for several propositions regarding the processing of information about people. The experiments leave little doubt that the processes attending other- and self-referent decisions are different. This difference is reflected in the consistent differences in RTs observed in Experiments 2 and 3. Not only did other-referent decisions require more time than self-referent ratings, but the RTs also showed some interesting interactions with the various independent variables. Of particular interest is the finding that words recalled from the other-referent condition were those requiring a greater amount of time for the rating to be performed. All five experiments revealed consistent patterns in the confidence and difficulty ratings, indicating that self-referent decisions were easier and produced more confidence in the given response. Within the constraints of the RT findings and the ratings data, the recall results from the five studies also support the contention that self- and other-referent decisions represent different processes. This two-process interpretation has a number of implications for the study of processing personal information.

The Self as a Cognitive Schema

First, the pattern of results supports the proposition that the self functions like a

schema. In all five studies, words receiving Yes ratings under the self-referent conditions were better recalled than No-rated words. This finding suggests that self-descriptive words, presumably those that "fit" with the self, produce strong and elaborate memory traces.[3] Such results would be expected if a schema is the basis of the self-referent decision.

Another aspect of the present data supports the self-as-schema view. Not only were self-referent decisions faster than other-referent judgments but they also showed an opposite pattern in regard to which adjectives were recalled. In the other-referent task, words with long RTs were recalled, whereas with the self-referent task, there is a hint that words with short RTs—presumably those that "fit" with the self—were those that were recalled. Again, a view that suggests the self is a schema predicts this pattern of results. When these two pieces of supporting evidence are combined with published data, an impressive case for the schema property of the self can be made. Some of this convergent evidence includes the following: (a) Greater amounts of information are available for terms in the self-concept, which is to be expected from a schema (Markus, 1977); (b) resistance to incorrect personal feedback is related to the degree of self-description (Markus, 1977), which demonstrates the activity of the self-schema during the processing of new information; and (c) several biases documented in recognition memory mirror those found for experimentally manipulated schemata (Rogers et al., in press).

All of this evidence suggests that the self is an important agent in the processing of personal information. Each of these lines of evidence can be predicted from a position that defines the self as an abstract cognitive structure that contains both general trait-like entries and some specific behavioral exemplars or instances. This memory structure is active during the input and interpretation of self-related information and provides a degree of "meaning" or embellishment to the incoming information. The self can be seen to act as a hook or interpretative frame for the encoding of personal data. The clustering analysis of the present data clearly indicate that the self is active during encoding. If the data were a retrieval phenomenon or simply reflected a listing of self-referent features, some degree of clustering in the output protocols should have been observed. Since this was not found, the present data reinforce the proposition that the self is a powerful agent during the encoding of personal information.

Processing Information About Unknown Others

A second aspect of the present data is the information they provide us about how we rate other people. The results for the other-referent conditions in the current experiments provide several suggestions about the mechanisms involved in making decisions about the psychological characteristics of another person. Ratings of an unfamiliar other person appear to involve a relatively inefficient frequency type of process, with the memory trace relating to the amount of effort that the rater puts into his or her response. This effort probably relates to attempting to form an impression of the other person. Secord and Backman (1974) suggest that the context, the target, *and* the observer contribute to the initial view of an unfamiliar other. People do use *contextual information* in forming impressions (e.g., Price, 1974). This fact sometimes emerges in the form of the context determining what information about a person is salient for a given situation. The *target person* elicits a considerable number of cues, even in a short exposure. Several researchers have provided categories of the kinds of information that might be involved (Beach & Wertheimer, 1961; Secord & Backman, 1974). Fiske and Cox (in press) suggest the following categories: (a) traits, (b) demographic parameters (e.g., race, sex), (c) spe-

---

[3] As indicated in Rogers et al. (1977), this difference fails to reach significance in individual studies. However, when we combine the previous research with the present five studies, we have nine separate verifications of Yes-rated words showing significantly better recall than No-rated words in the self-reference task, $p = 1/512 = .002$.

cific behaviors (e.g., verbal output *and* nonverbal cues—see Secord & Backman, 1974, p. 43), and (d) appearance (e.g., Jones & Gerard, 1967).

A third source of information about an unfamiliar other is the *observer*. Here we encounter implicit personality theories and stereotypes that prescribe various expectations or predictions about another person (e.g., Hastorf et al., 1970). These allow us to "fill in the holes" left by an inadequate opportunity to observe the target and thereby to embellish an incomplete memory representation. The observer's self may also be involved in the rating of another person. The possibilities here range from making relativistic appraisals of the other person (e.g., "He's more aggressive than I am—hence, he's really aggressive") to the imposition or projection of self-referent characteristics onto the other person (e.g., Koltuv's 1962 projectivist hypothesis). The increased availability of behavioral exemplars for self-descriptive qualities (Markus, 1977) suggests that such qualities may be used more often in judging other people.

All of these sources of information no doubt combine in subtle and complex ways to produce the subjects' ratings, and indeed the domain of person perception addresses these kinds of issues (Tagiuri, 1969). Although the present data do not permit statements about how these sources may combine, they do clearly indicate that the memory trace for the rating event is very different from that created during a self-referent rating. The other-referent data show that increased time performing the rating—which could be due to either seeking out additional information or combining available inputs—enhances the probability of subsequent recall. This finding suggests a frequency or strength model, in which increased input makes increments in some internal memory store and thereby facilitates subsequent performance.

Familiarity and a Model for Other Processing

A third proposition about the processing of personal information comes from manipulating the familiarity of the target for the other-referent rating. Apparently, familiarity with the target significantly modifies the processes involved in the judgment. Memory performance is enhanced after the subject comes to know the target (Experiments 4 and 5). This enhancement suggests that familiarity with the target permits a more efficient cognitive system for representing the other person than was available in the case of an unfamiliar target. The present data do not permit statements about whether these ratings are based on attempts to "type" a person (Cantor & Mischel, 1977), implicit personality theories (Schneider, 1973), or some form of self-involvement in the processing of information about others (Lemon & Warren, 1974). Rather, they do tell us that there are substantial changes in the cognitive system as we come to know a person better.

These changes relate to the increasing information we acquire about another person with increased familiarity. At some point during the accumulation of information about another person, an overload state must be encountered. That is, there must be a time when we can no longer keep track of all of the actually observed information we have about that other person. In order to cope with this, either the cognitive structure must change or we will have to ignore new information. If the cognitive structure changes, it probably shows an increasing degree of abstraction in content. Rather than remembering a series of specific instances, a general statement representing the commonality of the other's behaviors may be the stored element. To be certain, specific instances can be generated from this general statement, but the stored element is the generalization about the other person's behavior. Fiske and Cox (in press) present evidence consistent with this proposed interpretation. When individuals are asked to describe an unknown other (stranger), they rely heavily on specific behavioral exemplars or scripts, rather than abstracted traits. However, when asked to describe a known other, subjects generate descriptive protocols that indicate a higher frequency of abstracted traits than of specific observed behavior. Furthermore, traits are generated much earlier in the protocols for known others, suggesting that they do indeed

serve a summarizing function. For strangers, traits begin to emerge in the protocol only after the opportunity for observation has been provided. Hence, increased exposure to the target enhances the probability of abstraction.

The critical question is, *What guides the abstraction process that allows us to summarize the available information about the other person?* Implicit personality theories, stereotypes, and the kinds of attributes outlined by Fiske and Cox (in press) are no doubt part of this process. However, it is possible that the self is also very active here. The self is a cognitive structure that has evolved to help summarize available self-related information. This means it serves the same function as the abstraction process involved with other people. It is an intriguing possibility that how we summarize information about other people is bound up with our own view of self. Several lines of research evidence are compatible with this suggestion.

Self-Involvement in Other-Processing

First, it appears that the same traits may be used in the perception of other people as are used in self-perception. For example, Lemon and Warren (1974) found that personally relevant traits are used more often and earlier in free descriptions of other people (see also Shrauger & Patterson, 1974). This suggests that the self may shape our views of others.

Second, the research involving self-based consensus implies that people surmise population performance from their own behavior (Heider, 1958). More recently, Ross (1977) has formulated the notion of a "false-consensus bias" in attributions, which suggests that the self serves as a reference point in interpreting information about others. Briefly, the false-consensus bias postulates that laymen see their *own* behavioral choices and judgments as relatively common and appropriate to existing circumstances, while viewing alternate responses as uncommon, deviant, and inappropriate (Ross, 1977). An important corollary of this hypothesis is that a person judges those responses that differ from his or her *own* to be more revealing of another's stable disposition than responses that are similar (Ross, 1977). To the extent that persons use their own views to interpret information about others, the self is probably involved in processing information about others.

A third line of research and theory that speaks to the issue of the self being involved in the ratings of other people emerges under the idea that the self functions like a fixed cognitive reference point. Snygg and Combs (1959) suggested that the self is *the* reference point for the processing of *all* information. With respect to the processing of information about others, they suggested that "people are not really fat, unless they are fatter than we are" (Snygg & Combs, 1959, p. 145). Comparisons of others to the self also form the keystone of Festinger's (1954) "social comparison" theory and Jones and Gerard's (1967) "comparative appraisal" theory of self-development. Lemon and Warren (1974) and Shrauger and Patterson (1974) offer empirical evidence consistent with this notion of an implicit and automatic self–other comparison being an integral part of how we process information about other people.

These three lines of research are compatible with the notion that the self, as a cognitive structure, is involved in the processing of information about other people. This idea does not imply that the content of a cognitive structure for another person is the same as the content for the self. Rather, it suggests that the self is involved in the development of the cognitive representation of another person. The manner in which the self helps shape this development and the kinds of biases that may emerge are fodder for a considerable amount of future research. For the present, it appears that there may be a complex involvement of the self in the processing of information about other people.

In summary, the present data do verify that self- and other-referent tasks are performed using different processes when the target other is unfamiliar. There is also an indication that the cognitive structure associated with another person undergoes some substantive changes as familiarity increases.

However, these findings relate to only one kind of decision about other people and, as such, beg for further research to explore their generality. We hope that this research will be hardheaded experimental work that begins to shed light on the developmental sequence of the cognitive structures involved in processing information about other people.

References

Bartlett, F. C. *Remembering*. Cambridge, England: Cambridge University Press, 1932.

Beach, L., & Wertheimer, M. A free-response approach to the study of person cognition. *Journal of Abnormal and Social Psychology*, 1961, *62*, 367–374.

Bower, G. H., & Karlin, M. B. Depth of processing pictures of faces and recognition memory. *Journal of Experimental Psychology*, 1974, *103*, 751–757.

Bransford, J. D., & Franks, J. J. The abstraction of linguistic ideas. *Cognitive Psychology*, 1971, *2*, 331–350.

Campbell, D. T., & Stanley, J. C. *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally, 1967.

Cantor, N., & Mischel, W. Traits as prototypes: Effects on recognition memory. *Journal of Personality and Social Psychology*, 1977, *35*, 38–48.

Craik, F. I. M., & Tulving, E. Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 1975, *104*, 268–294.

Dalrymple-Alford, E. C. Measurement of clustering in free recall. *Psychological Bulletin*, 1970, *74*, 32–34.

Festinger, L. A theory of social comparison processes. *Human Relations*, 1954, *7*, 117–140.

Fiske, S. T., & Cox, M. G. Describing others: Person impressions as person concepts. *Journal of Personality*, in press.

Hastorf, A. H., Schneider, D. J., & Polefka, J. *Person perception*. Don Mills, Canada: Addison-Wesley, 1970.

Heider, F. *The psychology of interpersonal relations*. New York: Wiley, 1958.

Jackson, D. N. *A manual for the Personality Research Form*. Goshen, N.Y.: Research Psychologists Press, 1967.

Jones, E. E., & Gerard, H. B. *Foundations of social psychology*. New York: Wiley, 1967.

Klatzky, R. L. *Human memory: Structures and processes*. San Francisco: Freeman, 1975.

Koltuv, B. Some characteristics of intrajudge trait intercorrelations. *Psychological Monographs*, 1962, *76*(33, Whole No. 552).

Lemon, N., & Warren, N. Salience, centrality, and self-relevance of traits in construing others. *British Journal of Social and Clinical Psychology*, 1974, *13*, 119–124.

Markus, H. Self-schemata and processing information about the self. *Journal of Personality and Social Psychology*, 1977, *35*, 63–78.

Nisbett, R. E., Caputo, C., Legant, P., & Marecek, J. Behavior as seen by the actor and as seen by the observer. *Journal of Personality and Social Psychology*, 1973, *27*, 154–164.

Norman, W. T., & Goldberg, L. R. Raters, ratees, and randomness in personality structure. *Journal of Personality and Social Psychology*, 1966, *4*, 681–691.

Posner, M. I., & Keele, S. W. Retention of abstract ideas. *Journal of Experimental Psychology*, 1970, *83*, 304–308.

Postman, L. Short-term memory and incidental learning. In A. W. Melton (Ed.), *Categories of human learning*. New York: Academic Press, 1964.

Price, R. H. The taxonomic classification of behaviors and situations and the problem of behavior–environment congruence. *Human Relations*, 1974, *27*, 567–585.

Reed, S. K. Pattern recognition and categorization. *Cognitive Psychology*, 1972, *3*, 382–407.

Rogers, T. B. An analysis of two central stages underlying responding to personality items: The self-referent decision and response selection. *Journal of Research in Personality*, 1974, *8*, 128–138.

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology*, 1977, *35*, 677–688.

Rogers, T. B., Rogers, P. J., & Kuiper, N. A. Evidence for the self as a cognitive prototype: The "false alarms effect." *Bulletin of Personality and Social Psychology*, in press.

Ross, L. The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10). New York: Academic Press, 1977.

Schneider, D. J. Implicit personality theory: A review. *Psychological Bulletin*, 1973, *79*, 294–309.

Secord, P. F., & Backman, C. W. *Social psychology*. New York: McGraw-Hill, 1974.

Shrauger, J. S., & Patterson, M. B. Self-evaluation and the selection of dimensions for evaluating others. *Journal of Personality*, 1974, *42*, 569–585.

Snygg, D., & Combs, A. W. *Individual behavior: A perceptual approach to behavior* (Rev. ed.). New York: Harper, 1959.

Tagiuri, R. Person perception. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (Vol. 3). Reading, Mass.: Addison-Wesley, 1969.

Wiggins, J. S. *Personality and prediction*. Reading, Mass.: Addison-Wesley, 1973.

Received March 13, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 4, 515-521

Making Trust Easier and Harder Through Two Forms of Sequential Interaction

Philip Brickman, Lawrence J. Becker, and Sidney Castle
Northwestern University

If people in a trust dilemma take turns choosing first and second (alternating interaction), there is no longer a dilemma for the people choosing second. Their motives are completely revealed by their choices. In most real-world situations, however, although choices are sequential, each choice is both a response to another person's past move and a stimulus for the other person's next move (continuing interaction). It was hypothesized that cooperation would be harder to achieve in continuing interaction than in alternating interaction, since in continuing interaction, like simultaneous interaction, people's motives are ambiguous. Since males have been shown to be more concerned with communication in trust dilemmas, it was also hypothesized that alternating interaction would benefit males more than females. Data from two studies using a total of 42 male and 42 female dyads supported this reasoning. Thus, we can relax one of the most artificial constraints of the formal trust dilemma—simultaneous choice—and still retain its important psychological properties.

It is enormously important to understand when people will trust each other and when they will not. Experimental social psychology has invested great energy in studying this problem through what are called trust dilemmas or Prisoner's Dilemmas (see Wrightsman, O'Connor, & Baker, 1972, for an extensive review of this literature). In a trust dilemma each person is much better off competing if the other person cooperates (because one exploits the other person's generosity) and is also better off competing if the other person competes (because one defends oneself against exploitation by the other person). Thus, regardless of what the other person does, competing is always a more profitable choice. However—and this is the dilemma—both parties will be worse off if they each compete than they would be if they could both trust each other and cooperate.

The question is increasingly raised, however, as to whether people decide to trust someone else in laboratory experiments on the same basis that they do outside the laboratory (see Wrightsman, O'Connor, & Baker, 1972, also for a discussion of these issues). The present studies address themselves to one possible issue in this debate. In laboratory experiments on trust dilemmas, both parties to the dilemma are usually asked to decide whether to trust each other simultaneously and in ignorance of each other's choice. Mutual simultaneous choice is necessary for a situation to be considered a formal trust dilemma (Rapoport & Chammah, 1965a). However, in social interaction situations outside the laboratory, people rarely make choices that are precisely simultaneous and continue to be so over an extended period of time. Social interaction ordinarily consists of a sequence of events in which each party's action in turn stimulates subsequent action by the other party. If making choices in sequence removes the logical and psychological dilemma in trusting, the dynamic of the trust

The authors wish to thank Darwyn E. Linder for the helpful comments he made in a review of an earlier draft of this article.

Lawrence Becker is now at the Department of Psychology, Princeton University.

Requests for reprints should be sent to Philip Brickman, who is now at the Research Center for Group Dynamics, Institute for Social Research, University of Michigan, Ann Arbor, Michigan 48106.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3704-515$00.75

dilemma is applicable only to a very narrow and special set of circumstances.

We wish to demonstrate that there are two forms of sequential interaction. One of them removes the dilemma, to the extent that the dilemma exists because there is ambiguity in deciding why someone has chosen not to cooperate. The other form, however, continues to make it difficult to decide whether people are interested in exploitation or only in protecting themselves and is also a much more reasonable analogue to ordinary social interaction than either of the other two conditions we will be considering.

Ambiguity of Simultaneous Choice

Although choosing to compete is always more profitable in a trust dilemma than choosing to cooperate, the motivation for choosing to compete may be quite different depending upon whether the other party is expected to cooperate (in which case competing is a matter of exploitation) or expected to compete (in which case competing is a matter of self-defense). If one knows only that someone has chosen to compete, it is impossible to know whether his or her motivation was exploitation or self-defense. In a situation where both people are choosing simultaneously, it will be impossible to know with certainty whether a person is competing to gain advantage or to protect him- or herself. Kelley and Stahelski (1970b) have mapped out some of the problems that occur because people make different assumptions about what others will do, and McClintock (1972) has developed a sophisticated analysis of the different social motives potentially operative in trust dilemmas.

Clarity of Alternating First and Second Choices

The ambiguity involved in choosing to compete is fully removed if the person competing has already been informed whether the other party is competing or cooperating. If the other party is cooperating, the only reason to compete would be a desire to exploit. If the other party is competing, a desire to defend oneself is sufficient motivation to compete. Conversely, if the other party is competing, the only reason for cooperating in the second position would be a desire to create trust. There is no dilemma for a person choosing second in a sequential choice situation, and there is little difficulty in inferring the motives of a person in this position (cf. Kelley & Stahelski, 1970a). If people choose sequentially rather than simultaneously and take turns choosing first and second, each person should have a variety of occasions on which to signal the other person without ambiguity as to whether he or she is interested in exploitation, or only in self-defense, or actually in creating trust. If people compete in trust dilemmas in part merely because they cannot send or receive unambiguous messages of cooperative intent (cf. Deutsch, 1958, 1960) or because ambiguity serves as a smoke screen that makes it easier to get away with being exploitative, an interaction in which they alternate choosing first and choosing second should produce a greater degree of cooperation than a situation in which both parties choose simultaneously.

Direct tests of this hypothesis so far have yielded ambiguous results. Swinth (1967) found a trend toward more cooperation following an alternating-choice situation, but Wrightsman, Bruininks, Lucker, & O'Connor (1972) found no effect. Deutsch (1958, 1960) found more cooperation in simultaneous play, but he did not have parties alternate choosing first and choosing second. Oskamp (1974) also did not have parties alternate first choices. He, too, found more cooperation in simultaneous play, but only when the payoff matrix had a positive average expected payoff. When the average expected payoff was zero, there was more cooperation in sequential play.

Ambiguity of Continuing Interaction

Knowing whether the other person has just decided to compete or cooperate, however, may not remove the dilemma from one's decision of whether in turn to compete or cooperate. In ordinary social intercourse, every action is both a response to a previous action by the other party and a stimulus to a subsequent action by the other party. As a re-

sponse, the action of competing or cooperating may be taken as a second choice in a trust dilemma, fixing at that moment how much reward each party gets as a function of their joint current orientations toward competing and cooperating. As a stimulus, the action of competing or cooperating may be taken as a first choice in a trust dilemma, setting the stage for the other person to determine by his or her subsequent choice how much reward each party gets as a function of their current orientations.

In what may be called continuing interaction, since each move is both a first move (a stimulus) and a second move (a response), there is no longer a pure second move, and one can no longer infer unambiguously whether a person competes out of a bold desire to exploit or a cautious desire to protect him- or herself. Even if the other person has just chosen cooperatively, the subject cannot be sure that this is not a device to induce a cooperative choice that will subsequently be exploited to the net advantage of the other person. A continuing interaction of this sort conveys more information about participants' motives than simultaneous choosing, but less than an alternating situation in which people's moves count only as either first moves or second moves, not both. We would thus expect that the problem of whether to trust someone, as determined by the precision with which that person's motivation to cooperate can be estimated, would be easiest with alternating interaction, harder with continuing interaction, and hardest with simultaneous interaction. The present research is designed to test this analysis.

Sex Differences

Since sex differences have been an almost universal finding for trust dilemmas from the time they were explored by Rapoport and Chammah (1965b), they may be expected in the present research as well. Most past research has found that females are less likely to cooperate in trust dilemmas than are males (cf. Hottes & Kahn, 1974). The reasons for this finding are still somewhat obscure, but a number of provocative analyses (Bedell & Sistrunk, 1973; Conrath, 1972; Hottes & Kahn, 1974; Johnston, Markey, & Messé, 1973) have suggested that females are simply less aware that communicating cooperative intentions is a critical feature of the situation. If this is so, we would expect that the enhanced opportunity to communicate cooperative intentions in the continuing and alternating games would be of more benefit to males than to females.

The design for the present research was a 2×3 factorial using sex of subject and interaction structure (simultaneous, continuing, or alternating) as independent variables. Two studies were carried out. The purpose of Study 2 was to assess the reliability of the findings in Study 1, to remove a potential confound involving sex of experimenter, and to obtain explicit measures of subjects' attributions of intention for various choice sequences in each interaction condition. Since the procedures for the second study were the same as for the first study, except where noted, the two studies will be described together.

Method

Subjects and Experimenters

In Study 1, 20 male and 22 female undergraduates acted as experimenters. Same-sex students were paired off for a total of 10 male and 11 female experimenter teams. Each team of experimenters recruited and ran three pairs of undergraduate volunteers of the same sex as the team, one pair randomly assigned to each of the three experimental conditions. Thus 20 males and 22 females served in each of the three conditions.

In Study 2, six male and four female undergraduates acted as experimenters. Same-sex students were paired off for a total of two male and two female experimenter teams (one of the male teams had four members). Since one of the purposes of Study 2 was to avoid the confounding of subject and experimenter sex that was set up in Study 1, each team of experimenters in Study 2 recruited and ran six pairs of undergraduate volunteers, three male pairs and three female pairs. Each team randomly assigned one pair of each sex to each of three experimental conditions. Thus, eight males and eight females served in each of the three conditions. The order in which conditions were run by each experimenter team in both Study 1 and Study 2 was randomly determined for each team.

Procedure

The experiments were conducted in dormitories and other living units. Subjects were seated in sepa-

rate rooms, with care being taken to conceal each subject's identity from the other. The experimenters were also separated, each taking the responsibility for briefing one of the subjects.

Subjects were told that the study was one of how people made choices in a social situation. Their ability to handle some difficult social situations would be tested by having them play a special game. Evaluations would be based on the number of points each person earned in the game. The better a person handled such situations, the more points he or she would earn.

In the game, subjects were told, each player would have two possible moves, A or B, with the amount a person won or lost in each round depending both on the choice the person made and on the choice the other player made. To facilitate subjects' understanding of the game, they were handed a sheet of paper listing all the possible combinations of moves, along with their associated payoffs. The sheet showed that if both parties chose A, each would earn 10 points. If the subject chose A and the other person chose B, the other person would win 50 points and the subject would lose 50 points. If the subject chose B and the other person chose A, the subject would win 50 points and the other person would lose 50 points. If both parties chose B, each would lose 20 points. Subjects were quizzed to be sure that they understood these rules and were also told that they could keep their copy of the handout showing the various moves and payoffs and refer to it whenever they wished during the game.

At this point subjects were given a tally sheet on which columns were marked for them to record their move, the other person's move, their payoff, and the other person's payoff for each round. The manipulation of the interaction condition was then introduced, with players being instructed how to fill out their tally sheet on each round. To prevent end game effects, subjects were told only that they would play "a fair number" of rounds; they were actually allowed to play 18 rounds. Following play, subjects in Experiment 2 were asked to rate the extent to which they would infer either a desire to exploit or a desire to create trust from various moves made by the other player after their own previous move had been either competitive or cooperative. These ratings were made on scales ranging from 1 for "not at all" to 7 for "very much." Finally, subjects were brought together and allowed to talk about their experiences and to ask questions of the experimenters. The experimenters explained the study, gave subjects a feedback sheet that they could take with them to review at their leisure, and thanked the subjects for their participation.

Manipulation of Interaction Structure

In the simultaneous-choice condition, subjects were asked to write down their move for each round before being informed of the other player's move. After both subjects had made their choices, the experimenters informed each subject of what the other's choice had been.

In the alternating-choice condition, subjects were told that one would choose first and have his or her choice announced to the other, who would then in turn make a choice. Subjects were also told that they would take turns being in the position of moving first and that the person who was to start the game would be chosen by lot. It was suggested that subjects keep track of which player had chosen first on each round by circling on their tally sheet the choice made by the player who chose first.

In the continuing-choice condition, as in the alternating-choice condition, subjects were told that choices would be made in sequence and that it would be determined by lot who would start the game. It was then explained that each choice they made would always count double (except for the person starting the game)—once as the last move in one round and once as the first move in the next round. If the subject were moving last in, say, Round 7, he or she would write down that move once in the appropriate column for Round 7 and then once again directly below this, in the same column, as the move for Round 8. Similarly, when the other player's move for Round 8 was announced, the subject would write it down once as the last move for Round 8 and once again as the first move for Round 9. Again, subjects were asked to keep track of who had chosen first on each round by circling the choice of the player who had gone first. Subjects in all cases were quizzed to be sure that they understood the tally sheet and the sequence of moves before the game itself was started.

Results

Results of Study 1 were analyzed by means of 2×3 analyses of variance using sex of subject and interaction condition as independent variables. Sex of experimenter was added as a factor in the analysis of the choice behavior measures in Study 2. No significant main effect or interactions involving sex of experimenter were found.

The proportion of cooperative choices in a dyad in Study 1 and in Study 2 is shown in Table 1. In both studies there was a significant interaction between sex of subject and interaction condition: $F(2, 57) = 4.24$, $p < .019$, for Study 1, and $F(2, 12) = 4.00$, $p < .047$, for Study 2. In each case an analysis of simple main effects showed that interaction condition had a significant effect only for males: $F(2, 57) = 7.37$, $p < .005$, for Study 1, and $F(2, 12) = 8.43$, $p < .01$, for Study 2. In both studies a Newman-Keuls analysis indicated that males in the alternating condition had a significantly ($\alpha = .05$) higher proportion of cooperative choices in their dyads

than males in the simultaneous and continuing conditions and that between those latter two conditions there was no significant difference.

In Study 1 the proportion of cooperative moves in a dyad, given a cooperative prior move by the other party, was significantly greater for males, $F(1, 56) = 7.00$, $p < .011$. Although the interaction effect was not statistically significant, males in the alternating condition responded to a prior cooperative move by the other party with a much higher proportion of cooperative choices than males in the other two interaction conditions (see Table 2). The significant sex-of-subject effect on this measure was not replicated in Study 2. In neither study were there any significant treatment effects on the proportion of cooperative choices in a dyad given a competitive prior move by the other party.

Using the arc sine transformation suggested by Myers (1972) for proportions did not essentially change any of the results of the analyses of variance reported above.

Table 1
Means and Standard Deviations of the Proportion of Cooperative Choices in a Dyad

Sex of subject	Interaction condition: Simultaneous	Continuing	Alternating
Study 1[a]			
Male			
M	.34	.39	.69
SD	.21	.17	.28
Female			
M	.36	.29	.32
SD	.28	.14	.20
Study 2[b]			
Male			
M	.31	.18	.67
SD	.12	.13	.20
Female			
M	.49	.46	.48
SD	.24	.22	.13

[a] There were 10 male pairs and 11 female pairs in each interaction condition.
[b] To facilitate comparisons with Study 1, the data for male and female experimenters were combined. There were 4 male and 4 female pairs in each interaction condition.

Table 2
Means and Standard Deviations of the Proportion of Cooperative Moves in a Dyad Given a Cooperative Prior Move by the Other Party, Study 1

Sex of subject	Interaction condition: Simultaneous	Continuing	Alternating
Male			
M	.33	.47	.70
SD	.26	.31	.29
n	10	10	10
Female			
M	.32	.25	.34
SD	.33	.26	.27
n	10	11	11

The data collected in Study 2 on subjects' attributions of either a desire to exploit or a desire to create trust in hypothetical choice sequences are presented in Table 3. As can be seen, when a cooperative choice of their own is followed by a competitive choice by their partner, subjects are much readier to say that their partners want to exploit them in the alternating and continuing conditions than in the simultaneous condition, and especially in the alternating condition, $F(2, 42) = 6.99$, $p < .01$. When a competitive move of their own is followed by a cooperative move by their partner, subjects are much more willing to attribute a desire to create trust to their partner in the alternating condition than in either the simultaneous or the continuing conditions, $F(2, 42) = 3.34$, $p < .05$. When a cooperative move of their own is followed by a cooperative choice by their partner, subjects are also somewhat readier to say that their partners want to create trust in the alternating and continuing conditions than in the simultaneous condition, especially in the alternating condition, $F(2, 42) = 3.18$, $p < .06$. Finally, competition that follows competition is not seen as particularly indicative of a desire to exploit in any of the three conditions. There were no sex differences on these variables and no Sex of Subject × Interaction Condition interactions.

Table 3
Mean Attributions of Intentions to Other Person as a Function of Choice Sequence and Interaction Condition

Choice sequence		Interaction condition		
Own previous choice	Other's next choice	Simultaneous	Continuing	Alternating
Attributed desire to exploit as motive for competing				
Cooperate	Compete	3.00	4.56	5.06
Compete	Compete	2.12	2.50	2.31
Attributed desire to create trust as motive for cooperating				
Compete	Cooperate	3.19	3.63	5.06
Cooperate	Cooperate	4.12	5.12	5.56

Note. $n = 8$ for each interaction condition. All scales range from 1 for "not at all" to 7 for "very much."

Discussion

There are three main conclusions suggested by this research. The first is that trust dilemmas appear to generalize to continuing interactions, in which parties choose in turn but each party's choice counts as both a response to the partner's previous move and a stimulus for the partner's next move. The rational dilemma of whether to trust somebody and the attributional dilemma of how to interpret that person's intentions hold in modified form for continuing interaction as well as for simultaneous interaction, in which both parties choose at the same time in ignorance of each other's current choice. This result has important implications for the generalizability of the findings of laboratory studies of trust dilemmas and for the conduct of future laboratory studies. These studies have typically involved simultaneous interaction. Outside the laboratory, however, interactions are much more likely to be continuing than simultaneous. (Consider, for example, such phenomena as arms races, contract negotiations, courtship, seduction, and contract bridge.) By showing that trust dilemmas are relevant to continuing interactions as well as to simultaneous ones, the present study encourages us to believe that the dynamics of trust dilemmas underlie problems of suspicion and distrust outside the laboratory as well as inside it. Furthermore, by indicating that we can relax one of the most demanding constraints of the formal trust dilemma—simultaneous choice—and still retain many of the dilemma's important psychological properties, the present research opens the door to the use in future laboratory studies of better experimental analogues, namely those involving continuing interaction.

The second conclusion is that at least for males, trust is hard to achieve in part simply because it is hard to make unambiguous inferences about other people's motivations in the ordinary simultaneous-choice trust dilemma. Even if the subject has chosen cooperatively on his or her own previous move, a subsequent competitive move by the other party is not seen as indicating a clear desire to exploit in a simultaneous game (see Table 3). Even if the subject has chosen competitively on his or her own previous move, a subsequent cooperative move by the other party is not seen as indicating a clear desire to establish trust in a simultaneous game. On the other hand, both these inferences are clearly made for second moves in an alternating game, in which the first move is known in advance by the party making the second move. This knowledge proved sufficient to establish a much higher degree of cooperation among males in the alternating game than in the simultaneous game. Male subjects were apparently willing and able to signal their interest in cooperation in an alternating game and to respond cooperatively to such signals given by their partners. Where a disposition to cooperate exists, the present data indicate that we may take a long step toward resolving trust dilemmas merely by altering the structure of choice in a way that makes each party's motivation clear to the other party.

It should be noted that alternating interaction is a clear advantage for cooperation only when the question of inferring another's motives is paramount. This is not the case, for example, in a situation in which two parties unknowingly determine each other's fate, called the "minimal social situation." The main questions for the actors in the

minimal social situation are discovering which choice the other person prefers and communicating to the other person which choice they prefer. Kelley, Thibaut, Radloff, and Mundy (1962) have demonstrated that under these circumstances, a cooperative solution is more likely if the parties choose simultaneously rather than in alternation.

The third conclusion is that the basis for failure to achieve trust in a trust dilemma is to some extent different for males and females. Like males, females made more unequivocal inferences of motivation in an alternating interaction than in a simultaneous one. Unlike males, however, females were not thereby more cooperative under alternating than under simultaneous conditions. It would appear that the testing and communicating of cooperative intentions is a critical factor for males in resolving trust dilemmas, but not for females. This sex difference may account for Wrightsman, Bruininks, Lucker, & O'Connor's (1972) not finding increased cooperation in the alternating condition. They had twice as many female as male subjects and did not include sex of subject as a factor in the design. The puzzle of female distrust is all the greater because in other situations, females are more cooperative and more compassionate than males (Uesugi & Vinacke, 1963; Vinacke, 1959). We suspect that females are not defining compassion as relevant to the standard trust dilemma and that if they were induced to do so, this would be the key to elevating their tendency to cooperate.

References

Bedell, J., & Sistrunk, F. Power, opportunity costs, and sex in a mixed-motive game. *Journal of Personality and Social Psychology*, 1973, *25*, 219–226.

Conrath, D. W. Sex role and "cooperation" in the game of Chicken. *Journal of Conflict Resolution*, 1972, *16*, 433–443.

Deutsch, M. Trust and suspicion. *Journal of Conflict Resolution*, 1958, *2*, 265–279.

Deutsch, M. The effect of motivational orientation upon trust and suspicion. *Human Relations*, 1960, *13*, 123–139.

Hottes, J. H., & Kahn, A. Sex differences in a mixed-motive conflict situation. *Journal of Personality*, 1974, *42*, 260–275.

Johnston, J., Markey, C., & Messé, L. A. Sex difference in labeling effects on behavior in the Prisoner's Dilemma game. *Proceedings of the 81st Annual Convention of the American Psychological Association*, 1973, *8*, 323–324. (Summary)

Kelley, H. H., & Stahelski, A. J. The inference of intentions from moves in the Prisoner's Dilemma game. *Journal of Experimental Social Psychology*, 1970, *6*, 401–419. (a)

Kelley, H. H., & Stahelski, A. J. Social interaction basis of cooperators' and competitors' beliefs about others. *Journal of Personality and Social Psychology*, 1970, *16*, 66–91. (b)

Kelley, H. H., Thibaut, J. W., Radloff, R., & Mundy, D. The development of cooperation in the "minimal social situation." *Psychological Monographs*, 1962, *76*(1, Whole No. 538).

McClintock, C. G. Game behavior and social motivation in interpersonal settings. In C. G. McClintock (Ed.), *Experimental social psychology*. New York: Holt, Rinehart & Winston, 1972.

Myers, J. L. *Fundamentals of experimental design*. Boston: Allyn & Bacon, 1972.

Oskamp, S. Comparison of sequential and simultaneous responding, matrix, and strategy variables in a Prisoner's Dilemma game. *Journal of Conflict Resolution*, 1974, *18*, 107–116.

Rapoport, A., & Chammah, A. M. *Prisoner's Dilemma*. Ann Arbor: University of Michigan Press, 1965. (a)

Rapoport, A., & Chammah, A. M. Sex differences in factors contributing to the level of cooperation in the Prisoner's Dilemma game. *Journal of Personality and Social Psychology*, 1965, *2*, 831–838. (b)

Swinth, R. L. The establishment of the trust relationship. *Journal of Conflict Resolution*, 1967, *11*, 335–344.

Uesugi, T. K., & Vinacke, W. E. Strategy in a feminine game. *Sociometry*, 1963, *26*, 75–88.

Vinacke, W. E. Sex roles in a three-person game. *Sociometry*, 1959, *22*, 343–360.

Wrightsman, L. S., Jr., Bruininks, R. H., Lucker, W. G., & O'Connor, J. Effects of extensiveness of instructions upon cooperation in a Prisoner's Dilemma game. In L. S. Wrightsman, Jr., J. O'Connor, & N. J. Baker (Eds.), *Cooperation and competition: Readings on mixed-motive games*. Belmont, Calif.: Brooks/Cole, 1972.

Wrightsman, L. S., Jr., O'Connor, J., & Baker, N. J. (Eds.). *Cooperation and competition: Readings on mixed-motive games*. Belmont, Calif.: Brooks/Cole, 1972.

Received June 22, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 4, 522–533

Probability of Acceptance in Dating Choice

James Shanteau and Geraldine F. Nagy
Kansas State University

Three experiments examined the influence of probability of acceptance and physical attractiveness on dating choice. In Experiment 1, 15 females made preferential choices between two male dates, each described by a photograph and a statement of the probability of acceptance. Both probability and attractiveness of the date were found to be important in the subjects' judgments. Further, a multiplying rule for the combination of probability of acceptance and attractiveness was supported, in agreement with the subjective expected utility theory of risky decision making. In Experiment 2, the same females made preferential choices between pairs of dates described by photographs alone. The preferences for 11 of the 15 subjects could only be accounted for by using both probability and attractiveness. A multiplicative combination of probability and attractiveness was also found to be predictive of dating choices made by these subjects. Experiment 3 replicated the results of Experiment 2 with a new group of 14 females. This research provides evidence that (a) females are influenced by both probability of acceptance and physical attractiveness in dating choice and (b) they tend to combine probability with attractiveness in a multiplicative fashion.

Two factors have frequently been hypothesized to play an important role in dating choice (e.g., Walster, Aronson, Abrahams, & Rottman, 1966). One factor is the attractiveness of a prospective date, and the other is the perceived probability of being accepted by the date. The importance of attractiveness, particularly the physical appearance of the date, has been repeatedly demonstrated (Byrne, Ervin, & Lamberth, 1970; Kleck & Rubenstein, 1975; Walster et al., 1966). However, there has been no consistent evidence about the importance of probability of acceptance. The primary purpose of this research is to determine what influence probability has on dating choice.

Previous attempts to evaluate probability of acceptance have generally compared the dating choices of attractive and unattractive subjects. Unattractive subjects are assumed to have a lower expectancy of being accepted by any given date. Therefore, these subjects would presumably choose less attractive dates than would more attractive subjects. This should be particularly true when the saliency of rejection is emphasized.

The results using this between-subjects approach have not been conclusive. Berscheid, Dion, Walster, and Walster (1971) found that unattractive subjects did choose slightly less attractive dates than did attractive subjects. However, they failed to find any differential effects when rejection was made more salient. In comparison, Huston (1973) found no difference in the dates chosen by attractive and unattractive subjects. However, he did find that all subjects chose more attractive dates when acceptance was assured. Other studies have generally led to equally inconclusive results.

This research was supported in part by Public Health Service Grants MH 22610 and MH 26002 from the National Institute of Mental Health. The authors wish to thank Sue Ellingboe for her assistance in running pilot subjects and in analyzing data from the studies reported here. We also wish to thank Norman Anderson, Michael Birnbaum, John Carroll, and John Payne for their helpful comments on various stages of the research.

Requests for reprints should be sent to James Shanteau, Department of Psychology, Kansas State University, Manhattan, Kansas 66506.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3704-522$00.75

In contrast to previous efforts, the present research was designed to study the influence of probability of acceptance for individual subjects. That is, rather than studying the effects of probability by a comparison between groups of subjects, analyses were made by using a within-subjects approach. The research strategy is thus based on analyzing the impact that probability has on the cognitive processes used by individuals in making dating choices.

A second but related purpose of this research was to investigate how probability is used in dating choice. That is, assuming probability is important, how is it combined with other information such as physical attractiveness? Several approaches suggest that attractiveness and probability should combine by multiplying. For instance, level of aspiration theory predicts that the desirability of a goal should be multiplied by the probability of attaining that goal (Lewin, Dembo, Festinger, & Sears, 1944). According to this view, dating choice should depend on the product of the desirability (or attractiveness) of the date and the probability of acceptance (Walster et al., 1966).

From a rather different viewpoint, subjective expected utility theory states that choices between risky alternatives (such as gambles) should depend on the product of the payoff times the probability of attaining the payoff (Edwards, 1961). Support for this multiplying rule of risky decision making has been found in several types of gambling situations (Shanteau, 1974, 1975). However, the rule has not been evaluated in more complex social judgment settings such as dating choice.

Formally, the multiplying rule can be written as

$$R = P \times PA, \qquad (1)$$

where the desirability of a date (R) equals the product of the probability of acceptance (P) and the physical attractiveness (PA) of the date. Statistically, Equation 1 predicts a significant interaction between probability and attractiveness that should be concentrated in the bilinear (Linear × Linear) trend component (Anderson & Shanteau, 1970; Shanteau, 1977).

Thus, the present studies were designed to determine whether probability is important in dating and, if so, to evaluate how probability is combined with attractiveness to arrive at a dating choice. In the first of three experiments, probability of acceptance was explicitly provided along with a photograph of the prospective date. This procedure allowed a direct assessment of the impact of probability and attractiveness on dating choice. In the second and third experiments, only photographs were provided, with the probability unspecified. Since probability was left up to each subject to infer, these studies were intended to mirror more closely actual dating situations.

Experiment 1

In Experiment 1, female subjects were asked to choose between two male dates. Each date was described by a photograph and a verbal statement of the probability that he would agree to go out with the subject. Both the attractiveness of the photographs and the probability of acceptance were systematically varied in a factorial design. This design permitted a straightforward analysis of whether probability influenced the subjects' judgments. It also allowed a direct assessment of how probability is combined with physical attractiveness in making a dating choice.

Method

Subjects. Fifteen female undergraduates were each paid $1.50 per hour to participate in the research. The females were all unmarried, free to date, and between 17 and 22 years old. All subjects were run individually.

Stimuli. Subjects made preferential choices between two dates such as "Tom/Fairly likely" and "Joe/Unlikely." The names referred to photographs on a nearby display board; the verbal probability phrase represented the likelihood that the pictured male would be willing to go out with the subject on a date.

The stimulus information was presented on index cards with one date on the left side of the card and the other on the right. There were 140 data pairs from a 7 (left probability) × 5 (left photo) × 2 (right probability) × 2 (right photo) factorial design. For the left date, the seven probability phrases were Sure thing, Highly likely, Fairly likely,

Toss-up, Somewhat unlikely, Very unlikely, and No chance. These were combined with five photographs that varied from quite attractive to unattractive. For the right date, the probabilities were either Likely or Unlikely, and either an attractive or an unattractive photograph was used. The left and right dates were counterbalanced so that each photograph and probability appeared equally often on each side of the card.

The photographs were candid yearbook shots and were selected on the basis of preliminary ratings to cover a wide range of physical attractiveness. The names identifying the photographs were the seven most common male names (John, Bob, Bill, Jim, Tom, Joe, and Dick) listed by Battig and Montague (1969, p. 35); the names were randomly assigned to the photographs. The verbal probability phrases were selected on the basis of normative ratings (Lichtenstein & Newman, 1967) to cover a wide range; these same phrases have also been used in prior research (Shanteau, 1974, 1975).

Procedure. The subject's task was to compare the two dates on a stimulus card and to "make a choice as to which you would prefer." Although the subject was aware that she would not actually go out on a date, an attempt was made to make the task meaningful. For instance, photographs were taken of all subjects comparable to the pictures of male dates. The subjects were to assume that the probability phrase represented the male's reaction after seeing her photograph.

The subject indicated her date preference by locating a pointer on a 40-cm unmarked response bar. The left end of the bar was defined as a "very strong preference for the date on the left," and the right end was defined as a "very strong preference for the date on the right." The middle of the scale represented "no preference between the two dates." The ends of the response bar were also exemplified by anchor stimuli that were more extreme than any of the experimental stimuli. These anchors were used to decrease the possibility of floor or ceiling effects.

After presenting each dating pair, the experimenter recorded the subject's response to the nearest millimeter, using a ruler on the rear of the scale. The scale ranged from +100 on the left to −100 on the right.

The experiment consisted of three daily 1-hour sessions. The initial part of the first session was devoted to instructions and practice. Following the practice, the subject was required to correctly summarize the instructions back to the experimenter. The entire set of experimental stimuli was then presented. Sessions 2 and 3 began with a brief summary of the instructions. In Session 2, the complete set of stimuli was presented for a second time; two more replications were then administered in Session 3. Prior to each replication, the stimulus cards were shuffled to provide a different random presentation order.

Results

The group mean preference responses for the left dates averaged over the right dates are plotted in Figure 1. For example, the top right point corresponds to the mean preference for the photograph of Tom when it was paired with a Sure thing probability; the mean response in this case was roughly +65.

The names have been spaced along the horizontal axis according to their marginal means and reflect the attractiveness of the corresponding photos: the farther to the right, the more attractive the photograph. Thus, Tom and Bob were seen as highly attractive, John and Jim were of intermediate attractiveness, and Bill viewed as fairly unattractive.

Each curve in Figure 1 corresponds to the listed probability. As can be seen, preference for dates increases as the probability becomes more certain. For instance, "Tom/Sure thing" has a much higher preference than "Tom/No chance."

The results in Figure 1 not only suggest that probability of acceptance is important but also bear on the combination rule used to integrate probability with attractiveness. The multiplying model in Equation 1 predicts that the seven curves in Figure 1 should form a fan of straight diverging lines (Anderson & Shanteau, 1970). As can be seen, the lines are very nearly linear and diverge in the predicted fashion. Similar-appearing results were obtained for all 15 females when individual-subject data were plotted.

Statistical analyses. The results for each subject were analyzed by separate analyses of variance. For all subjects, the probability and photograph main effects were both highly significant. In addition, all 15 subjects revealed a significant Photograph × Probability interaction as shown in the second column of Table 1. While these results are expected from Equation 1, the model makes precise predictions about the form of the interaction. In particular, the Photograph × Probability interaction should be concentrated in the bilinear trend component. This means that after the bilinear term is removed, the residual component should be nonsignificant. Using the procedure described in Shanteau

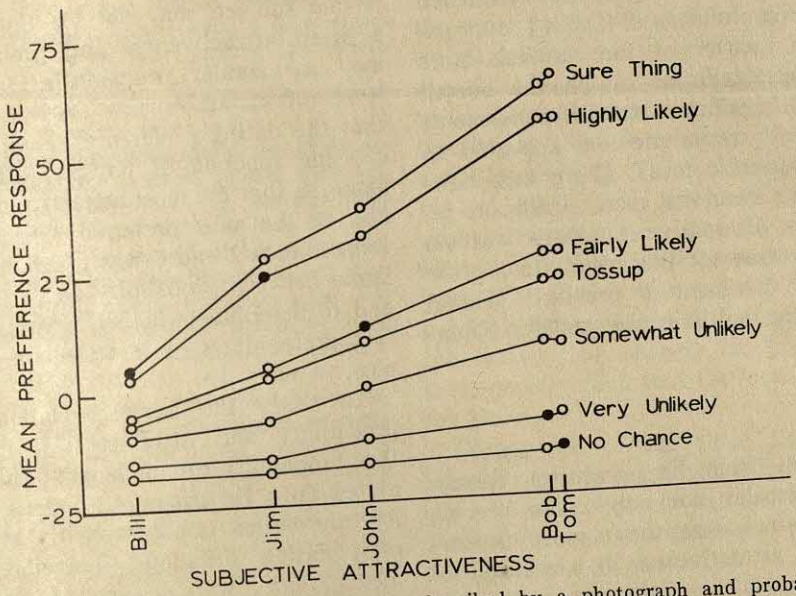


Figure 1. Mean preference responses for dates described by a photograph and probability of acceptance, Experiment 1. (Each point is averaged over four alternative dates. Photographs are spaced along the horizontal axis according to their marginal means; probabilities are listed as curve parameters. Filled points are discussed in the text.)

(1977), 11 of the 15 subjects were found to have nonsignificant residual components, as listed in the third column of Table 1. While the remaining 4 subjects showed significant residuals, the Fs are quite small, and plots of individual data revealed no systematic pattern of deviations from bilinearity.

Similar analyses were also performed sepa-

Table 1
F *Ratios for Interaction and Residual Terms Collapsed Over All Alternatives, and for Each Alternative Separately*

Subject No.	Photo × Probability interaction (df = 24, 420)	Overall residual (df = 23, 420)	Residual for alternative 1 (df = 23, 105)	Residual for alternative 2 (df = 23, 105)	Residual for alternative 3 (df = 23, 105)	Residual for alternative 4 (df = 23, 105)
1	4.79*	1.45	2.29*	1.02	.54	1.27
2	95.75*	.81	.89	.96	2.71*	1.18
3	23.84*	1.93	6.12*	.75	1.30	.93
4	20.45*	.93	3.09*	2.26*	1.38	.91
5	41.74*	1.83	.97	.90	1.27	1.68
6	29.09*	2.18*	1.69	1.40	.96	2.43*
7	8.28*	2.50*	.90	1.05	1.20	5.69*
8	16.45*	1.53	2.03*	1.98*	.62	1.67
9	15.58*	1.50	1.28	.93	.59	2.01*
10	19.10*	.94	.49	2.04*	.64	2.00*
11	40.25*	1.20	1.44	2.14*	2.01*	2.10*
12	36.93*	1.51	1.02	1.44	.54	1.05
13	17.50*	2.00*	2.64*	.66	1.07	1.96*
14	24.92*	3.34*	3.47*	1.37	2.43*	1.00
15	29.11*	1.63	.81	1.37	1.70	2.57*

* $p < .05$.

rately for each choice alternative. As listed in the last four columns of Table 1, approximately one quarter of the residuals were found to be significant. As in the overall analysis, the significant residual components were of small magnitude and appeared to have no systematic locus. There was, however, a slight trend for more significant residuals with Alternative 4; there was no identifiable reason for this result. In all, the results of Experiment 1 provided general support for the multiplicative model in Equation 1.

Discussion

The results from Experiment 1 clearly show that females not only make use of probability information, they appear to combine it with attractiveness in a multiplicative fashion. Thus, a date who is low on either factor (i.e., very unattractive or very unlikely to accept a date) will be seen as undesirable.

It is commonly believed that there is an inverse relation between attractiveness and probability (e.g., Walster et al., 1966). This means that highly attractive dates should have a low probability of acceptance, and unattractive dates should have a high probability. If the multiplying rule holds, then this inverse relation between attractiveness and probability implies that dates of intermediate physical attractiveness (and intermediate probability of acceptance) will be most preferred.

This effect is graphically illustrated in Figure 1. Five points have been filled based on an assumed negative relation between attractiveness and probability. For example, the least attractive date (Bill) was given the highest probability (Sure thing), whereas the most attractive date (Tom) was given the lowest probability (No chance). As can be seen, the filled points peak for dates of intermediate attractiveness. In this example, moderately attractive Jim with a Highly likely probability would be the most desirable.

Of course, this prediction depends on the presence of a negative relation between attractiveness and probability. One alternative is that subjects may not see any connection between attractiveness and probability (or may see similar probabilities for different levels of attractiveness). This would imply that the dating preferences would be an increasing function of physical attractiveness alone, so that the most attractive dates would also be the most preferred. One purpose of Experiments 2 and 3 was to examine the relation between probability and attractiveness and to determine whether subjects would in fact prefer dates of intermediate attractiveness.

To study the natural relation between probability and attractiveness, it is clear that probability cannot be artificially manipulated as in Experiment 1. Indeed, probability information is not explicitly available in most dating situations. Therefore, Experiments 2 and 3 were designed so that subjects would have to infer or evaluate probability based on other information (i.e., a photograph).

Experiments 2 and 3

In Experiments 2 and 3, subjects made preferential choices between pairs of dates described by photographs alone. This allowed a direct assessment of whether factors other than physical attractiveness influence dating choice. One additional purpose of Experiments 2 and 3 was to evaluate the multiplying rule for attractiveness and probability in a setting where probability was not explicitly provided.

Experiment 2 was run as a continuation of Experiment 1 using the same subjects. However, it is possible that the explicit presentation of probabilities in the first experiment may have alerted these subjects to the use of probability in the second experiment. Therefore, Experiment 3 was run on a new group of female subjects who had not gone through Experiment 1 initially.

Method

Subjects. In Experiment 2, the 15 females used in Experiment 1 returned for one additional session. In Experiment 3, 14 new subjects with similar characteristics were run under the same conditions.

Procedure. Subjects were initially presented with the seven photographs of males used in Experiment 1 and were asked to carefully consider the dating potential of each of the males. It was then explained that several facets of dating desirability would be examined in the experiment.

Subjects began by estimating the probability that each of the males would accept a date with her after viewing her photograph. In making these estimates, subjects were instructed to keep in mind that the male would be assured of her willingness to accept a date; this was done because several pilot subjects indicated that unattractive males might not choose them because the males would fear rejection.

Subjects responded by locating a pointer along a 40-cm unmarked bar. The left end of the bar was defined as "No chance" and the right end was defined as "Sure thing." The experimenter read the responses from the rear face of the scale using a 0–100 scale. The stimuli were presented four times in all, with a different random order each time.

Subjects were then asked to rate the physical attractiveness of each of the pictured males. The subject's task was to "disregard any other feelings you may have about the picture while making this rating—tell me only how physically attractive you consider this person to be." This instruction was provided because some pilot subjects stated that they were taking into account the possibility of reciprocal interest by the males. The basic experimental procedure was the same as for the probability ratings, except that the ends of the scale were defined as "Extremely unattractive" and "Extremely attractive."

Finally, subjects made preferential choices between pairs of dates in which each date was described by just a photograph. As in Experiment 1, subjects were aware that they would not actually go on a date. However, subjects were instructed to assume that they were in an actual dating situation and to make choices taking into account all factors that might be important in choosing a date. Furthermore, it was emphasized that in this dating situation, like any other, it would be unclear whether their dating choice would reciprocate.

Twenty-one pairs of dates were formed from all possible combinations of the seven photographs. For each pair, subjects responded by first giving the name of the chosen date. They then indicated the degree of their preference by using the same response scale as in Experiment 1. The scale was defined as before, except that anchor stimuli were not used to define the ends of the scale. The 21 choice pairs were presented four times, with a different shuffled order each time. Finally, each subject was debriefed and interviewed regarding her thoughts on dating and her approach to making dating choices.

Results of Experiment 2

Results for three representative subjects are graphically presented in Figure 2. The spacing along the horizontal axis for each of the three panels reflects the subject's physical attractiveness ratings of the photographs. Subject (S)1, for example, considered Dick to be the most attractive, Joe to be least attractive, and Bob to be of roughly intermediate attractiveness. The starred points at the top of each panel represent the subject's probability estimates (e.g., S1 assigned the lowest probability to Jim and the highest to Joe). The filled points indicate the subject's mean preference for each date relative to all others (e.g., S1 showed the greatest mean preference for Tom and the lowest preference for Joe).

Comparison of the three panels reveals several types of individual differences. In the physical attractiveness ratings, for instance, Bob was given the highest rating by S5 and S13, whereas S1 rated Bob as only moderately attractive. There were also considerable differences in subjects' probability ratings, with S1 and S5 giving low probability estimates to highly attractive dates, whereas S13 rated all dates fairly high on probability. Perhaps the most pronounced differences were in the preference values; although Dick was highly preferred by S13, he was one of the least preferred by S1 and S5. Individual differences of comparable magnitude were observed for the remaining subjects as well.

In spite of these differences, subjects revealed some important similarities in the shapes of their preference curves. The left and center panels of Figure 2 are representative of results obtained for 11 of the 15 subjects; the plots show a peaked preference curve with the greatest preference for dates of intermediate attractiveness. Further, there is a marked similarity between the probability and the preference curves. Thus, these 11 subjects preferred dates of intermediate attractiveness, and it appeared that probability was related to these intermediate preferences.

As exemplified by S13, the remaining four subjects revealed a rather different preference pattern. These subjects showed an increasing curve and preferred the most attractive dates; moreover, their probability estimates were generally flat, with little relation to the

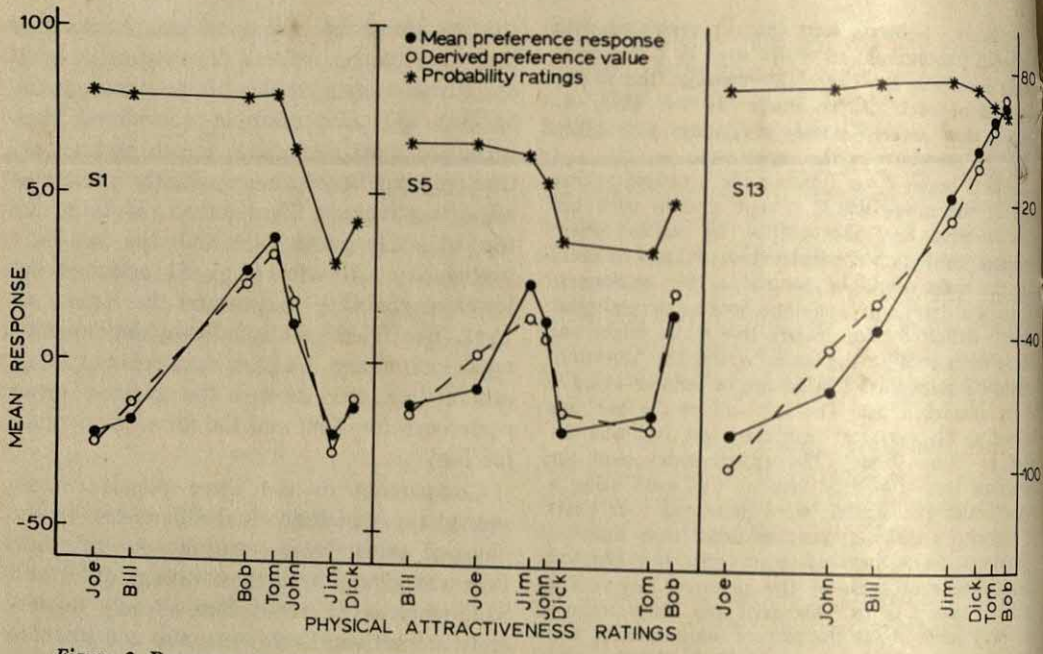


Figure 2. Responses to dates described only by photographs for three typical subjects (S), Experiment 2. (Horizontal spacing reflects physical attractiveness ratings for each date. Stars give probability-of-acceptance ratings for each date. Filled points show mean preferences for each date compared to all others. Open points give values derived from a multiple regression model.)

preference curve. Therefore, these four subjects appeared to be basing their preferences primarily on the attractiveness of the dates.

Statistical analyses. In order to quantitatively describe each individual's responses, separate multiple regression analyses were performed. The basic strategy was to derive a regression equation to predict each subject's preference values based on (a) her probability ratings for each date, (b) her physical attractiveness ratings for each date, and (c) the cross-product values derived by multiplying probability and attractiveness ratings together. Thus, the multiple regression model was based on

$$J = \beta_1 (\text{Physical Attractiveness}) + \beta_2 (\text{Probability}) + \beta_3 (\text{Physical Attractiveness} \times \text{Probability}). \quad (2)$$

The subject's mean preference judgment (J) is expressed as a function of the attractiveness and probability ratings; the weighting constants (β) represent standardized weights derived from a multiple regression analysis.

The prediction terms that accounted for a significant proportion of variance in the preferences of each subject are listed in the second column of Table 2.[1] The terms are listed in their order of variance accounted for. The subjects were placed into two groups depending on whether they preferred dates of intermediate attractiveness (peaked curve) or dates of greatest attractiveness (increasing curve).

As expected, the physical attractiveness

[1] When predictor terms are intercorrelated, as is the case here, the squared semipartial correlation is the best indicant of a term's significance (Darlington, 1969). The squared semipartial correlation measures the contribution of a predictor term after the variance accounted for by all preceding terms has been removed. Thus, when predictors are intercorrelated, the semipartial correlation depends heavily on the order in which terms are entered into the regression analysis. To alleviate this problem, the order should be determined on the basis of theory or some other predetermined criterion (Kerlinger & Pedhazur, 1973). In this experiment, the shapes of the preference curves were predictive of certain combination rules (e.g., a multiplicative rule for subjects with a peaked preference curve) and were therefore used to determine the order in which terms were entered into the regression analysis.

term accounted for the most variance for three of the four subjects who chose the most attractive dates. Of more interest, the cross product was dominant for 9 of the 11 subjects who preferred dates of intermediate attractiveness; for these subjects, probability did appear to influence dating preferences in the form of a multiplicative combination of attractiveness and probability.

Of additional interest, several of the subjects who chose dates of intermediate attractiveness also revealed a significant attractiveness term (e.g., S5). These subjects generally had preference curves in which the greatest preference was for a date of intermediate physical attractiveness, but there was also a moderately high preference for the most attractive date (as seen for S5 in Figure 2). This may represent a compromise strategy between a simple peaked strategy (e.g., S1) and an increasing preference strategy (e.g., S13).

To determine whether each subject's regression equation could adequately describe her preference judgments, the discrepancies between actual and derived preferences were examined. As illustrated by the open circles in Figure 2, the discrepancies were visually quite small. This result suggests that the predictor terms listed in Table 2 for each subject provided a reasonable account of the preferences.

At a statistical level, however, analyses of the discrepancies revealed small but consistent significant differences. When tested against within-cell error, significant discrepancies were observed for 10 of the 15 subjects, including S5 and S13 in Figure 2. Examination of these discrepancies generally revealed that they were quite small. Therefore, it appears that the regression equations provide at least a reasonable first approximation to the preferences.

To further examine the usefulness of the

Table 2
Significant Predictor Terms and Number of Correctly Predicted Choices From Regression Equations and Attractiveness Ratings for Subjects Showing Peaked and Increasing Preference Curves

Subject No.	Significant predictor terms	Number of correctly predicted choices from regression equation	Number of correctly predicted choices from attractiveness ratings
	Peaked curve		
1	(PA × Prob.), PA	21	12
2	PA, (PA × Prob.), Prob.	17	15
3	(PA × Prob.)	18	14
5	(PA × Prob.), PA	18	11
6	Prob.	15	15
7	(PA × Prob.), Prob.	17	16
8	(PA × Prob.), PA	18	17
9	(PA × Prob.), PA	21	21
10	(PA × Prob.)	19	16
12	(PA × Prob.), PA	18	18
14	(PA × Prob.), PA	21	20
	Increasing curve		
4	PA	19	19
11	Prob., (PA × Prob.)	21	20
13	PA	21	21
15	PA, (PA × Prob.)	21	20

Note. PA = physical attractiveness; Prob. = probability.

derived regression equations, three post hoc analyses were conducted. The first was based on the ability of each subject's equation to account for her dating choices. The number of choices that could be correctly predicted is presented in the third column of Table 2. Out of a total of 21 choices, an average of 19.0 choices were accounted for by the regression equation (which are based on attractiveness and probability ratings). This ranged from a low of 15 to a high of all 21 correct. More detailed inspection revealed few errors of any magnitude, and most errors occurred when there was little or no difference in the preference ratings. Thus, the choices of most subjects could be accounted for quite well by the predictions obtained from the regression equations.

The second post hoc analysis involved a comparison between choice predictions based on the regression equations and choice predictions based on physical attractiveness values alone. This analysis bears directly on the issue of whether probability is an important component in subjects' preferences. The number of choices that could be correctly predicted based on attractiveness ratings is shown in the right column of Table 2. Out of a total of 21 choices, attractiveness alone was able to account for an average of 17.0 correct choices, with a low of 11 and a high of 21 correct. Many of the errors, however, were quite large. As expected, attractiveness alone was able to account for the choices of the four subjects whose preference curves increased with attractiveness (e.g., S13). In comparison, attractiveness alone was inferior to the regression equation for most of the 11 subjects who preferred dates of intermediate attractiveness (e.g., S1 and S5).

The final post hoc analysis was designed to examine the credibility of the direct ratings of attractiveness. The regression model in Equation 2 depends crucially on these ratings, and any bias or distortion in the ratings can greatly influence the form of the regression equation (Schmidt, 1973). However, substitution of attractiveness values derived for each subject in Experiment 1 into Equation 2 produced no substantial differences in the resulting regression equations. Thus, the present direct rating procedure apparently produced reasonable estimates of individual subjective values.[2]

Results of Experiment 3

The results obtained from the new set of subjects substantially replicated the findings of Experiment 2. For instance, analysis of the preference curves revealed that 9 of the 14 subjects preferred dates of intermediate attractiveness, whereas five subjects preferred the most attractive dates. Comparison of actual preferences with the preferences derived from the regression equation produced only small visual differences; however, the discrepancies were significant for 11 of the 14 subjects. Finally, 18.1 of the 21 choices could be correctly predicted from the regression equations. In all, the results from Experiment 3 were quite similar to those of Experiment 2. This suggests that the results are not dependent on the subjects having prior experience in Experiment 1.

General Discussion

There were two questions addressed in this research: (a) Is probability of acceptance important in choosing a date? (b) How is probability combined with physical attractiveness in making a choice? The discussion will consider the implications of the present findings for these two questions.

Is Probability Important?

The answer from all three of the present experiments is clear and unequivocal: Yes, probability is important in choosing a date.

---

[2] This analysis also bears on an important issue involving scale constants. The addition of constants to any of the terms in Equation 2 can potentially change the form of the derived regression equation. Moreover, since only the direct ratings of probability can be assumed to be on a ratio scale (because probability has a meaningful zero value), the interval scale of attractiveness values may be particularly suspect. The fit of the multiplying model in Experiment 1, however, provides attractiveness estimates on a ratio scale (Shanteau, 1974). Since the substitution of these values into the regression analysis did not change the basic results, the derived regression equations are apparently not subject to problems of scale constants.

Probability was used by subjects when it was explicitly provided in Experiment 1 and when it was left for subjects to infer in Experiments 2 and 3. There is little doubt, therefore, that the females in this research used probability of acceptance in making their dating choices.

The present findings on the importance of probability contrast with previous research that reports little or no effect of probability (e.g., Berscheid et al., 1971; Huston, 1973). There are, however, several design and procedure differences that may account for the discrepancy in results.

To begin with, the present research was designed to isolate the effects of probability without concern for the subject's own level of physical attractiveness. In contrast, prior studies have generally studied probability by comparing the dating choices of attractive and unattractive subjects. However, if there is a small or a diffuse relation between subjects' attractiveness and their use of probability, then this research approach may not be sensitive to the effects of probability.

An attempt was made in the present study to determine whether there was any connection between a subject's attractiveness and her use of probability.[3] However, no evidence was found of any relationship between a subject's appearance and her probability inferences. A further argument against comparing groups comes from the widespread individual differences in Experiments 2 and 3. These differences suggest that subjects' perceptions of probability and attractiveness may vary quite widely. Therefore, analyses based on comparing group differences between attractive and unattractive subjects may well be incapable of detecting the presence of probability effects.

Another difference from past research was that the present subjects did not actually expect to go out with their dating choice. Previous studies, in contrast, have generally attempted to simulate actual dating situations in various ways (Huston & Levinger, 1978). It may be, then, that the present subjects were isolated from "real world" demands, and they might have used probability in the laboratory when in fact they would not normally use such information.

Several lines of evidence, however, suggest that the subjects in this research did approach the present tasks in a manner analogous to choosing real dates: (a) Subjects' verbal reports indicated that they were thinking about the present tasks in much the same way as they would think about real dates; (b) many subjects became quite personally involved in the research, even to the point of wanting to know how they could find out more about the pictured males; and (c) the widespread individual differences between subjects suggest that each subject was carefully evaluating the dates by her own standards and not simply following some stereotypic set of peer standards. Therefore, it would appear that probability was an important factor for the present subjects in their general approach to dating.

How Is Probability Used?

The answer is that, at least to a first approximation, probability is combined with physical attractiveness by multiplying. The visual evidence in support of multiplying, especially in Figure 1, seems quite compelling. However, the presence of small but consistent statistical discrepancies in Experiments 2 and 3 means that the model may not be entirely adequate.

One possible explanation for these discrepancies from multiplying is that subjects may use other information besides probability and

---

[3] If the subject's own physical attractiveness is related to her expectancy of acceptance, then the subject's attractiveness should be related to the dating strategy employed. For example, subjects who always chose the most attractive dates might be expected to be more attractive than subjects who chose dates of intermediate physical attractiveness. Contrary to this prediction, no such relation was observed. Ratings of the subjects' photographs by 60 male undergraduates revealed little or no relationship between the physical attractiveness of a subject and the dating strategy she employed. For example, S1 and S13 in Figure 2 were nearly equal in rated attractiveness, but employed very different dating strategies. It appears, therefore, that the subject's attractiveness does not determine her expectancy for acceptance. Of course, the possibility remains that self-perceived attractiveness (as opposed to outward attractiveness) may be related to probability.

attractiveness in making their choices. One conjecture, by Shanteau and Nagy (1976), is that females may also rely on their inferred compatibility with prospective dates. Evidence for this possibility came from comments by the subjects to the effect that "he is (or is not) my type," and so "we would (or would not) get along together." Therefore, the present discrepancies from multiplying may indicate that more than attractiveness and probability information was being combined when subjects were making choices.

The use of factors other than probability and attractiveness might explain why the multiplying model fit well in Experiment 1 but less well in Experiments 2 and 3. In the first experiment, probability and attractiveness were explicitly manipulated, and the data supported multiplying; there was little opportunity in this experiment for other variables to have any effect. In Experiments 2 and 3, however, only physical attractiveness was manipulated, and subjects were encouraged to use any other information they might actually rely on in choosing dates. Thus, there was an opportunity for compatibility (or other factors) to influence the choices; this would then decrease the fit of the multiplying model, which is based only on attractiveness and probability.

Nevertheless, the model still did impressively well in accounting for subjects' dating choices in Experiments 2 and 3. The derived model accounted for over 90% of the subjects' choices; in comparison, a model based on physical attractiveness alone was able to account for only 75% of the choices for subjects who preferred dates of intermediate physical attractiveness. Thus, combining probability and attractiveness information does an appreciably better job than attractiveness alone. Including information on compatibility or other factors may, of course, increase the percentage even more.

Finally, it is appropriate to make some comment on the general research strategy used in this study. The progression from controlled to more naturalistic research in a single study has been praised by some (Byrne et al., 1970) and criticized by others (Ebbessen & Konečni, 1975). In the present research, this strategy proved quite useful in, among other things, evaluating the strengths and weaknesses of the multiplying rule. The results from the three experiments suggest that probability and attractiveness probably do combine by multiplying, but in more naturalistic settings other factors may enter in as well.

The progression of studies was also useful in empirically deriving the prediction from Experiment 1 that subjects might prefer dates of intermediate attractiveness. This prediction was borne out in Experiments 2 and 3 for roughly two thirds of the subjects. Even the deviations from this pattern were interesting, since they indicate that a substantial minority of females may rely primarily on physical attractiveness in choosing dates.

In all, the present research illustrates how a combination of controlled and more naturalistic research may be more beneficial than either approach taken alone. Together with other studies using a similar strategy (e.g., Phelps & Shanteau, 1978), this study suggests that combining controlled and naturalistic research may prove quite useful in future research involving complex human judgment.

References

Anderson, N. H., & Shanteau, J. C. Information integration in risky decision making. *Journal of Experimental Psychology*, 1970, *84*, 441–451.

Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph*, 1969, *80*(3, Pt. 2).

Berscheid, E., Dion, K., Walster, E., & Walster, G. M. Physical attractiveness and dating choice: A test of the matching hypotheses. *Journal of Experimental Social Psychology*, 1971, *7*, 173–189.

Byrne, D., Ervin, C. F., & Lamberth, J. Continuity between the experimental study of attraction and real-life computer dating. *Journal of Personality and Social Psychology*, 1970, *16*, 157–165.

Darlington, R. Multiple regression in psychological research and practice. *Psychological Bulletin*, 1969, *69*, 161–182.

Ebbesen, E. B., & Konečni, V. J. Decision making and information integration in the courts: The setting of bail. *Journal of Personality and Social Psychology*, 1975, *32*, 805–821.

Edwards, W. Behavioral decision theory. *Annual Review of Psychology*, 1961, *12*, 473–498.

Huston, T. L. Ambiguity of acceptance, social desirability, and dating choice. *Journal of Experimental Social Psychology,* 1973, *9,* 32–42.

Huston, T. L., & Levinger, G. Interpersonal attraction and relationships. *Annual Review of Psychology,* 1978, *29,* 115–156.

Kerlinger, F., & Pedhazur, E. *Multiple regression in behavioral research.* New York: Holt, Rinehart & Winston, 1973.

Kleck, R. E., & Rubenstein, C. Physical attractiveness, perceived attitude similarity, and interpersonal attraction in an opposite-sex encounter. *Journal of Personality and Social Psychology,* 1975, *31,* 107–114.

Lewin, K., Dembo, T., Festinger, L., & Sears, P. Level of aspiration. In J. McV. Hunt (Ed.), *Personality and the behavior disorders* (Vol. 1). New York: Ronald Press, 1944.

Lichtenstein, S., & Newman, J. R. Empirical scaling of common verbal phrases associated with numerical probabilities. *Psychonomic Science,* 1967, *9,* 563–564.

Phelps, R. H., & Shanteau, J. Livestock judges: How much information can an expert use? *Organizational Behavior and Human Performance,* 1978, *21,* 209–219.

Schmidt, F. L. Implications of a measurement problem for expectancy theory research. *Organizational Behavior and Human Performance,* 1973, *10,* 243–251.

Shanteau, J. Component processes in risky decision making. *Journal of Experimental Psychology,* 1974, *103,* 680–691.

Shanteau, J. An information integration analysis of risky decision making. In M. Kaplan & S. Schwartz (Eds.), *Human judgment and decision processes.* New York: Academic Press, 1975.

Shanteau, J. POLYLIN: A FORTRAN IV program for the analysis of multiplicative (multilinear) trend components of interactions. *Behavior Research Methods & Instrumentation,* 1977, *9,* 381–382.

Shanteau, J., & Nagy, G. Decisions made about other people: A human judgment analysis of dating choice. In J. Carroll & J. Payne (Eds.), *Cognition and social judgment.* Potomac, Md.: Erlbaum, 1976.

Walster, E., Aronson, V., Abrahams, D., & Rottmann, L. Importance of physical attractiveness in dating behavior. *Journal of Personality and Social Psychology,* 1966, *4,* 508–516.

Received December 19, 1977 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 4, 534–550

Consequences of Responsiveness in Dyadic Interaction: Effects of Probability of Response and Proportion of Content-Related Responses on Interpersonal Attraction

Deborah Davis
University of Nevada, Reno

William T. Perkowitz
Loyola University

"Responsiveness" is defined in terms of two sequential response contingencies: (a) the probability with which each person in an interaction responds to the communicative behaviors of the other, and (b) the proportion of responses that are related in content to the preceding behaviors of the other. Two experiments examined the effects of responsiveness in a verbal exchange on attraction. Under the guise of a study of the "acquaintanceship process," 176 male and female subjects exchanged information about themselves with another subject (actually a same-sex confederate) by taking turns choosing and answering one of either two or three questions about themselves on each trial. For Experiment 1, subjects were required to answer on all trials, whereas the probability and frequency with which the confederate responded to the subject were orthogonally manipulated. For Experiment 2, the proportion of content-related responses was varied. The confederate answered the same question as the subject on either 80% or 20% of the trials. Both the probability of response and the proportion of content-related responses were positively related to (a) attraction to the confederate, (b) subjects' perceptions of the confederate's attraction to themselves, and (c) the degree to which subjects felt that they and the confederate had become acquainted with one another.

Social interaction has been described as "the very stuff of human life" (Goldschmidt, 1972, p. 59). Certainly, at the least, it may be described as the "very stuff" of human relationships. Relationships are formed and maintained through a series of interactions and are often formed simply for the sake of interaction.

Despite the importance of interaction, by far the majority of empirical research on love, attraction, and friendship has focused on the effects of static characteristics of others (e.g., similarity, physical appearance, personality or behavior descriptions, etc.). Although such characteristics are undeniably important, adequate analysis of the determinants of attraction must also relate the process of interaction to attraction. The present article is concerned with one dimension of interaction—responsiveness—which, it is argued, functions as a major determinant of the quality of interaction and, hence, attraction.

Because the present concept of responsiveness is new to social psychology, before describing two empirical studies relating responsiveness to attraction we will (a) define responsiveness and distinguish it from other related concepts, (b) describe the functions of responsiveness and their implications for attraction, and (c) summarize relevant literature.

The Concept of Responsiveness

Responsiveness is defined in terms of two sequential contingencies: (a) the probability with which each participant responds (either

We are indebted to David Biscardi and Rita Hargood, who served as confederates, to Mark McDade, who built the apparatus for Experiment 1, and to Roger Bakeman, Timothy C. Brock, Thomas M. Ostrom, and Robert S. Wyer, Jr., for their comments on previous drafts of this manuscript.

Requests for reprints should be addressed to Deborah Davis, Department of Psychology, University of Nevada, Reno, Nevada 89557.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3704-534$00.75

verbally or nonverbally) to the communicative behaviors of the other and (b) the proportion of responses that are related in content to the preceding communicative behavior of the other. It should be noted that the first contingency (probability of response) is applicable to both related and unrelated responses. It is assumed that an unrelated response is better than no response. With regard to the second contingency (relatedness of response content), three conceptual distinctions should be emphasized.

Perception of relatedness versus logic or intent of the speaker. We would like to stress that what is important for attraction is the *perception* that the other's responses are related to one's own communications. Responsiveness is "in the ear of the beholder," so to speak. If a response is perceived to be unrelated, it will be functionally eqivalent to any other unrelated response, regardless of whether it was logically related, or intended by the speaker to address the preceding communication.

Content versus causality. The relationship between the *content* of successive responses should not be confused with the nature of the *causal* relationship between them. Jones and Gerard (1967), for example, classify interactions into four categories that vary in the extent to which each participant's behaviors are *determined* by the preceding behavior of the other versus by his or her own "plan" for the interaction. Likewise, Mehrabian's (1969, 1970, 1971) definition of responsiveness focuses on the extent to which one's behaviors are influenced by those of the other. For Mehrabian, responsiveness is indexed by the amount of *change* in one's behaviors (e.g., facial and vocal expressions, rate and volume of speech, etc.) resulting from another's presence or behavior.

In contrast, the present concept of relatedness of response content is concerned only with the degree to which the response is perceived to address itself to the content of the preceding message. Responses that are determined or influenced by the other's preceding behavior may nevertheless be unrelated in content—as in the case of unrelated associations that may be triggered by the other's behavior. These responses will, however, be more frequently characterized by related content than those not influenced by the preceding behavior. On the other hand, responses that are not causally related may be related in content. In this case, the frequency with which related responses will occur will depend upon the frequency with which responses that are related in content arise independently.

Relatedness versus valence. A related response need not be positive. It may be friendly, hostile, agreeing, disagreeing, defiant, cooperative, and so on. A punch in the nose and a pat on the back may be equally related to a preceding behavior, although different in valence. We are concerned here only with the degree to which each behavior addresses the content of the preceding behavior, not with the way in which it does so. This is not to say that valence has no effect on attraction, but rather that valence and relatedness are independent in their effects.

The Functions of Responsiveness—Why is it Important?

When two persons interact, each expects the other to be responsive—to respond to his or her communicative behaviors, and to respond with content that addresses itself to the intended message. As Ruesch and Bateson (1951) have pointed out, each behavior in an interaction has a certain "demand" aspect, requiring a response of a certain kind. Certainly, in order to fulfill the demand, the response must address the content of the preceding behavior. We expect a question to be followed by an answer, or an expression of disgust with the weather by a comment on the same. Changes of subject are, of course, acceptable, but only after at least a brief related response. Such mutual responsiveness between the participants will serve at least four functions for the interaction, each with potential implications for attraction.

Maintenance of interaction. Perhaps most fundamental of the functions of responsiveness is its role in the maintenance of interaction. Although interactions may persist in the face of occasional failure to respond (as indeed occurs in virtually all interactions), repeated failure to respond will typically result in cessation of interaction. Likewise,

irrelevant responses do not facilitate interaction and, in fact, are frequently used to either terminate an interaction or to change the subject (Watzlawick, Beavin, & Jackson, 1967). Attraction has been shown to be a positive function of the opportunity for interaction (e.g., Ebbesen, Kjos, & Konečni, 1976; Festinger, Schachter, & Back, 1950) and of the frequency of interaction (Brockner & Swap, 1976; Byrne, 1961; Kipness, 1957; Saegert, Swap, & Zajonc, 1973). Through its influence on the maintenance of interaction, responsiveness should likewise facilitate attraction.

Predictability, control, and stress. To the extent that another is responsive, his or her behavior will be in some degree predictable. Although the exact content of a response is typically not predictable, one may at least expect (a) that the other *will* respond and (b) that the response will address the content of the preceding message. If the other is unresponsive, even this degree of predictability is lost.

Similarly, the degree to which the other is responsive will affect the magnitude of control one may exert over the course of the interaction. Clearly, one cannot often exert complete control. Neither the exact content of the other's responses nor the specific nature of the behaviors the other will initiate will typically be controllable. However, responsiveness from the other will allow some degree of control, in that one can (a) cause the other to respond and (b) determine the general topic area (albeit not the specific content) of the response.

Because of the constant uncertainty associated with unpredictability and the frustration associated with lack of control, interaction with an unresponsive other will be stressful. Arnold, Veitch, and Arkkelin (Note 1), for example, have recently demonstrated that social situations that are perceived as most uncomfortable are those characterized by both high arousal and lack of control. Since others who are associated with stress should be less attractive than those who are not, again, it follows that responsiveness will influence attraction.

Facilitation of interaction goals. To the extent that responsiveness facilitates control over the course of the interaction, it should also facilitate interaction goals. Most behaviors or communicative acts in interaction have a purpose. It may be as simple as the maintenance of interaction or as complex as an attempt to manipulate or persuade. Whatever the goal of the act, response from the other is necessary, but not sufficient, for its fulfillment. Although response may or may not facilitate the initiator's purpose, failure to respond will almost certainly frustrate it. Similarly, related responses will be more likely to facilitate interaction goals than will irrelevant responses. Since others who are associated with reward (facilitation of interaction goals) should be more attractive than those associated with punishment (frustration), responsiveness would again be expected to facilitate attraction.

Communication of interpersonal affect. The degree of responsiveness between participants in an interaction reflects the relationship between them—their interest in one another, as well as in their respective communications. When interacting with another who frequently either fails to respond or responds irrelevantly, one is tempted to conclude that the other is interested neither in oneself nor in what one has to say. Certainly, one is left with the feeling that no real "interaction" has taken place, and that no "relationship" has been developed. Responsive interaction, on the other hand, will facilitate the perception of a "relationship" between the participants as well as the perception of mutual interest and attraction.

As Heider (1958) has argued, the perception of a "unit" relationship should produce a corresponding sentiment relationship (attraction). Similarly, attraction to another is facilitated by the perception that he or she is attracted to oneself (e.g., Backman & Secord, 1959; Harvey, Kelley, & Shapiro, 1957; Jones, Gergen, & Davis, 1962). Responsiveness, to the extent that it strengthens the perception of a bond between the participants and of mutual interest and attraction, should likewise facilitate attraction.

Empirical Evidence

Although research on attraction has not dealt explicitly with responsiveness, evidence

of its importance has appeared in several lines of research. Responsiveness is defined in terms of two response contingencies—probability of response and relatedness of response content. The existing literature relevant to the two response contingencies is summarized below.

Probability of response. Response to another's behavior may be as simple as an orienting response or other indication of attention or as complex as an elaborate compliment. The power of attention as a reinforcer has been well documented by the countless reports of the use of "differential attention" as a technique for behavior modification (see Hersen & Barlow, 1976, pp. 339–352). Verbal and nonverbal responses indicating attention to another have also been shown to facilitate attraction. Research on eye contact (cf. Ellsworth & Ludwig, 1972), for example, has demonstrated that under appropriate circumstances, visual attention to another will facilitate attraction. Likewise, Rosenfeld's (1966) work on affiliation has documented the importance of verbal indicators of attention or acknowledgment. Rosenfeld recorded various verbal and nonverbal behaviors of subjects instructed to either seek or avoid approval from a naive interlocutor (interlocutress). Approval seekers used more "recognitions" (e.g., brief utterances such as "Mmhmm," "Hmmm," "Yeah," "No kidding?" "Really?", etc.) than approval avoiders; and furthermore, "recognition" scores were positively correlated with approval actually received from subjects' naive partners. Finally, research on physical pleasuring (Davis & Brock, in press; Davis & Martin, 1978; Davis, Rainey, & Brock, 1976) has shown that attraction to a recipient of physical pleasure is facilitated by responsiveness. Recipients who responded to receiving pleasure (by expressing enjoyment) were liked better than those who remained silent.

The importance of response has been further documented by studies of mother–infant interaction. Maternal responsiveness has been shown to facilitate both attachment to the mother and the development of behavioral competence for the infant (see Appleton, Clifton, & Goldberg, 1975, and Martin, 1975, for reviews of this literature). Konner (1975) has also stressed the importance of responsiveness, pointing out that children often prefer to interact with older children or adults, rather than peers, because the former are more "contingently responsive"—presumably with respect to both probability and relatedness of response.

Finally, in a series of studies of responsiveness and social attraction between rodents, Latané and his colleagues have demonstrated that while static qualities of the animals (e.g., color, texture, smell, etc.) do not affect attraction, attraction is strongly influenced by factors affecting the capacity for movement and response (see reviews by Latané & Hothersall, 1972; Werner & Latané, 1974). Rats are relatively unattracted, for example, to caged, stuffed, or anesthetized rats, whose capacity for response is restricted, and to unresponsive objects (e.g., tennis balls, warm water bottles, Plexiglas tubes, etc.), but are quite attracted to a responsive human hand. Additionally, attraction is facilitated when the capacity for response is enhanced by injection of adrenaline or caffeine, and it is inhibited when rats are made sluggish by injection of chlorpromazine or alcohol.

While the results of the research described above are consistent with the hypothesis that attraction will vary positively with the probability of response, the various manipulations (e.g., anesthetizing, stuffing, eye contact) and measures (e.g., recognitions) have involved initiative as well as responsive behaviors, and have not varied the probability of response independently of response valence, and in relation to specified units of perceiver behavior.

Relatedness of response content. Although the concept of relatedness of response content has not been prominent in social psychological theories of love and attraction, it has played an important role in other areas of research.

The problem of unresponsive interaction, under the label "disqualification," has been prominent in discussions of mental illness and family and marital dysfunction. "Disqualifications" are responses that "invalidate" the messages that evoke them (Danziger, 1976). Such invalidation may be achieved by changing the subject, failure to respond, ignoring the message, misunderstanding, interruption,

tangentialization, literal interpretation of metaphors and vice versa, and so forth—essentially, by irrelevant or unrelated responses. The concept of disqualification has been central to the double-bind theory of schizophrenia (e.g., Bateson, Jackson, Haley, & Weakland, 1956; Sluzki, Beavin, Tarnopolsky, & Vernon, 1967), and interactions marked by frequent disqualifications have been indicted as contributors to inadequate development of the self-concept (e.g., Laing, 1961) and have been shown to characterize dysfunctional marriages and families (e.g., Watzlawick et al., 1967; Williams, 1969; Follingstad & Haynes, Note 2).

Deriving from the client-centered tradition, responsiveness has played an important role in counseling and psychotherapeutic theory and technique—under the label "empathy." "Empathic understanding," for example, as defined by Carkhuff and Berenson (1967), is defined in terms of the extent to which the content of the therapist's responses reflects understanding of, and addresses itself to, the true meaning (both verbal and emotional) of the client's communications. The ability of the therapist to facilitate the client's problem solving and growth, they argue, is directly related to the depth of the counselor's empathic understanding.

Finally, in contrast to the intimacy of therapist–client or family and marital relationships, Petty and Brock (1976) have studied the effects of relevant versus irrelevant responses from a public speaker to a heckler. The degree to which the speaker's responses were related in content to the heckler's comments significantly affected the audience's agreement with, and reactions to, the speaker. Relevant responses produced significantly greater agreement with the message than irrelevant responses, and the responsive speaker was perceived as more competent than the unresponsive one.

Again, although the literature summarized above is suggestive of the relationship between responsiveness and attraction, direct empirical evidence is lacking. The present two experiments were designed to test the hypothesis that attraction will vary positively with responsiveness. Experiment 1 dealt with the first component (probability of response) and Experiment 2 with the second component (relatedness of response content). As argued earlier, responsiveness will serve to communicate interest and attraction, as well as to strengthen the perception of a "relationship" between participants. Thus, it was also expected that responsiveness would affect subjects' perception of the other's attraction toward them and the degree to which they felt that they and the other had become acquainted with one another.

Experiment 1: Probability of Response

Proper manipulation of the probability of response requires identification of the behavioral unit that "demands" response. While in the "passive" or "listener" role, each participant in an interaction will emit a number of brief responses such as Rosenfeld's (1966) recognitions (e.g., smiles, head nods and shakes, gestures, facial expressions, brief utterances such as "I see," "Yeah," etc.), which provide important feedback for the other. Such responses have been variously termed "back-channel responses" (Duncan, 1975; Duncan & Fiske, 1977; Krauss, Garlock, Bricker, & McMahon, 1977; Yngve, Note 3), "listener responses" (Dittman & Llewellyn, 1967, 1968), "concurrent feedback" (Krauss & Weinheimer, 1966), or "signals of continued attention" (Fries, 1952). While the concept of responsiveness is intended to include such "back-channel" responses, because of the difficulty of identifying the unit of behavior that "demands" them, the present study dealt only with the probability of response at those junctures at which the exchange of the "active" (or speaker) role is demanded.

Under the guise of a study of the "acquaintanceship process," subjects exchanged information about themselves with another "subject" (actually a tape recording) by taking turns choosing and answering one of two questions on each of 6, 12, or 24 trials. Subjects were instructed to answer a question on each trial. However, they were told that their partner would be allowed to choose, for each trial, whether or not to answer a question. The tape responded for either 33% or 66% of the trials.

Table 1
Design of Experiment 1

Frequency of response	Probability of response	
	33%	66%
4	12 trials	6 trials
8	24 trials	12 trials

In the present situation, if the number of trials were held constant, probability and frequency of response would be confounded. Thus, differences in attraction as a function of probability of response, if obtained, could be interpreted as resulting from either differing amounts of information about the stimulus person (see Sloan & Ostrom, 1974) or different frequencies of exposure to him or her (e.g., to the sound of his or her voice; see Zajonc, 1968). In order to separate the effects of the two variables, probability (33% versus 66%) and frequency (4 versus 8) of response were factorially combined in the present experiment, while the number of trials varied between conditions.

Method

Subjects. Subjects were 96 Caucasian introductory psychology students, 48 males and 48 females, who participated for partial fulfillment of course requirements.

Design. In a 2 (probability of response) × 2 (frequency of response) × 2 (sex of subject) design, subjects chose and answered one of two questions on each of 6, 12, or 24 trials. Subjects were responded to on either 33% or 66% of the trials by a same-sex confederate's tape-recorded answer to the same question as that chosen by the subject. In order not to confound probability and frequency of response, the total number of trials was varied in relation to the two variables: In the 33% condition, the tape responded on either 4 of 12 or 8 of 24 trials; and in the 66% condition, the tape responded on either 4 of 6 or 8 of 12 trials. (The design of the experiment is illustrated in Table 1.)

Questions were designed to elicit information about the subject (e.g., "How do you usually spend summers?" "Do you keep house plants?" "What do you expect to be doing at 65?" "What would you do if you suddenly inherited a million dollars?" "Did you work during high school?" "Where do you spend most of your time outside of class?") and were designed to be neutral in content in order to minimize the effects of similarity.

The order of pairs of questions was Latin square counterbalanced for each condition. For the 6- and 12-trial conditions, the 24 pairs of questions were divided into four and two subgroups, respectively. The order of the 6 or 12 pairs was then Latin square counterbalanced within each subgroup, and each subject saw one of the orders within one of the subgroups.

Apparatus. The subjects' cubicle contained an intercom box approximately 60.96 cm wide × 30.48 cm high × 30.48 cm deep. Centered on its front were two panels labeled "Talk" and "Listen," and affixed to it were two buttons, one attached by wire and the other on the box under the "Talk" panel. The two buttons were used by the subjects to signal when ready to begin speaking and when finished speaking, respectively.

Subjects were instructed to signal when ready to begin speaking by pushing the button attached by wire. This, they believed, would cause the other's "Listen" panel to illuminate. When finished speaking, subjects pushed the panel button, which they believed would signal the other to begin by causing his or her "Talk" panel to illuminate. For the responsive trials, the panel button activated a control box in an adjacent room, which, after a fixed delay, caused the subject's "Listen" panel to illuminate and automatically initiated the appropriate taped response. For unresponsive trials, the experimenter pushed a button that caused the light to sequence back through the "Listen" panel to the "Talk" panel without activating the taped response.

Responses to the 48 questions were recorded separately on 96 cassette reels, 48 for a male and 48 for a female confederate. In this way, the experimenter could easily insert the appropriate tape after hearing the subject's choice of question. The tape was always an answer to the same question as the one chosen by the subject.

Procedure. Upon arrival, subjects were seated in a small cubicle containing an intercom system and a set of written instructions. The experimenter explained that another subject was expected, instructed the subject to read the written instructions, and left him or her to wait, closing the door behind him. After a brief delay, the experimenter pretended to lead another "subject" to the adjacent cubicle, audibly instructed the "subject" to read the instructions, and closed the door. The true subject's instructions were labeled "Instructions–Person A," and read as follows:

> This experiment is concerned with the acquaintanceship process as it occurs in brief encounters with another person, and as it occurs without the use of nonverbal cues such as physical appearance, dress style, eye contact, body posture, etc.
>
> You will be communicating with another person using the intercom system (the box in front of you). Both of you will have a booklet, with two questions on each page. You (Person A) will begin by choosing and answering one of the two questions on the first page of the booklet, and you will be required to answer one of the two questions on *all* of the trials.

You may begin your answer when the light is in the panel labeled "Talk." When you are ready to answer, push the button attached by wire to the box. This will activate the intercom and signal the other person that you are going to talk, by causing his or her "Listen" panel to illuminate.

Please begin by telling the other person which question you are going to answer (A or B), and then give your answer. When you have finished, push the red button under the "Talk" panel. This will signal the other person that you are finished by causing his or her "Talk" panel to illuminate.

Person B will then have the option of either responding or not responding on each trial. If Person B chooses to respond, when he or she is ready to talk, he (she) will push the button attached by wire to the box, again activating the intercom, and causing your "Listen" panel to illuminate. When finished, he or she will push the red button under the "Talk" panel, causing your "Talk" light to come on. If Person B chooses not to respond, he or she will just push the red button, and the lights will sequence back to your "Talk" panel.

Please do not communicate in any way other than by answering the questions.

The written instructions were accompanied by diagrams of the subject's own box and "Person B's" box, with further description of their operation.

When subjects had been allowed enough time to read the instructions, the experimenter opened both doors and, from the hallway facing the two cubicles, reviewed and elaborated upon the written instructions for the subject (Person A) and the nonexistent counterpart (Person B). Subjects were reminded that they were to answer on all trials, whereas their partner would have the option to respond or not on each trial. No explanation for this difference was offered. When the verbal instructions were complete, the experimenter instructed the subject to begin when the "Talk" panel illuminated, and closed both doors.

Upon completion of the trials, subjects responded to a questionnaire designed to assess their attraction to the confederate (eight items), perceptions of the confederate's attraction to them (two items), the degree to which they felt they and the confederate had become acquainted (two items), and their reactions to the experiment (two items). The questionnaire also included five items reflecting subjects' perception of confederate responsiveness and five filler items. Subjects also responded to a question designed to assess their awareness of the purpose of the experiment: "What do you think the experimenter expects to learn from the experiment?" "What was he interested in, and what were his hypotheses?" No subject guessed the purpose of the experiment. Subjects were debriefed by mail at the end of the semester.

Results

Attraction to confederate. It was predicted that subjects would like the confederate who responded on 66% of the trials better than the one who responded on only 33%. The postexperimental questionnaire included eight questions [1] designed to measure liking of the confederate (see Table 2). The first five columns of Table 2 indicate, respectively, the dependent measure of attraction and the means for the four treatment combinations.

The predicted main effect of probability of response was highly significant for seven out of the eight questions and in the predicted direction for the eighth. Columns 6 and 7 of Table 2 indicate the Fs and significance levels, respectively. Analysis of variance on the mean of the eight questions also yielded a significant main effect of probability of response, $F(1, 88) = 14.89$, $p < .001$.

There were no significant main effects of either sex or frequency of response, and in no case did probability of response interact with either variable.

Perceptions of confederate reactions to subject. Two of the postexperimental questionnaire items reflected subjects' perceptions of how much the confederate liked them ("How much do you think your partner likes you?") and the confederate's interest in their answers to the questions ("Did you think your partner was interested in the information you sent him/her?"). Although subjects tended to believe that the responsive confederate liked them more than the unresponsive one (Ms = 5.23 and 4.90, respectively) and that he or she was more interested in their answers to the questions (Ms = 4.40 and 3.71, respectively), the main effect of probability of response was only marginally significant for the two questions separately, Fs(1, 88) = 1.81 and 3.44, ps $< .17$ and $< .06$, respectively, and for their mean, $F(1, 88) = 3.67$, $p < .06$. Again, there were no significant effects involving frequency of response or sex.

Perceptions of acquaintanceship. Responsiveness affected the degree to which subjects felt that they and the confederate had become

[1] All questionnaire ratings were made on 11-point scales.

Table 2
Attraction to Confederate as a Function of Probability and Frequency of Response

Questionnaire item	Probability of response: 33% — Frequency of response: 4	33% — 8	66% — 4	66% — 8	Main effect, probability of response: $F(1, 88)$	p
How much do you like your partner?	5.31	5.71	6.38	6.63	12.89	.001
How well do you think you would get along with your partner?	5.08	5.00	6.17	6.63	14.85	.001
Were you interested in the information your partner sent you?	5.46	5.79	6.71	6.92	11.90	.001
How much would you enjoy a casual conversation with your partner, over a beer or coke?	6.21	5.38	7.08	6.96	12.72	.001
How much would you enjoy working on a problem with your partner?	5.08	5.42	6.63	6.29	11.70	.001
Is your partner the type of person you could become close friends with?	4.58	4.54	5.33	5.37	5.63	.02
Is your partner a good listener?	5.58	5.42	6.13	5.67	1.28	.26
How friendly is your partner?	5.63	5.92	6.79	6.46	4.97	.03
M	5.35	5.40	6.40	6.37	14.89	.001

Note. Ratings ranged from 0 ("not at all _____") to 10 ("very _____").

acquainted with one another. Responses to the postexperimental question "How well did you feel that you got to know your partner?" indicated that subjects felt they got to know the confederate who responded on 66% of the trials ($M = 4.09$) better than the one who responded on only 33% of the trials ($M = 3.16$); $F(1, 88) = 8.11$, $p < .005$.

In contrast, in response to the question "How well do you think that your partner got to know you?" subjects indicated that the confederate got to know them better in the 33% condition than in the 66% condition (Ms = 4.90 and 4.19, respectively), although the effect was only marginally significant, $F(1, 88) = 3.17$, $p < .08$. The latter effect is not surprising, however, since subjects answered twice as many questions in the 33% as in the 66% condition (i.e., for 12 and 24 versus 6 and 12 trials). In fact, the two 12-trial conditions did not differ from each other (4.83 versus 4.67), whereas the 6-trial condition was much lower (3.70) and the 24-trial condition higher (4.96).

Again, there were no main effects of either sex or frequency of response and no interactions of either variable with probability of response. It is interesting that frequency of response failed to affect the extent to which subjects felt that they had become acquainted with the confederate, despite the fact that they heard the confederate answer twice as many questions in the high frequency condition. On the other hand, probability of response did affect perceived acquaintanceship, even though the number of answers subjects heard from the confederate was constant across the two conditions.

Ratings of responsiveness. Five trait ratings reflected subjects' perceptions of confederate responsiveness. Again the main effect of probability of response was significant, indicating that responsive confederates (66% response) were perceived as more responsive, $F(1, 88) = 26.24$, $p < .001$; attentive, $F(1, 88) = 6.26$, $p < .02$; talkative, $F(1, 88) = 33.95$, $p < .001$; active, $F(1, 88) = 5.13$, $p < .025$; and sympathetic, $F(1, 88) = 3.80$, $p <$

.05, than the more unresponsive (33% response) confederate. There were no main effects of sex or frequency of response and no interactions of sex with probability of response for any of the above ratings. Although probability and frequency of response interacted for the trait "active," $F(1, 88) = 4.59$, $p < .04$, the two variables did not significantly interact for the other four ratings.

Differences in number of trials. In order to independently manipulate probability and frequency of response, the number of trials varied between conditions (see Table 1). The two 33% conditions involved 12 and 24 trials, whereas the two 66% conditions involved 6 and 12 trials. The effect of probability of response may be tested, however, while holding the number of trials constant.

T tests were performed to test the difference between the two cells involving 12 trials (frequency = 4, probability = 33% versus frequency = 8, probability = 66%) on the means of the three major dependent variable categories. The comparison was significant for all three: attraction to the confederate, $t(46) = 14.48$, $p < .001$; perceptions of confederate attraction to the subject, $t(46) = 9.88$, $p < .001$; and the degree to which subjects felt they got to know the confederate, $t(46) = 15.96$, $p < .001$.

Although probability and frequency of response are confounded for the above comparison, it is doubtful that differences in frequency of response accounted for the results. The overall main effect of frequency of response was not significant, and for the other possible comparison (frequency = 8, probability = 33% versus frequency = 4, probability = 66%) the means were in the opposite direction (i.e., frequency of response was inversely related to the three dependent variables). The failure of frequency of response to affect attraction is, of course, consistent with existing evidence that attraction varies with the proportion, but not number, of positive reinforcers (e.g., Byrne & Nelson, 1965).

Discussion

The results of Experiment 1 are clearly consistent with the hypothesis that attraction will vary positively with responsiveness. Subjects liked the confederate who responded on 66% of the trials more, believed that he or she liked them more, and felt that they got to know him/her better than the confederate who responded on only 33% of the trials. However, since only related responses were employed (the confederate always answered the same question as the subject), there is no direct evidence that the effect of probability of response would hold for unrelated responses as well. Indirect evidence is available, however, through comparison of the present results with those of Experiment 2, which will be described in the general discussion of the two experiments.

Experiment 2: Proportion of Related Responses

Experiment 2 was designed to test the hypothesis that attraction will be a positive function of the proportion of related responses. As in Experiment 1, subjects exchanged information about themselves with another subject (actually a confederate) by taking turns choosing and answering questions about themselves. The subject and confederate chose one of three questions on each of 10 trials. Relatedness was varied through the confederate's choice of question. He or she chose either the same question as the subject (related response) or one of the two different questions (unrelated response). The same question was chosen on either 80% or 20% of the trials.

For half of the subjects, the confederate answered after the subject on each trial, in which case his or her choice could be perceived as a response to the subject's answer. For the other half, in order to control for the possibility that perceived similarity (as a result of choosing the same questions) could affect attraction, the confederate answered first. In this case, his or her responses could in no way be perceived as contingent upon those of the subject and, thus, could not be perceived as a response to the subject. Therefore, for the confederate-first condition, the proportion of related responses should not affect attraction.

Method

Subjects. Subjects were 80 Caucasian introductory psychology students, 40 males and 40 females, who

participated for partial fulfillment of course requirements.

Design. In a 2 (sex of dyad) × 2 (order) × 2 (proportion of related responses) factorial design, subjects and a same-sex confederate[2] alternately chose and responded to one of three questions on each of 10 trials. Using prearranged responses, the confederate answered the same question as the subject on either 80% or 20% of the trials. For half of the subjects, the confederate answered first on all 10 trials, and for the other half, the subject answered first for the 10 trials.

As in Experiment 1, questions were designed to elicit information about the subject (e.g., "What did you do the last time it was a nice day?" "What family traditions do you have around holidays?" "What are your parents' occupations?" "What mode of transportation do you use to get around town?" "What classes are you taking this semester?") and to be neutral and noncontroversial in content. In addition, the three questions for each trial were scaled and equated for intimacy level.

The order of the 10 triads of questions was Latin square counterbalanced, and the two responses that represented the 20% (either same or different) always occurred on Trials 2 and 8.

Procedure. Upon arrival, subjects were seated in a small cubicle containing an intercom system. The experimenter explained that another "subject" was expected and left the subject to wait, closing the door behind him. After a brief delay, another "subject" (actually a same-sex confederate) was audibly led to the adjacent cubicle. The experimenter then opened both doors and instructed both subject and confederate from the hallway facing the two rooms. At no time during the experiment did the subject and confederate come into visual contact.

Subjects were told that the purpose of the experiment was to study the "acquaintanceship process" as it occurs through the exchange of information alone, without the use of the physical and nonverbal cues normally present in everyday interactions. The experimenter explained that in this case, the participants would be communicating over the intercom system and getting acquainted simply by taking turns answering questions about themselves.

Both subject and confederate were given a 10-page booklet containing three questions per page and were instructed to take turns choosing and responding to *one* of three questions available on each of the 10 trials. The experimenter explained that they should begin each turn by announcing which question they were going to answer and signal when finished by saying "OK." It was stressed that conversation should be strictly confined to answering the questions. No other conversation, such as extraneous questions or comments, was permitted.

For the confederate-first condition, in order for the confederate to make the appropriate number of "same" responses, subjects were instructed to choose and indicate their choice of questions on a "choice form," which was collected by the experimenter before the trials began and used to choose the confederates' responses. Subjects also circled their choices on their own booklets, and the experimenter stressed that they were not to alter their choices after the experiment began. Subjects believed that the same procedure was followed by the confederate.

This procedure was employed in order to ensure that for the confederate-first condition, the confederate's choices would not be perceived as contingent upon those of the subject. In the subject-first condition, choice forms were not employed, and both the subjects' and confederates' responses were chosen as the trials progressed.

When the instructions and the choice forms (for the confederate-first condition) were complete, the experimenter allowed the subject and confederate to greet each other in order to become familiar with the intercom system, instructed either the subject or the confederate to begin, and closed both doors.

To keep the confederate blind to condition, a visual cue from the experimenter, who was monitoring the subject's responses via headphones located in the confederate's cubicle, informed the confederate when, and to which question, to give the prearranged response.[3]

Upon completion of the trials, subjects completed the same questionnaire as the one employed in Experiment 1, and again were debriefed by mail at the end of the semester.

Results

Attraction to confederate. It was hypothesized that the proportion of related responses would influence attraction in the subject-first condition, but not in the confederate-first condition. Again, the postexperimental questionnaire included eight questions designed to measure liking of the confederate (see Table 3). The first five columns of Table 3 indicate, respectively, the dependent measure of attraction and the means for the four treatment combinations.

The predicted interaction of order with

---

[2] One confederate of each sex was employed.

[3] If there tended to be distinct patterns for subjects' choices among the three questions for the 10 triads, then there would also be distinct patterns for the questions answered by the confederate. In other words, subjects who answered with 80% related responses would typically answer different questions than those who responded with 80% unrelated responses. Differences in attraction might then be explained by differences in questions answered by the confederate. In order to examine this possibility, the experimenter kept a record of all subject and confederate choices. Analyses of this record indicated that subjects chose all questions approximately equally often and that confederate choices did not differ across conditions.

proportion of related responses was statistically significant for six of the eight questions (see column 6 of Table 3) and for the mean of the eight questions, $F(1, 72) = 8.78$, $p < .004$. For the other two questions, the interaction failed to reach significance by only a small margin.

The simple effect tests for proportion of related responses within the subject-first and confederate-first conditions (columns 7 and 8 of Table 3) indicated that as predicted, related responses facilitated attraction when the subject answered first, but did not significantly affect attraction when the confederate answered first. There were no interactions involving sex for any of the questionnaire items.

Perceptions of confederate reactions to subject. Two of the postexperimental questionnaire items reflected subjects' perceptions of how much the confederate liked them and the confederate's interest in their answers. Again, the proportion of related responses interacted with order, such that when subjects answered first, they believed that a confederate who answered the same question on 80% as opposed to 20% of the trials liked them more (Ms = 6.55 and 5.55, respectively), simple $F(1, 72) = 5.05$, $p < .03$, and was more interested in their answers (Ms = 6.45 and 4.40, respectively), simple $F(1, 72) = 13.54$, $p < .001$; but when the confederate answered first, the proportion of related responses had no effect, Fs$(1, 72) < 1.53$, *ns*. The interaction was significant for both questions, Fs$(1, 72) = 6.07$ and 8.15, ps $< .02$ and $< .006$, respectively, and for their mean, $F(1, 72) = 6.59$, $p < .01$.

Perceptions of acquaintanceship. The proportion of related responses affected the extent to which subjects felt that they and the confederate became acquainted with one another and the extent to which they enjoyed the experiment, again only in the subject-first condition. When subjects answered first, they felt that they got to know the confederate better (Ms = 6.60 and 4.95), simple $F(1, 72) = 8.07$, $p < .006$, and that the confederate got to know them better (Ms = 6.30 and 4.45), simple $F(1, 72) = 11.73$, $p < .001$, when the confederate answered the same question for 80%, as opposed to 20%, of the trials. There were no significant differences for the confederate-first condition. The interaction was significant for both questions, (Fs$(1, 72) = 5.35$ and 5.86, ps $< .03$ and $< .02$, respectively).

Similarity. Neither order nor the proportion of related responses affected subjects' ratings of their partner's similarity to themselves. This result, in combination with the absence of an effect of relatedness for the confederate-first condition, argues against the alternative explanation of the results that the proportion of related responses affected attraction through its effect on perceived similarity.

Ratings of responsiveness. Subjects rated the confederate as more "responsive" when he or she answered 80% of the same questions rather than 20%, $F(1, 72) = 3.49$, $p < .07$, but only for the subject-first condition, interaction $F(1, 72) = 3.21$, $p < .08$. These effects were only marginally significant, however, and responsiveness did not affect ratings of "attentive," "talkative," "active," and "sympathetic," as it did in Experiment 1. This difference probably occurred because subjects' personal definitions of these adjectives are more tied to probability than to relatedness of response.

General Discussion

The results of the two experiments clearly supported the hypothesis that attraction is facilitated by responsiveness. For Experiment 1, probability of response facilitated attraction, as well as the perception of acquaintanceship with the confederate. Subjects also tended to believe that the confederate who responded with high probability liked them more and was more interested in their answers to the questions than the one who responded with low probability.

For Experiment 2, those in the subject-first condition not only liked the confederate who gave 80% related responses more, but also felt that the confederate liked them more and was more interested in their answers and that the two of them had become better acquainted with one another. These results are especially impressive, given the large amount of information (constant across conditions) that

Table 3
Attraction to Confederate as a Function of Order and Proportion of Related Responses

Questionnaire item	Order: Subject first, Proportion of related responses 80%	Order: Subject first, Proportion of related responses 20%	Order: Confederate first, Proportion of related responses 80%	Order: Confederate first, Proportion of related responses 20%	Interaction of Order × Proportion of Related Responses $F(1, 72)$	Interaction of Order × Proportion of Related Responses p	Simple effects proportion of related responses: Within subject-first condition $F(1, 72)$	Within subject-first condition p	Simple effects proportion of related responses: Within confederate-first condition $F(1, 72)$	Within confederate-first condition p
How much do you like your partner?	7.40	6.80	7.25	7.95	4.19	.04	1.79	.18	2.43	.13
How well do you think you would get along with your partner?	6.55	5.90	6.70	7.55	3.51	.07	1.32	.25	2.25	.14
Were you interested in the information your partner sent you?	7.70	6.50	7.25	7.70	4.16	.05	4.40	.04	<1	.44
How much would you enjoy a casual conversation with your partner, over a beer or coke?	7.95	6.85	7.20	7.85	5.09	.03	4.02	.05	1.41	.24
How much would you enjoy working on a problem with your partner?	7.55	6.50	7.00	7.90	6.55	.02	3.80	.06	2.79	.10
Is your partner the type of person you could become close friends with?	6.55	4.95	6.50	7.00	6.62	.02	7.69	.007	<1	.39
Is your partner a good listener?	7.80	6.10	6.95	8.10	13.26	.001	9.44	.003	4.32	.04
How friendly is your partner?	7.85	7.00	8.00	8.25	3.24	.08	3.87	.05	<1	.57
M	7.42	6.33	7.11	7.79	8.78	.004	6.43	.02	2.74	.10

Note. Ratings ranged from 0 ("not at all _____") to 10 ("very _____").

was conveyed by the confederate's answers to the 10 questions and by his or her style of verbal interaction—information that would doubtless carry a great deal of weight for impressions formed of the confederate. It is also impressive that responsiveness affected attraction in the present situation, where relevant responses should not have been necessarily expected or demanded by the situation, and where unrelated responses would not reflect negatively upon the character of the confederate.

As mentioned earlier, since Experiment 1 employed only related responses, there was no direct evidence that the effect of probability of response would hold for unrelated responses. The subject-first, 20% related condition of Experiment 2 represents a comparison in which the probability of *related* response was even lower (20%) than for the low-probability-of-response (33%) condition in Experiment 1. Despite the fact that the former condition involved lower probability of related responses, attraction to the confederate ($M = 6.33$), perceptions of confederate attraction to the subject ($M = 4.98$), and the degree to which subjects felt that they had gotten to know the confederate ($M = 4.95$) were greater than the comparable values for the 33% response condition of Experiment 1 (Ms = 5.38, 4.30, and 3.13, respectively). While comparisons between experiments are tenuous, it is plausible to argue that the differences were due to the relative probabilities of *unrelated* responses (80% versus 0%).

Perceived Acquaintanceship

Perhaps one of the most interesting results of the two experiments was the relationship between responsiveness and perceived acquaintanceship. For Experiment 1, probability of response affected the degree to which subjects felt that they became acquainted with the confederate, even though the number of answers they heard from the confederate was constant across the two conditions. Frequency of response, on the other hand, did not affect perceived acquaintanceship, despite the fact that subjects heard the confederate answer twice as many questions in the high frequency condition. For Experiment 2, the proportion of related responses affected perceived acquaintanceship: Subjects felt that they became better acquainted with the confederate and that the confederate had become better acquainted with them when the confederate responded with 80% related answers. Apparently, responsiveness affected something more basic than attraction, namely, the perception of a "bond" or "relationship" between the subject and confederate.

Literature on pathological communication (e.g., Danziger, 1976; Laing, 1961; Ruesch & Bateson, 1951; Sluzki et al., 1967; Watzlawick et al., 1967) suggests that the relationship between the successive communicative acts in an interaction serves to define the relationship between the participants. As long as a response addresses the content of the preceding message—whether positively, negatively, with agreement or disagreement—it is assumed to provide confirmation of the other as a person. In contrast, an irrelevant response or failure to respond—a "disqualification"—is assumed to communicate the interpersonal message "You do not exist." Danziger (1976) has expressed the point well:

> In recognizing the meaning and intent of a message and addressing myself to its implied demand I also recognize the originator of the message as a person. In disqualifying a message and refusing to recognize its nature I invalidate the claim to self-definition that it implies. If I show by my inappropriate response that a message has not registered I am at that moment refusing confirmation of the personality that is the source of the message. (p. 139)

Certainly, in order to "get to know" an individual, one must first recognize his or her claim to personhood. It may be that for the present studies, responsiveness communicated to the subject that the confederate was reacting to him or her as a person, thereby inducing the perception of greater acquaintanceship. At the least, the data indicate that responsiveness communicated to the subjects that the confederate liked them and was interested in their answers to the questions.

The results for the confederate-first condition in Experiment 2 attest to the importance of the perception of contingency between the responses, both for the definition of the relationship (or the perception of acquaintanceship) and for attraction. Since the responses

were chosen before the interaction began, and since the subject answered after the confederate, the confederate's responses could not have been perceived as responses to the subject and consequently could not have been interpreted as a comment on their "relationship." Thus, not surprisingly, the proportion of related responses affected neither attraction to the confederate, nor subjects' perceptions of how much the confederate liked them, nor the degree to which subjects felt that they and the confederate became acquainted with one another.

A final piece of evidence concerning the importance of related responses came from pilot subjects in the confederate-first 20% related condition. Until the instructions were modified to strictly and redundantly forbid that the choice of questions be altered during the interaction, subjects frequently changed responses that should have been different from the confederate's in order to answer the same question. When confronted, they complained, "But how can we get to know each other if we just keep answering different questions that have nothing to do with each other?" Apparently, subjects felt that it was necessary to *be* responsive in order to become acquainted with the other person.

Implications for Self-Disclosure Research

Research on self-disclosure (e.g., Altman, 1973; Altman & Taylor, 1973) has been concerned with the content (i.e., intimacy level) of information exchanged in dyadic interaction. The present analysis suggests that the sequential *process* of disclosure may be as important as the specific content. As suggested above, greater feelings of acquaintanceship and attraction, as well as greater perception of control, will develop from interactions in which each participant's responses address the content of the other's preceding communication. Thus, it would be expected that those disclosures which most directly address previous disclosures or other behaviors of the perceiver will produce greatest attraction.

The majority of self-disclosure research has focused on the absolute intimacy level of another's disclosures, independently of the perceiver's own. However, it may be argued that disclosures at the *same* level of intimacy will be perceived to be directly related to the preceding communication with greater probability than those at a different level of intimacy. It may also be argued that one will feel more in control of the intimacy level of the interaction to the extent that another's disclosures are of the same level of intimacy as one's own. If so, another will be more attractive the more closely his or her disclosures *match* the intimacy level of the perceiver's.

Directions for Future Research

Although the results of the present two experiments were clearly consistent with the hypothesis that attraction is facilitated by responsiveness, the experimental interactions were highly structured and in many ways quite different from more naturalistic interaction. Both the format of the interaction (i.e., taking turns answering questions) and the range of alternative responses (choice among two or three topics) were defined by the experimenter. In addition, unresponsiveness was specifically condoned by the experimenter. For Experiment 1, the confederate was told that he or she could choose not to respond, and for Experiment 2, that he or she could answer any of the three questions. Because of the structure of the situation, the attributions resulting from the confederate's unresponsiveness—both regarding his or her social graces and his or her feelings toward the subject—should have been less extreme than those concerning unresponsiveness in more naturalistic interactions. Responsiveness should also have been less related to maintenance of interaction, predictability, control, stress, and facilitation of interaction goals in the present situation. Thus, one of the first priorities of future research should be to manipulate responsiveness in the context of less structured more naturalistic interactions, where it would be expected to have even greater impact on attraction.

Given the importance of responsiveness for human interaction, future research might also profitably explore the determinants of two persons' capacity for mutually responsive interaction. At least five conditions are necessary for

responsive interaction. In order to be responsive, one must first attend to the other person, both in order to detect behaviors that demand response and in order to process those behaviors with sufficient accuracy to be able to respond with related content. Many situational characteristics, such as other persons present, task demands, noise, or other distracting stimuli, will affect the degree to which attention is available for the person with whom one is interacting. Internal demands for attention, such as self-consciousness, preoccupation with irrelevant thoughts or with bodily sensations, and so forth will also reduce attention to the other person. Finally, stimulus characteristics of the other person may influence the amount of attention devoted to him or her. It is reasonable to assume, for example, that one will devote more attention to those whom one likes, those who are of high status or in control of rewards, or those who are physically attractive.

A second condition necessary for responsive interaction is the motivation to be responsive. Responsiveness has two effects, which may or may not be desirable to the participants: (a) It tends to prolong the interaction and (b) it tends to lead the participants to feel more of a relationship (or more intimacy) with one another. If either or both of the two effects are undesirable to the participants, the motivation to be responsive will be correspondingly low. As Watzlawick et al. (1967) have pointed out, unrelated or irrelevant responses (disqualifications) are commonly employed as a means to either terminate an interaction or to change the subject. Thus, the motivation to be responsive should be affected by such factors as the attractiveness of the other person and the degree to which he or she controls rewards, as well as by temporary factors affecting the desire to interact at a specific time or place or on a particular topic.

Given that a person has attended to the other and is motivated to be responsive, he or she must have the capacity or energy to respond at the rate "demanded" by the other. When interacting with another who talks rapidly and at length, one must have the energy to respond rapidly and at length. If the "interaction rates" of two persons are not matched, the more active participant will feel frustrated by the other's unresponsiveness, while the less active participant will feel drained by the constant demand for response. Chapple (1939, 1974) has argued for years that two persons' capacity for smooth interaction will be determined in part by the match of their activity levels. Similarly, Latané and Hothersall (1972) have argued that the relative preference of rodents for within- versus between-species others results in part from incompatible activity levels, or "interaction rates," between species.

In order to respond with related content, one must first accurately interpret the meaning of the other person's behavior or communication. Accuracy of understanding will be affected, in part, by the content of the communication. The more ambiguous the subject matter, the more difficult it will be to interpret accurately, and therefore the less responsive will be the interaction. The cognitive skills of the participants will also affect the capacity for responsive interaction. The ability to accurately understand communications will be determined both by general verbal skills and by the ability to take on the frame of reference of the other person. Thus, the capacity for responsiveness will increase with the intelligence, general knowledgeability, cognitive development, and cognitive complexity of the participants, and with their familiarity with one another.

Understanding will also be facilitated by many forms of similarity. It would be expected, for example, that any shared characteristic (such as race, culture or subculture, socioeconomic class, profession, interests, cognitive structure, and even shared experience) that would lead two persons to share ways of thinking or of perceiving the world, of using language, or of interpreting behavior would facilitate understanding and therefore responsiveness. Similarity of culture (Triandis, 1964, 1975), socioeconomic class (Harms, 1961), and cognitive structure (Runkel, 1956; Shibuya, 1962; Triandis, 1959, 1960a, 1960b) have, in fact, been shown to facilitate communication.

The fifth, and final, condition necessary for responsive interaction is that the participants

possess the response repertoire necessary for related responses. Again, similarity, as well as general knowledgeability, will be important. Since persons most often initiate conversations dealing with topics in which they are interested or knowledgeable, or with which they are currently concerned, persons who share similar interests, knowledge, or topics of concern will more often engage in responsive interaction than those who do not.

Conclusions

The present demonstration of the importance of responsiveness illustrates the utility of exploration of interaction-based determinants of attraction. Such sequential response contingencies as those described here may be profitably explored as both independent and dependent variables, as illustrated above. As a dependent variable, responsiveness or other sequential response contingencies may allow identification and greater understanding of variables that may influence attraction either partially or entirely through their influence on interaction (e.g., similarity).

Reference Notes

1. Arnold, J., Veitch, R., & Arkkelin, D. *An explication of the meaning of discomfort in social situations*. Paper presented at the meeting of the Midwestern Psychological Association, Chicago, May 1978.
2. Follingstad, D., & Haynes, S. *Assessment and intervention of marital dysfunction*. Paper presented at the meeting of the Association for Advancement of Behavior Therapy, New York, 1976.
3. Yngve, V. H. *On getting a word in edgewise*. Paper presented at the meeting of the Chicago Linguistic Society, Chicago, 1970.

References

Altman, I. Reciprocity of interpersonal exchange. *Journal for the Theory of Social Behavior*, 1973, *3*, 249–261.

Altman, I., & Taylor, D. A. *Social penetration*. New York: Holt, Rinehart & Winston, 1973.

Appleton, T., Clifton, R., & Goldberg, S. The development of behavioral competence in infancy. In F. Horowitz (Ed.), *Review of child development research* (Vol. 4). Chicago: University of Chicago Press, 1975.

Backman, C. W., & Secord, P. F. The effect of perceived liking on interpersonal attraction. *Human Relations*, 1959, *12*, 379–384.

Bateson, G., Jackson, D. D., Haley, J., & Weakland, J. Toward a theory of schizophrenia. *Behavioral Science*, 1956, *1*, 251–264.

Brockner, J., & Swap, W. C. Effects of repeated exposure and attitudinal similarity on self-disclosure and interpersonal attraction. *Journal of Personality and Social Psychology*, 1976, *33*, 531–540.

Byrne, D. The influence of propinquity and opportunities for interaction on classroom relationships. *Human Relations*, 1961, *14*, 63–69.

Byrne, D., & Nelson, D. Attraction as a linear proportion of positive reinforcements. *Journal of Personality and Social Psychology*, 1965, *1*, 659–663.

Carkhuff, R. R., & Berenson, B. G. *Beyond counseling and therapy*. New York: Holt, Rinehart & Winston, 1967.

Chapple, E. D. Quantitative analysis of the interaction of individuals. *Proceedings of the National Academy of Science*, 1939, *25* (February), 58–67.

Chapple, E. D. *Culture and biological man: Explorations in behavioral anthropology*. New York: Holt, Rinehart & Winston, 1974.

Danziger, K. *Interpersonal communication*. New York: Pergamon Press, 1976.

Davis, D., & Brock, T. C. Effects of the recipient's status and responsiveness on physical pleasuring between heterosexual strangers. *Journal of Experimental Social Psychology*, in press.

Davis, D., & Martin, H. J. When pleasure begets pleasure: Recipient responsiveness as a determinant of physical pleasuring between heterosexual dating couples and strangers. *Journal of Personality and Social Psychology*, 1978, *36*, 767–777.

Davis, D., Rainey, H. G., & Brock, T. C. Interpersonal physical pleasuring: Effects of sex combinations, recipient attributes, and anticipated future interaction. *Journal of Personality and Social Psychology*, 1976, *33*, 89–106.

Dittman, A., & Llewellyn, L. G. The phonemic clause as a unit of speech decoding. *Journal of Personality and Social Psychology*, 1967, *6*, 341–349.

Dittman, A., & Llewellyn, L. G. Relationship between vocalizations and head nods as listener responses. *Journal of Personality and Social Psychology*, 1968, *9*, 79–84.

Duncan, S. On the structure of speaker–auditor interaction during speaking turns. *Language in Society*, 1975, *2*, 161–180.

Duncan, S., & Fiske, D. W. *Face-to-face interaction: Research, methods, and theory*. New York: Wiley, 1977.

Ebbesen, E. B., Kjos, G. L., & Konečni, V. J. Spatial ecology: Its effects on the choice of friends and enemies. *Journal of Experimental Social Psychology*, 1976, *12*, 505–518.

Ellsworth, P. C., & Ludwig, L. M. Visual behavior in social interaction. *The Journal of Communication*, 1972, *22*, 375–403.

Festinger, L., Schachter, S., & Back, K. *Social pressures in informal groups: A study of a housing community*. New York: Harper, 1950.

Fries, C. C. *The structure of English: An introduction to the construction of English sentences.* New York: Harcourt, Brace, 1952.

Goldschmidt, W. H. An ethnography of encounters: A methodology for the enquiry into the relation between the individual and society. *Current Anthropology,* 1972, *13,* 59–78.

Harms, L. S. Listener comprehension of speakers of three status groups. *Language and Speech,* 1961, *4,* 109–122.

Harvey, O. J., Kelley, H. H., & Shapiro, M. M. Reactions to unfavorable evaluations of oneself made by other persons. *Journal of Personality,* 1957, *25,* 393–411.

Heider, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

Hersen, M., & Barlow, D. H. *Single case experimental designs: Strategies for studying behavior change.* New York: Pergamon Press, 1976.

Jones, E. E., & Gerard, H. B. *Foundations of social psychology.* New York: Wiley, 1967.

Jones, E. E., Gergen, K. J., & Davis, K. E. Some determinants of reactions to being approved or disapproved as a person. *Psychological Monographs,* 1962, *76*(2, Whole No. 521).

Kipness, D. M. Interaction between members of bomber crews as a determinant of sociometric choice. *Human Relations,* 1957, *10,* 263–270.

Konner, M. Relations among infants and juveniles in comparative perspective. In M. Lewis & L. Rosenblum (Eds.), *Friendship and peer relations.* New York: Wiley, 1975.

Krauss, R. M., Garlock, C. M., Bricker, P. D., & McMahon, L. E. The role of audible and visible back-channel responses in interpersonal communication. *Journal of Personality and Social Psychology,* 1977, *35,* 523–529.

Krauss, R. M., & Weinheimer, S. Concurrent feedback, confirmation, and the encoding of referents in verbal interaction. *Journal of Personality and Social Psychology,* 1966, *4,* 342–346.

Laing, R. D. *The self and others.* London: Tavistock, 1961.

Latané, B., & Hothersall, D. Social attraction in animals. In P. C. Dodwell (Ed.), *New horizons in psychology II.* Harmondsworth, England: Penguin, 1972.

Martin, B. Parent–child relations. In F. Horowitz (Ed.), *Review of child development research* (Vol. 4). Chicago: University of Chicago Press, 1975.

Mehrabian, A. Some referents and measures of nonverbal behavior. *Behavior Research Methods & Instrumentation,* 1969, *1,* 203–207.

Mehrabian, A. A semantic space for nonverbal behavior. *Journal of Consulting and Clinical Psychology,* 1970, *35,* 248–257.

Mehrabian, A. *Silent messages.* Belmont, Calif.: Wadsworth, 1971.

Petty, R. E., & Brock, T. C. Effects of responding or not responding to hecklers on audience agreement with a speaker. *Journal of Applied Social Psychology,* 1976, *6,* 1–17.

Rosenfeld, H. M. Approval-seeking and approval-inducing functions of verbal and nonverbal responses in the dyad. *Journal of Personality and Social Psychology,* 1966, *6,* 597–605.

Ruesch, J., & Bateson, G. *Communication: The social matrix of psychiatry.* New York: Norton, 19

Runkel, P. J. Cognitive similarity in facilitating communication. *Sociometry,* 1956, *19,* 178–191.

Saegert, S., Swap, W., & Zajonc, R. B. Exposure, context, and interpersonal attraction. *Journal of Personality and Social Psychology,* 1973, *25,* 234–24

Shibuya, Y. A. A study in the relationship between cognitive similarity and communication effectiveness. *Japanese Journal of Psychological Research,* 1962, *4,* 173–177.

Sloan, L. R., & Ostrom, T. M. Amount of information and interpersonal judgment. *Journal of Personality and Social Psychology,* 1974, *29,* 23–29.

Sluzki, C. E., Beavin, J., Tarnopolsky, A., & Verne E. Transactional disqualification: Research the double bind. *Archives of General Psychiatry,* 1967, *16,* 494–504.

Triandis, H. C. Cognitive similarity and interpersonal communication in industry. *Journal of Applied Psychology,* 1959, *43,* 321–326.

Triandis, H. C. Cognitive similarity and communication in a dyad. *Human Relations,* 1960, *13,* 17 183. (a)

Triandis, H. C. Some determinants of interpersonal communication. *Human Relations,* 1960, *13,* 27 287. (b)

Triandis, H. C. Cultural influences upon cognitive processes. In L. Berkowitz (Ed.), *Advances experimental social psychology* (Vol. 1). Ne York: Academic Press, 1964.

Triandis, H. C. Culture training, cognitive complexity and interpersonal attitudes. In R. Brislin, S Bochner, & W. Lonner (Eds.), *Cultural perspectives on learning.* New York: Halsted/Wiley, 197 39–77.

Watzlawick, P., Beavin, J. H., & Jackson, D. *Pragmatics of human communication: A study interactional patterns, pathologies and paradox* New York: Norton, 1967.

Werner, C., & Latané, B. Interaction motivates attraction: Rats are fond of fondling. *Journal Personality and Social Psychology,* 1974, *29,* 32 334.

Williams, J. R. Feedback techniques in marriage counseling. In B. N. Ard & C. C. Ard (Eds.) *Handbook of marriage counseling.* Palo Alto Calif.: Science and Behavior Books, 1969.

Zajonc, R. B. Attitudinal effects of mere exposure *Journal of Personality and Social Psychology Monograph,* 1968, *9*(2, Pt. 2).

Received December 27, 1977

Journal of Personality and Social Psychology
1979, Vol. 37, No. 4, 551-563

Behavioral Change in a Constant Environment: Shift to More Difficult Tasks With Constant Probability of Success

Julius Kuhl
Ruhr University, Bochum, West Germany

Virginia Blankenship
University of Michigan (Ann Arbor)

The present experiment was designed as a first attempt to test hypotheses derived from Atkinson & Birch's dynamic theory of action, which proposes a theoretical reorientation from an *episodic* to a *dynamic* view of motivation. Traditional episodic theories of achievement motivation predict constant risk preference over a series of free choices from various difficulty levels when the assumed situation-specific determinant (probability of success) remains constant. In contrast to this, the dynamic theory predicts a shift to more and more difficult tasks, both for success-oriented and for failure-oriented subjects. In addition, the dynamic theory predicts that the initial ambivalence between very easy and very difficult tasks, predicted by traditional theory of achievement motivation for failure-oriented subjects, is quickly replaced by a consistent preference for very easy tasks in that motive group. The results of the present experiment support the predictions derived from the dynamic theory.

When subjects are allowed to choose tasks from a given set of different difficulty levels over an extended period of time, a shift of preference to more and more difficult tasks is the common finding (e.g., Atkinson, Bastian, Earl, & Litwin, 1960; Atkinson & Feather, 1966). The cognitive theory of achievement motivation (Atkinson, 1957; Atkinson & Feather, 1966; Raynor, 1969) and its reformulation within an attributional theory of motivation (Heckhausen, 1973, 1977; Weiner et al., 1971) attribute such shifts in risk preference to respective changes in the relevant situation-specific determinant of risk preference, namely, subjective probability of success. According to this hypothesis, subjects move to more difficult levels during a series of free choices among various difficulty levels because subjective probability of success increases as a result of subjective improvement. With probability of success held constant, no systematic shift in preferred difficulty would be expected. In recent years, a dynamic theory of motivation has been developed that explains behavioral change even if the situation-specific determinants remain constant (Atkinson & Birch, 1970, 1974). The experiment to be described here was designed to test this hypothesis, which was derived from *The Dynamics of Action* (Atkinson & Birch, 1970), stating that shifts to more and more difficult tasks occur even if subjective probability of success is held constant (i.e., has stabilized after a sufficiently long practice period).

This article was prepared during an exchange visit granted to the first author by the German Science Foundation.

The authors wish to thank John W. Atkinson for numerous helpful suggestions and Marjan Kuhl-de Heus and Glenn Gallop for serving as assistant experimenters.

Requests for reprints should be sent to Virginia Blankenship, Department of Psychology, University of Michigan, Ann Arbor, Michigan 48109.

The Dynamics of Action

The basic concepts and assumptions of this theory (Atkinson & Birch, 1970, 1974) are

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3704-551$00.75

summarized in a recent article on computer simulation of imaginative behavior (Atkinson, Bongort, & Price, 1977):

> If a certain kind of activity has been intrinsically satisfying or previously rewarded in a particular situation, there will be an *instigating force* (F) for that activity, attributable in part to strength of motive in the person and in part to the magnitude of incentive for that activity in that situation. This will cause a more or less rapid arousal and increase in the strength of an inclination to engage in that activity, an *action tendency* (T), depending on the magnitude of the force. If a certain kind of activity has been frustrated or punished in the past, there will be an *inhibitory force* (I) and a more or less rapid growth in the strength of a disinclination to act. This is what we now call a *negaction tendency* (N) and conceive as a tendency *not* to do it. The duration of these forces will determine how strong the action tendency or negaction tendency becomes. The latter, the tendency not to do something, will produce *resistance* to the activity. It opposes, blocks, dampens; that is, it subtracts from the action tendency to determine the *resultant action tendency* ($\bar{T} = T - N$). The resultant action tendency competes with other action tendencies for other incompatible activities. The strongest of them is expressed in behavior. The expression of an action tendency in behavior is what reduces it. Engaging in activity produces a *consummatory force* (C), which depends in part on the consummatory value (c) of the particular activity and in part on the strength of tendency being expressed in behavior (i.e., $C = c\bar{T}$). Similarly, the resistance to an action tendency, produced by the opposition of a negaction tendency, constitutes an analogous *force of resistance* (R), which reduces, in a comparable way, the strength of the negaction tendency.

The force of resistance (R) is a function of the negaction tendency and the resistance value (r) of the particular activity (i.e., $R = r \cdot N$). In order to derive predictions from the theory concerning changes in risk preference during a series of free choices among various difficulty levels, definitions have to be made coordinating the basic parameters (F, I, c, and r) to the assumed antecedents in achievement-related situations.

Coordinating Definitions

The instigating force (F_i) to choose Difficulty Level i is defined by the product of the motive to succeed (M_S), the subjective probability of success (P_{s_i}), and the incentive value of success (I_{s_i}), which is considered the complement of P_s ($I_s = 1 - P_s$):

$$F_i = M_S \times P_{s_i} \times (1 - P_{s_i}). \quad (1)$$

Equation 1 differs from the original theory of achievement motivation (Atkinson, 1957; Atkinson & Feather, 1966) in relating the product of motivational determinants to a parameter controlling the *rate of increase* (F_i) rather than the absolute strength (T_i) of a motivational tendency (cf. Atkinson & Birch, 1970, 1974).

Similarly, the inhibitory force (I_i) controlling the rate of increase in the tendency to avoid working on Difficulty Level i is defined as

$$I_i = M_F \times (-P_{s_i}) \times (1 - P_{s_i}), \quad (2)$$

where M_F describes the motive to avoid failure.

Consummatory value (c), that is, the rate of reduction in the tendency to continue choosing a given level of difficulty, is assumed to be greater for success than for failure ($c_s > c_f$). This is consistent with results obtained by Weiner (1965) showing that for success-oriented subjects, level of performance and persistence are greater following failure than following success. The average consummatory value ($\bar{c}$) in a series of repeated attempts at the same level of difficulty may be assumed to be a function of the *objective* probability of success (p_s):

$$\bar{c} = c_s \cdot p_s + c_f \cdot p_f. \quad (3)$$

In the experiment to be described below, objective and subjective probability of success are assumed to be identical.

Similarly, it is assumed here that the resistance value (r), which controls the reduction of negaction tendencies, varies as a function of success and failure. A plausible theoretical assumption would be that resistance as a result of fear of failure is reduced after success to a greater extent than after failure. This is consistent with results suggesting that failure-oriented subjects show better performance and greater persistence after success than after failure (Weiner, 1965). The average force of resistance ($\bar{r}$) in a series of repeated attempts at the same difficulty may be assumed to be a function

of the objective probability of success (p_s):

$$\bar{r} = r_s \cdot p_s + r_f \cdot p_f. \quad (4)$$

Computer simulations of the dynamic theory using Equations 3 and 4 for defining average c and r values, respectively, resulted in predictions that were equivalent to those obtained in trial-by-trial simulations using the appropriate c_s or c_f value after each success or failure outcome, respectively (Kuhl & Blankenship, in press). It was concluded from this result that Equations 3 and 4 may be used to derive predictions from the dynamic theory.

Predicted Change in Risk Preference

Equations 1 to 4 may be used to derive predictions concerning changes in risk preference in a series of free choices from a given set of different difficulty levels. Though mathematical derivations are not possible because of the complexity of the theory and the situation it is applied to, the implications of the theory for a given situation may be found by computer simulations of the theory. Simulations of the dynamics of action yielded the following hypotheses when the basic parameters (F_i, I_i, c_i, r_i) were defined as suggested by Equations 1 to 4. A description of the simulation program (Bongort, Note 1), the simulation procedure, and the simulated results can be found elsewhere (Kuhl & Blankenship, in press).

Hypothesis 1. For success-oriented subjects ($M_S > M_F$), the slope of the regression line describing preferred difficulty ($1 - P_s$) as a function of time (trials) is greater than zero, indicating a gradual shift to more difficult tasks.

Hypothesis 2. The intercept of this regression line does not differ from the intermediate difficulty level ($P_s = .5$). This indicates a preference for intermediately difficult tasks ($P_S = .5$) at the first trial when $M_S > M_F$.

Hypothesis 3. For failure-oriented subjects ($M_F > M_S$), the slope of the regression line describing preferred difficulty ($1 - P_s$) as a function of time (trials) is greater than zero, indicating a gradual shift to more and more difficult tasks.

Hypothesis 4. The intercept of this regression line is at a difficulty level lower than the intermediate one ($P_s > .5$). This indicates a preference for tasks easier than intermediately difficult at the first trial when $M_F > M_S$.

Hypothesis 5. For subjects with low motivation ($M_S \approx M_F$), the slope of the regression line describing preferred difficulty ($1 - P_S$) as a function of time (trials) does not differ significantly from zero.

Hypothesis 6. The mean slope of the regression line is greater for success- than for failure-oriented subjects.

These simulated predictions were confirmed by trial-by-trial simulations using specific c_s or c_f values after each success or failure outcome rather than average c and r values based on Equations 3 and 4.

Initial Risk Preference

Hypothesis 2 corresponds to the well-known implication of the original theory of achievement motivation: Success-oriented subjects prefer intermediately difficult tasks (Atkinson, 1957). For failure-oriented subjects, the simulations of the dynamics of action yielded ambivalence between the easiest and the most difficult task (Kuhl & Blankenship, in press), as predicted by the original theory (Atkinson & Feather, 1966). This prediction, however, described the very first trials only. After the first few trials, a clear-cut preference for the easiest difficulty level is predicted—because of the reduction of negaction tendency, which is assumed to be greater for easier than for more difficult tasks when $r_s > r_f$ (Equation 4). Since the hypotheses are based on the assumption of stabilized P_s values, a practice period is necessary before the beginning of the main phase of the experiment. Hypothesis 4 states the prediction for the main phase, assuming that the short period of ambivalence between extreme difficulty levels has already been finished during the practice period (Kuhl & Blankenship, in press).

Temporal Change in Risk Preference

Hypotheses 1 and 3 contain the main implications to be tested in this experiment.

In contrast to the original theory, a shift to more and more difficult tasks is predicted for both success- and failure-oriented subjects even when P_s is assumed to have stabilized at the end of the practice period. This is explained as a result of consummation, which is assumed to be greater at easy than at more difficult tasks when $c_s > c_f$ (Equation 3). Hypothesis 3 is also in contrast to predictions derived from a recent application of some concepts of the dynamics of action (Revelle & Michaels, 1976). Based on a mistaken interpretation of the coordinating definitions made in the dynamics of action (Atkinson & Birch, 1970), Revelle & Michaels (1976, p. 396) predict that failure-oriented subjects will show a strategy that is the reverse of what is expected from success-oriented subjects (Kuhl & Blankenship, in press). According to the dynamics of action, failure-oriented subjects show—like success-oriented subjects—a trend to increasingly more difficult tasks, because resistance due to negaction tendencies is more quickly reduced for easier tasks than for more difficult tasks (cf. Equation 4). As a result, a tendency to choose more and more difficult tasks will emerge, after a delay that is proportional to the time required to overcome initial resistance.

Forced Choice

Revelle & Michaels's discussion of a special case of the dynamics of action is based on the assumption of continuous operation of consummatory forces, as if a subject is working on all difficulty levels at the same time. This is, of course, an unrealistic assumption for a situation in which a person makes a choice of one among tasks that differ in difficulty. The ideal case of working on all difficulty levels at the same time can be approximated, however, by forcing the subject to work successively on all difficulty levels included in the experiment before repeating the series. Such a "forced-choice" situation is introduced in the experiment described below.

For this experimental condition, the dynamics of action as well as Revelle & Michaels's special case derived from it predict a trend to increasingly more difficult tasks provided initial tendency strengths (T_i) are close to zero. When a considerable number of practice trials precede the main phase of the experiment, however, both theories predict this trend to weaken to the extent that T_is approach the respective asymptotic levels. This change in trend can be seen from Figure 1, which is based on a computer simulation of the forced-choice condition using initial tendency strengths markedly above zero. These predictions are in contrast to what is expected for the free-choice condition: Computer simulations showed that in this situation, the dynamics of action predict the trend to more difficult tasks, both when T_is are close to zero and when T_is are close to the respective (F/c) value (asymptotic level; Kuhl & Blankenship, in press).

The reason for this implication of the theory seems to be related to a positive-feedback effect typical of the free-choice condition. Under free-choice conditions, the strength of a tendency rises considerably higher than the asymptotic level it would approach if the expression of any competing tendency were prevented: Whenever a tendency is in the "dormant" state (i.e., another tendency has gained dominance), it is no longer a function of consummatory force (the necessary determinant of its asymptotic level in the dormant state) and may, as a result, rise above its asymptotic level if the latter is lower than the momentary strength of the dominant tendency. In sum, the dynamics of action predict a stronger trend to more difficult tasks in the free-choice than in the forced-choice condition, whereas Revelle & Michaels's (1976) theory predicts no such interaction, since it is limited to the forced-choice paradigm.

Method

Subjects

Forty male and 37 female subjects were obtained from the introductory psychology course at the University of Michigan (Ann Arbor) during the 1976 fall semester. Participation in a psychology experiment was a requirement of the course. All 77 subjects were tested in group sessions. Of these, 64 subjects (32 males, 32 females) were sub-

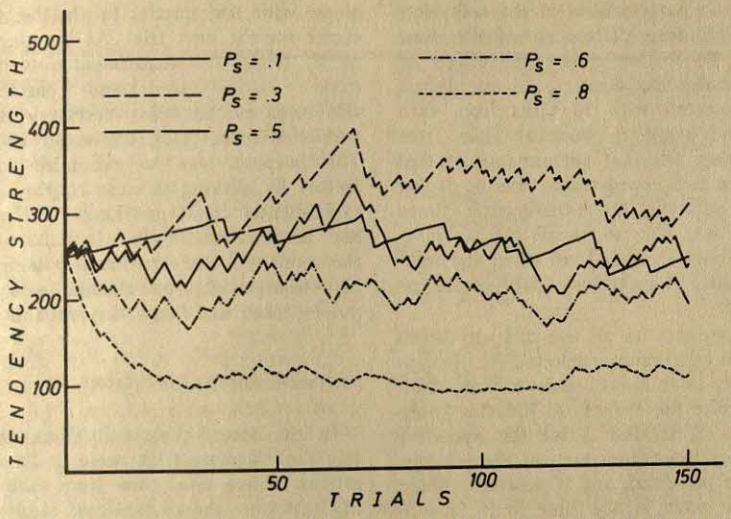


Figure 1. Simulated choice tendencies for the five probability of success (P_s) levels for a forced-choice situation.

sequently selected at random and assigned at random to the two experimental groups for individual testing.

Group Sessions

The two experimenters administered a Thematic Apperception Test (TAT; McClelland, Atkinson, Clark, & Lowell, 1953) using sex-specific verbal cues following Horner (1974) and a short form of the Test Anxiety Questionnaire (TAQ; Mandler & Sarason, 1952) in three group sessions that varied in size from 19 to 29 subjects. The TAT Need for Achievement Scale (n Ach) was administered with standard neutral instructions.

Individual Sessions

Instructions for both experimental groups (forced choice and free choice) were the same in the first part of the individual sessions. The subjects were told that it was the purpose of the experiment to get information about how people feel about a new series of tasks. In order to give knowledgeable answers, they would have to work on a series of these tasks first. Subsequently, subjects were instructed in the task, which was a modified version of a perceptual reasoning task used by Feather (1963). Figure 2 illustrates an example of the task, which had been studied extensively by Kuhl (1977). The objective of the task is to connect the numbered dots in order with an imaginary line starting at the arrow and proceeding to the highest-numbered dot. The correct answer to the task is the number of units (line segments) in the shortest line that will connect the dots in sequential order without violating any of the following rules: (a) The line may not cross, touch, or retrace itself. (b) The borderline of the grid may not be touched or traced. (c) Only the lines within the grid may be traced. Subjects were told that the tasks were similar to items on typical IQ tests and that their opinions were being sought concerning the difficulty of the tasks, problems they saw regarding the tasks, and so on, which would be elicited on a questionnaire to be administered at the end of a prolonged session during which the subjects would gain sufficient familiarity with the tasks.

One puzzle from each difficulty level was solved verbally by the subjects so that the experimenter could check the subjects' understanding and appli-

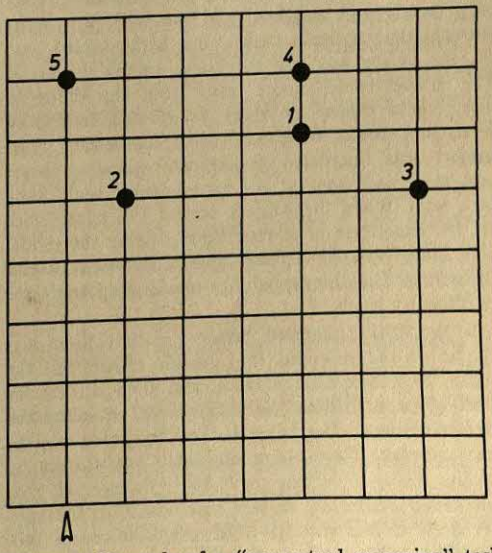


Figure 2. Example of a "perceptual reasoning" task.

cation of the rules. Characteristics of the task were discussed to avoid sudden changes in solution time due to learning. To further reduce the effects of learning on the probability levels to be established, subjects were presented with 10 tasks from each difficulty level and urged to work at tasks from each level until they felt that performance at that level had stabilized and improvement was no longer forthcoming. In addition, the experimenter timed the subjects and was able to verify the stabilization of solution times at each level when the solution times for three consecutive trials varied no more than $\pm \frac{1}{2}$ sec.

When the performance on all five difficulty levels had stabilized, the experimenter selected for the upcoming probability-establishing trials a time limit that would maximize the chances of inducing probabilities of .8, .6, .5, .3, and .1 for the respective difficulty levels. Six tasks from each of the five difficulty levels were presented, and if necessary, slight adjustments were made in the time limit to produce 5, 4, 3, 2, and 1 successes at the respective difficulty levels. The answers were then checked in front of the subject, and the corresponding probabilities, .83, .67, .50, .33, and .17, were computed. These odd and asymmetric probabilities were chosen to maximize credibility of personal probabilities. Labels with these personal probability values were placed in view next to the stacks of corresponding tasks.

Free-Choice Condition

The next phase of the experiment varied, depending on whether the subject had randomly been assigned to the free-choice situation or the forced-choice situation. In the free-choice condition, 50 tasks from each of the five difficulty levels were placed in front of the subject. The personal probability levels were in view, and the subject was instructed that 50 more tasks were to be solved, but that he or she was free to choose among the tasks. The subject could choose tasks from any difficulty level, could choose as many as wanted from any level, and could switch as often as desired. The subject was instructed to indicate when a choice had been made so that the experimenter could time each trial. When the subject turned the puzzle face up, the experimenter started timing, using the modal time limit (varying among subjects between 10 and 20 sec) used in the probability-establishing trials for all difficulty levels.

If the subject finished before the time limit was up, he or she recorded the answer, transferred the puzzle to a box on the table, and then turned the sheet of paper behind the puzzle over, revealing the correct answer. The subject then indicated on the sheet whether the answer obtained was correct by circling *S* for a success or *F* for a failure. If time was called before an answer was obtained, the subject drew a line through that trial, indicating failure. The subject then deposited the answer key, along with the puzzle, in the box and made the choice for the next trial. At the start of this group of 50 trials, the experimenter turned his or her chair so that it faced away from the subject. This was explained as being necessary to keep the experimeter from being distracted from timekeeping. The purpose was to minimize social-desirability factors by making it clear to the subject that the experimenter could not know which difficulty level had been chosen. Only after the subject had left the room did the experimenter learn which choices had been made, by examining the sequence of puzzles that had been deposited in the box.

Forced-Choice Condition

In the forced-choice situation, the subject was told that 50 more trials were to be completed in 10 groups of five tasks, one from each difficulty level. Subjects were shown tables of random permutations and told that the order of the difficulty levels in each group of 5 trials would be random. Again, the experimenter turned his or her chair to face away from the subject. The subject worked the puzzles with the modal time limit established in the probability-establishing trials, checking the answers with the keys behind the puzzles in the same way as the free-choice subjects did. At the end of 5 trials, the subject was asked to indicate on the answer sheet which level of task—easy, moderately easy, intermediate, moderately difficult, or difficult—would be chosen if a free choice were allowed. It was emphasized that this choice would not be acted upon. Each subject in the forced-choice situation made 10 of these choices.

The Questionnaire

At the end of the last 50 trials, subjects in both the free-choice and forced-choice conditions were given the questionnaire that had been mentioned earlier. The main purpose of most questions was to maintain the credibility of the earlier statement about the purpose of the experiment. The questions referred to standard setting, expended effort, and expectation of success. The question regarding the predicted number of successes given 10 more trials at each difficulty level is particularly pertinent to this study, since it was used to select those subjects for whom the manipulation of situational factors (time limit, success and failure feedback, etc.) was successful in inducing and maintaining the personal probabilities established with the subject in the early portion of the experiment. If the average of the probabilities reported by the subject on the questionnaire was more than .1 higher or lower than the average of the established probabilities (.50), the subject was excluded from further analysis. On the basis of this criterion, 25% of the subjects were excluded, 4 because of too low and 12 because of too high subjective probabilities.

Treatment of the Data

Only after the individual sessions had been completed were the TAT protocols scored for n Ach by the second author, who had acquired a reliability correlation of .88 with expert-scored testing materials (Atkinson, 1958). The use of code numbers on the TAT protocols and TAQs further decreased the possibility of experimenter knowledge of subject classification during individual sessions. The TAQ was scored as described by Mandler and Cowen (1958) except that 5 intervals instead of 10 were used. The rank order correlation between TAT and TAQ scores was $-.09$ ($n = 64$), a nonsignificant relationship. Because there was a significant difference between the male and female scores on the TAQ, $F(1, 60) = 9.53$, $p < .01$, data were analyzed separately for males and females. Within each sex, subjects were ranked on both measures, with the lowest score given a rank of 1, and so on. Differences between ranks were obtained by algebraically subtracting the subject's rank on the TAQ from his or her rank on n Ach. Those for whom this calculation produced a high positive number were relatively higher on n Ach than on test anxiety and were considered more nearly $M_S > M_F$ individuals. Subjects with resulting high negative rank differences were relatively higher on test anxiety than on n Ach and were considered more nearly $M_F > M_S$ individuals.

Results

Stability of P_s

To check the assumption of stabilized subjective probability of success during the main phase of the experiment, the mean P_s values were computed from P_s reported at the end of the experiment for each difficulty level. The P_s values were computed for each group of subjects and averaged across the five difficulty levels. As shown in Table 1, the average P_s values are very close to the average subjective probabilities ($P_s = .50$) induced at the five difficulty levels prior to the main phase. A three-way analysis of variance did not show any significant differences in mean P_s between groups. These findings were confirmed by separate analyses of the five difficulty levels using the Newman-Keuls algorithm: None of the tests performed indicated a significant increase in subjective probability over the respective induced value.

P_s Preferences

Since frequency of response is commonly considered an index of motivational preference (e.g., Atkinson & Birch, 1970), the mode of P_s values associated with the preferred difficulty levels was determined for each successive block of five trials from the free-choice situation, and the slope and intercept of the regression line describing P_s preference as a function of time (trials) was computed. This was done for each person within each of the four groups in the free-choice situation. A four-way analysis of variance (Sex × Choice × Motive × Blocks) performed on preferred P_s values revealed a significant main effect of the trial blocks, $F(9, 360) = 4.6$, $p < .0001$, and a significant interaction between choice conditions and trial effect. A trend analysis performed separately for the free-choice and the forced-choice group revealed a highly significant linear trend in the free-choice group, $F(1, 207) = 36.7$, $p < 10^{-7}$, and a significant linear trend in the forced-choice group, $F(1, 207) = 5.1$, $p < .05$. Both trends indicate the expected shift to increasingly more difficult tasks during the course of the experiment. No other trends were significant.

Figure 3 shows the mean preferred P_s levels for the 10 sets of five trials for success- and failure-oriented males. The mean slope of the regression line describing the linear trend to more difficult choices was $M = .43$ in the success-oriented male group and $M = .27$ in the failure-oriented male group (cf. Table 2). The two coefficients depart

Table 1
Mean Probability of Success Reported at the End of the Experiment

Group	Free choice		Forced choice	
	M	*SD*	*M*	*SD*
Males				
$M_S > M_F$	.50	.06	.39[a]	.04
$M_F > M_S$	.46	.04	.52	.06
Females				
$M_S > M_F$	.51	.10	.51	.11
$M_F > M_S$	.52	.06	.50	.13

Note. M_S = motive to succeed; M_F = motive to avoid failure.
[a] Departs significantly (.05 level) from induced average probability of success ($P_s = .50$).

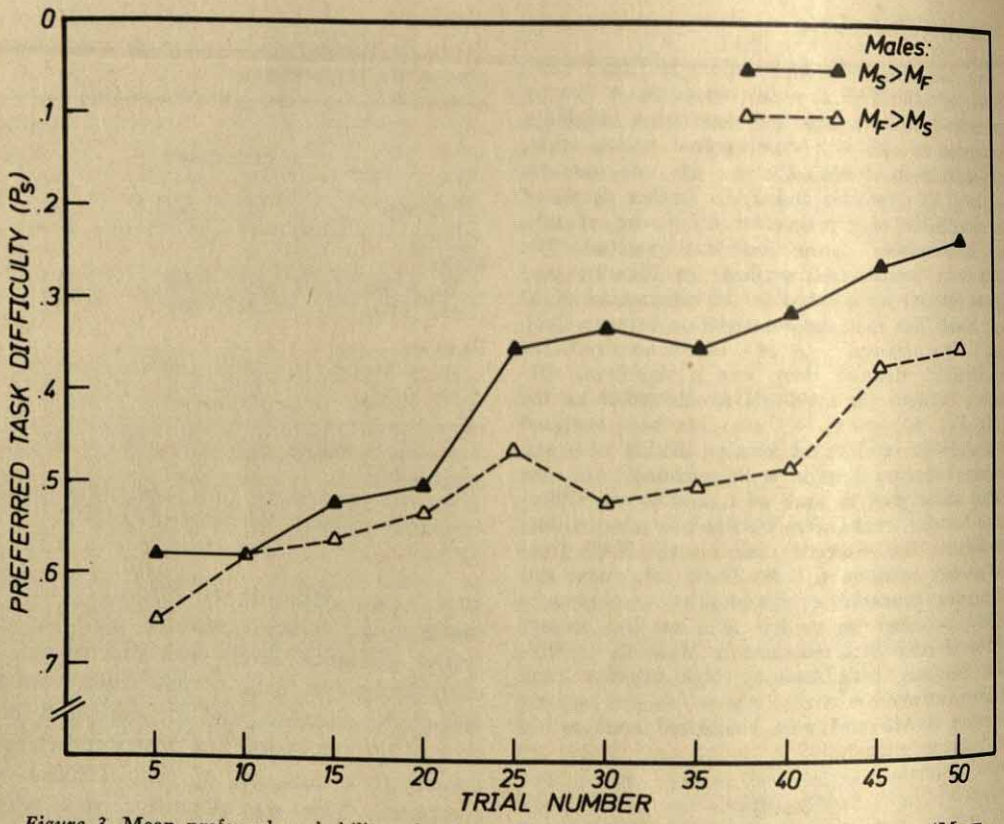


Figure 3. Mean preferred probability of success (P_s) level as a function of motive type (M_S = motive to succeed; M_F = motive to avoid failure) and trials (free choice, males).

significantly from zero, as was expected by Hypotheses 1 and 3; $t(5) = 3.46$, $p < .01$, for success-oriented males, and $t(5) = 4.80$, $p < .0025$, for failure-oriented males. Similar results were obtained for female subjects (cf. Figure 4). The mean slope of the regression line is $M = .26$ for success-oriented females and $M = .22$ for failure-oriented females. The first value approaches statistical significance, $t(5) = 1.55$, $p < .09$, whereas the second value is significantly above zero, $t(5) = 2.59$, $p < .025$. The differences in mean slopes between male and female groups are not significant; $t(10) = 1.24$, *ns*, for success-oriented subjects, and $t(10) = .99$, *ns*, for failure-oriented subjects. The mean slopes from pooled male and female groups are $M = .32$ for success-oriented subjects, and $M = .31$ for failure-oriented subjects. Both values depart significantly from zero; $t(11) = 3.51$, $p < .0025$, and $t(11) = 4.43$, $p < .0005$, respectively.

Table 2
Mean Slopes of Regression Lines for Eight Subject Groups

Group	Free choice		Forced choice	
	Male	Female	Male	Female
$M_S > M_F$	.43[a]	.26	−.03	.23
$M_F > M_S$	.27[a]	.22[a]	.06	.01

Note. $n = 6$ in each cell. M_S = motive to succeed; M_F = motive to avoid failure.
[a] Significantly above zero (.05 level).

The variance of the probabilities preferred by failure-oriented subjects on the first trial of the main phase does not differ significantly from the respective variance for success-oriented subjects, as would be expected if there were any ambivalence between low and high difficulty in failure-oriented subjects, $F(5, 5) = 3.02$, *ns*.

The hypotheses regarding preferred difficulty level at the beginning of the main phase of the experiment (Hypotheses 2 and 4) were confirmed. As can be seen from Table 3, the mean intercepts of the regression line are .63 and .59 for success-oriented males and females, respectively. These values do not differ significantly from the expected value ($P_s = .5$); $t(5) = 2.37$, *ns*, for males, and $t(5) = 1.02$, *ns*, for females. For failure-oriented subjects, the mean intercepts are .65 for males and .71 for females. These values are, as expected (Hypothesis 4), significantly above the intermediate P_s level; $t(5) = 3.37$, $p < .01$, for males, and $t(5) = 3.18$, $p < .05$, for females. Similar results were obtained when first-trial data rather than intercepts were analyzed: The mean P_s level chosen at the first trial was .57 for success-oriented subjects, which does not differ significantly from the expected value (.5); $t(11) = 1.7$, *ns*. For failure-oriented subjects, the mean P_s level chosen at the first trial was .66. This value is, as expected (Hypothesis 4), significantly higher than the intermediate P_s level; $t(11) = 4.17$, $p < .001$. The difference in mean P_s level between motive types approaches significance, $t(22) = 1.53$, $p < .07$ (cf. Hypotheses 3 and 4).

Table 3
Mean Intercepts of Regression Lines for Eight Subject Groups

Group	Free choice		Forced choice	
	Male	Female	Male	Female
$M_S > M_F$	.59	.63	.44	.61
$M_F > M_S$	.65[a]	.71[a]	.49	.57

Note. $n = 6$ in each cell. M_S = motive to succeed; M_F = motive to avoid failure.
[a] Significantly different from probability of success (P_s) = .50.

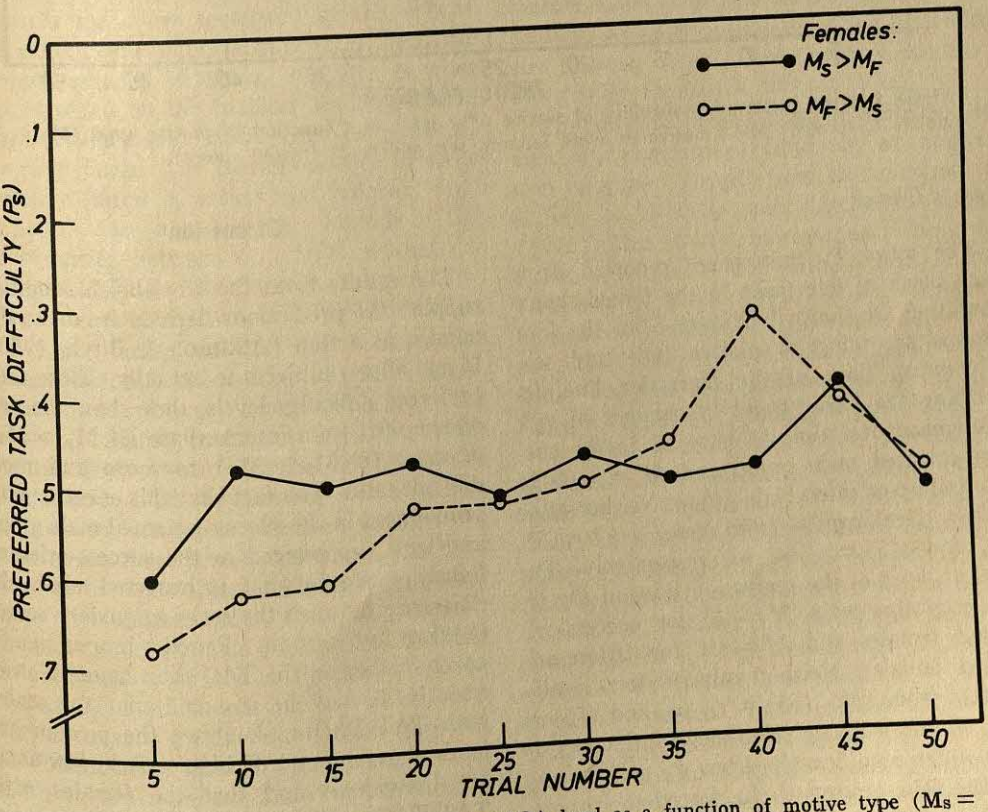


Figure 4. Mean preferred probability of success (P_s) level as a function of motive type (M_S = motive to succeed; M_F = motive to avoid failure) and trials (free choice, females).

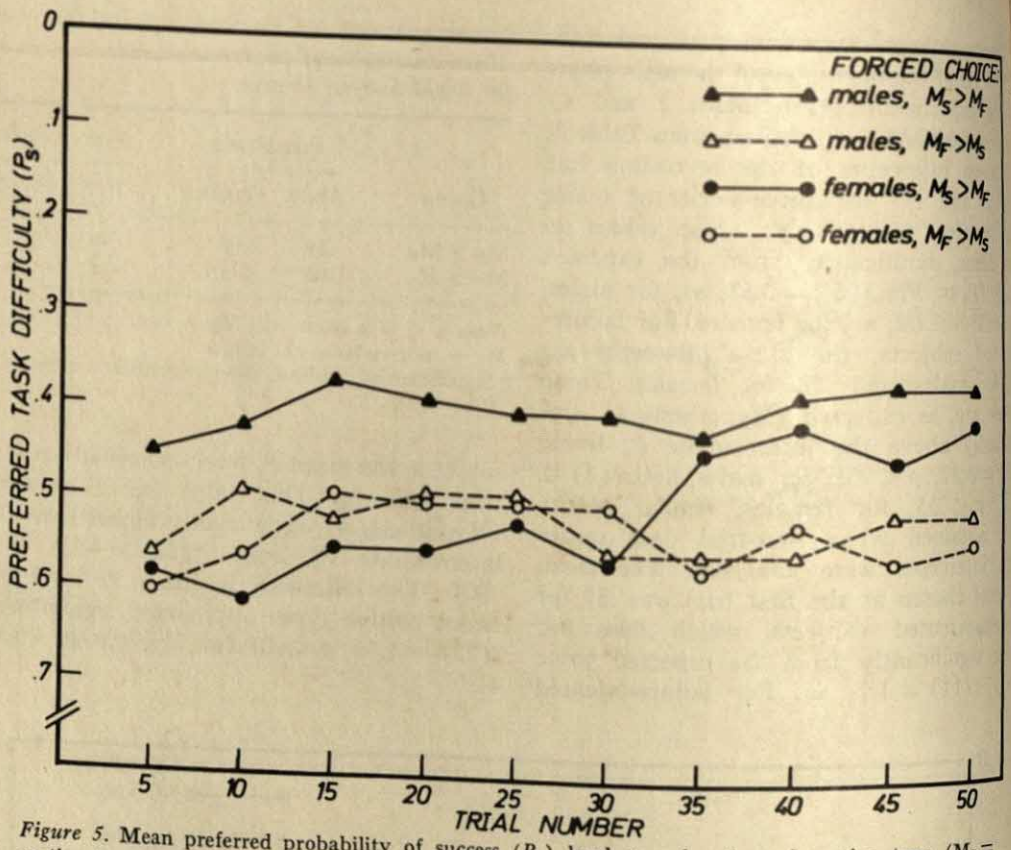


Figure 5. Mean preferred probability of success (P_s) level as a function of motive type (M_S = motive to succeed; M_F = motive to avoid failure), sex, and trials (forced choice).

Forced Choice

The mean P_s preferences reported after each block of five tasks in the forced-choice condition are shown in Figure 5 for the four groups according to motive type and sex. The mean slope of the regression line describing the linear trend in changes of risk preference over time is $M = .03$ in the success-oriented male group and $M = -.06$ in the failure-oriented male group. Neither value differs significantly from zero, $t(5) = .57$, *ns*, and $t(5) = -.64$, *ns*, respectively. The mean slopes of the regression lines in the female groups were $M = .23$ for success-oriented females and $M = .01$ for failure-oriented females. Neither value differs significantly from zero, $t(5) = .6$, *ns*, and $t(5) = .21$, *ns*, respectively. The mean intercepts do not differ significantly from $P_s = .5$ in any of the four groups of the forced-choice condition (cf. Table 3).

Discussion

The results from the free-choice condition support the predictions derived from the dynamics of action (Atkinson & Birch, 1970). If one allows subjects to actually choose their preferred difficulty levels, they show the predicted shift from intermediate (if $M_S > M_F$) or easy (if $M_F > M_S$) to more and more difficult tasks. The fact that this effect is most pronounced in the success-oriented male group and least pronounced in the success-oriented female group is what is expected when the difference between the sexes in anxiety scores is taken into account: Females scored significantly higher on the TAQ than males. Consequently, it may be assumed that the males with TAT-TAQ rank above the median are more characteristic of high n Ach, low-anxiety individuals and that the females with TAT-TAQ rank below the median are more characteristic of low n Ach, high-anxiety indi-

viduals. Females with TAT-TAQ rank above the median may be assumed to have low motivation ($M_S \approx M_F$), since their positive motivation may be dampened by negative motivation, which is significantly greater for females than for males. According to this argument, the insignificant trend obtained by success-oriented females is in line with Hypothesis 5.

The results suggest that there are factors influencing temporal change in risk preference that cannot be explained by changes in P_S. The important implication of this conclusion is that the shift to harder tasks on repeated trials cannot be explained by traditional episodic theories of achievement motivation (Atkinson, 1957; Heckhausen, 1977; Weiner et al., 1971), because subjective probability of success (P_s) was held constant throughout the main phase of the experiment. This is suggested by average P_s judgments obtained at the end of the experiment (Table 1). In the free-choice condition, none of the reported P_s values measured at the end of the experiment differs significantly from the average induced value of .50. This is what was expected on the basis of the assumption that P_s values were stable over trials during the main phase. The correspondence between *objective* initial p_s values and reported final P_s values does not, however, exclude initial discrepancies between objective and subjective probabilities, nor does it exclude increments of latent (not reported) probabilities.

We are facing an old and fundamental methodological problem here: Repeated measurements of subjects' reports do not solve the problem, because they may constitute reactive conditions destroying the validity of the conclusions. The application of objective methods for assessing variables requires, in most cases, more knowledge about the relationship between objective indexes and their subjective determinants than we actually have. Accordingly, our method of checking the assumption of stable P_s values can be viewed as a preliminary compromise only. Yet, the results (Table 1) do show that one possible attempt to falsify the assumption of stable P_s values failed. The conclusiveness of future studies attempting to check hypotheses like ours may be increased if we succeed in developing objective (rather than self-report) techniques of assessing subjective probabilities. A measure of decision time has already yielded some promising results in this direction (Schneider, 1973).

In spite of the encouraging results discussed so far, the validity and generality of the conclusions still have to be investigated in further studies. Results obtained from similar experiments with different tasks and different types of subjects are consistent, though, with the present data in suggesting that there are reasons for the tendency to choose increasingly more difficult tasks that cannot be explained on the basis of respective shifts in expectancy of success alone (Schneider & Posse, 1978; Kuhl, Note 2). These studies also demonstrate that the obtained results cannot be explained by effects due to the spatial arrangement or order of presentation of the various difficulty levels. The explanation offered by the dynamics of action is based on the assumption of greater consummatory value of success than of failure, which leads to relatively quick reduction of tendencies to work on easy tasks because success is more likely than with difficult tasks.

The shift to harder tasks in spite of constant P_s was also predicted by an application of some concepts from the dynamics of action by Revelle and Michaels's theory (1976). This theory, however, sacrifices some of the predictive power of the original theory (Atkinson & Birch, 1970) by neglecting the motivational effects of success and failure on negaction tendencies in failure-oriented subjects. The fact that failure-oriented subjects show a significant though weaker trend to more difficult tasks contradicts Revelle and Michaels's reverse-strategy hypothesis for this motive type. In addition, the limited dynamic theory includes the assumption of constantly active consummation of all competing tendencies, no matter whether a tendency is expressed in behavior. This simplifying assumption has the implication that the theory can only be tested in a situation in which constantly present consummation of all tendencies is approximated by forcing the subjects to work on consecutive series consisting of tasks of all difficulty levels (Revelle, Note 3). The significant interaction between choice

condition and trial effect (compare Figures 2 and 3 with Figure 4) stresses the predictive power of the original theory, which predicted the differential effects in both the free-choice and the forced-choice conditions. The generality of the simplified dynamic theory is limited to the forced-choice situation. The main focus of the dynamics of action, however, is directed towards the "ecologically more representative" (Brunswick, 1952) situations allowing free expression of preferred tendencies over extended periods of time.

A final word may be said about a *phenomenological* interpretation of the results. The shift to more difficult tasks in spite of constant P_s may be described in terms of a "cognitive strategy" involving a shift to a higher difficulty level after one or more successes on a given difficulty level have been experienced. Cognitive interpretations of behavioral data have been contrasted to "mechanistic" explanations (i.e., explanations that do not describe cognitive mediators explicitly; Weiner, 1972). According to our view regarding this controversy, the difference between a phenomenological and a *functional* explanation, as we prefer to call the one offered in this article, is not a difference between contradicting theoretical positions, but a difference between two epistemological perspectives on the same phenomenon. Accordingly, we believe that the functional explanation offered by the dynamics of action in terms of differential consummatory values after success and failure is compatible with a phenomenological explanation describing the cognitive processes that may be coordinated to the functional parameters discussed here.

Reference Notes

1. Bongort, K. *Most recent revision of computer program for dynamics of action.* Unpublished program, University of Michigan, 1975.
2. Kuhl, J. *Zeittrends bei der Schwierigkeitswahl von Anagrammaufgaben.* Unpublished manuscript, Ruhr University, Bochum, West Germany, 1978.
3. Revelle, W. Personal communication, 1976.

References

Atkinson, J. W. Motivational determinants of risk-taking behavior. *Psychological Review,* 1957, *64,* 359–372.

Atkinson, J. W. (Ed.). *Motives in fantasy, action, and society.* Princeton, N.J.: Van Nostrand, 1958.

Atkinson, J. W., Bastian, J. R., Earl, R. W., & Litwin, G. H. The achievement motive, goal setting, and probability preferences. *Journal of Abnormal and Social Psychology,* 1960, *60,* 27–36.

Atkinson, J. W., & Birch, D. *The dynamics of action.* New York: Wiley, 1970.

Atkinson, J. W., & Birch, D. The dynamics of achievement-oriented activity. In J. W. Atkinson & J. O. Raynor (Eds.), *Motivation and achievement.* Washington, D.C.: Winston & Sons, 1974.

Atkinson, J. W., Bongort, K., & Price, L. H. Explorations using computer simulation to simulate thematic apperceptive measurement of motivation. *Motivation and Emotion,* 1977, *1,* 1–27.

Atkinson, J. W., & Feather, N. T. (Eds.). *A theory of achievement motivation.* New York: Wiley, 1966.

Brunswick, E. *The conceptual framework of psychology.* Chicago: University of Chicago Press, 1952.

Feather, N. T. Persistence at a difficult task with alternative task of intermediate difficulty. *Journal of Abnormal and Social Psychology,* 1963, *66,* 604–609.

Heckhausen, H. Intervening cognitions in motivation. In D. E. Berlyne & K. B. Madsen (Eds.), *Pleasure, reward, preference.* New York: Academic Press, 1973.

Heckhausen, H. Achievement motivation and its constructs: A cognitive model. *Motivation and Emotion,* 1977, *1,* 283–329.

Horner, M. The measurement and behavioral implications of fear of success in women. In J. W. Atkinson & J. O. Raynor (Eds.), *Motivation and achievement.* Washington, D.C.: Winston & Sons, 1974.

Kuhl, J. *Mess- und prozesstheoretische Analysen einiger Person- und Situationsparameter der Leistungsmotivation.* Bonn, West Germany: Bouvier, 1977.

Kuhl, J., & Blankenship, V. The dynamic theory of achievement motivation: From episodic to dynamic thinking. *Psychological Review,* in press.

Mandler, G., & Cowen, J. E. Test anxiety questionnaires. *Journal of Consulting Psychology,* 1958, *22,* 228–229.

Mandler, G., & Sarason, S. B. A study of anxiety and learning. *Journal of Abnormal and Social Psychology,* 1952, *47,* 166–173.

McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

Raynor, J. O. Future orientation and motivation of immediate activity: An elaboration of the theory of achievement motivation. *Psychological Review,* 1969, *76,* 606–610.

Revelle, W., & Michaels, E. J. The theory of achievement motivation revisited: The implications of inertial tendencies. *Psychological Review,* 1976, *83,* 394–404.

Schneider, K. *Motivation unter Erfolgsrisiko*. Göttingen, West Germany: Hogrefe, 1973.

Schneider, K., & Posse, N. Subjektive Unsicherheit, Kausalattribuierungen und Aufgabenwahl. *Zeitschrift für experimentelle und angewandte Psychologie*, 1978, *25*, 302–320.

Weiner, B. Effects of unsatisfied achievement motivation on persistence and subsequent performance. *Journal of Personality*, 1965, *33*, 428–442.

Weiner, B. *Theories of motivation*. Chicago: Rand McNally, 1972.

Weiner, B., Frieze, I., Kukla, A., Reed, L., Rest, S., & Rosenbaum, R. M. *Perceiving the causes of success and failure*. New York: General Learning Press, 1971.

Received February 6, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 4, 564-571

Effects of Paced Respiration and Expectations on Physiological and Psychological Responses to Threat

Kevin D. McCaul, Sheldon Solomon, and David S. Holmes
University of Kansas

While waiting to receive electric shocks, 105 males either (a) regulated their breathing at one half the normal rate, (b) regulated their breathing at the normal rate, or (c) did not regulate their breathing rate. Half of the subjects in each breathing condition were told that their breathing task would aid them in relaxing, whereas the other half were not given that expectation. Subjects in a no-threat condition were not threatened with shocks, did not regulate their breathing, and were not provided with expectations. The results indicated that slowing respiration rate reduced physiological arousal as measured by skin resistance and finger pulse volume (but not heart rate) and reduced self-reports of anxiety. Expectations did not influence arousal. These data provide evidence for the effectiveness of paced respiration as a coping strategy, and they resolve the conflicting findings of previous investigations.

Considerable attention is being devoted to determining the effectiveness of various strategies for controlling the physiological arousal and anxiety experienced during threatening situations. One such strategy involves the control of respiration. Controlling respiration may be effective for reducing physiological arousal because there are strong general relationships among the various physiological response systems (cf. Davies & Neilson, 1967; Levenson, 1976; Obrist, Webb, & Sutterer, 1969; Stroufe, 1971). Control over respiration may also be effective for reducing anxiety, because if it does reduce physiological arousal, there would be fewer cues to stimulate self-reported anxiety (Schachter, 1964).

Surprisingly, only two experiments have been conducted specifically to test the effects of controlled respiration on persons' responses to threat. In the first investigation (Harris, Katkin, Lick, & Habberfield, 1976), subjects participated in either (a) a *paced-respiration* condition in which they were instructed to synchronize their respiration to a light that came on and went off at a rate of 8 cycles per minute, a rate half that of the normal resting respiration rate; (b) an *attention* control condition in which they simply counted the light cycles and signaled the experimenter after every 10 cycles; or (c) a *baseline* control condition in which they were not given any instructions concerning the flashing light. All subjects were then presented with two tones that they were told might be followed by electrical shocks (the first one was). The investigators reported that when the first tone was presented, subjects in the paced-respiration condition evidenced smaller electrodermal responses (though not smaller heart rate responses) than did subjects in the other two conditions. Regrettably, these results must be interpreted with caution, because the investigation suffered from a number of methodological and statistical problems (see Holmes, McCaul, & Solomon, 1978).

In the second investigation (Holmes et al., 1978) subjects were first asked to rest for a

This research was supported by biomedical research and general research fund grants from the University of Kansas to David S. Holmes. The authors would like to thank Rusty Brown and B. Kent Houston for their assistance with the project.

Requests for reprints should be sent to Kevin D. McCaul, who is now at the Department of Psychology, North Dakota State University, Fargo, North Dakota 58102.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3704-564$00.75

30-min. period during which their respiratory patterns were recorded. Subjects then participated in either (a) a *respiration-tracing* condition in which they were shown a record of their resting respiratory patterns and were instructed to breathe in such a way that the pen tracing their present respiratory pattern traced the line representing their resting respiratory pattern—a procedure that required subjects to duplicate in all respects (rate, depth, and phasing of rate and depth) their resting respiratory activity; (b) an *attention-tracing* control condition in which they were shown a record of their resting respiratory patterns and were instructed to manipulate a knob in such a way that a pen connected to the knob traced the line representing their resting respiratory pattern; or (c) a *no-tracing* control condition in which they were not shown a record of their resting respiratory pattern and simply sat quietly. Once the respiration manipulation was completed, half of the subjects in each condition were threatened with electrical shocks, whereas the other half were not. Contrary to what was expected, analyses of heart rate and self-report data indicated that the respiration manipulation was not effective for decreasing arousal. (Unfortunately, apparently because of technical problems with the equipment, in this experiment the skin resistance measure was insensitive to changes in arousal and thus could not be used for testing the hypothesis.) In contrast to the first experiment, then, the second experiment offered no evidence for the effectiveness of controlled respiration for reducing arousal under threatening conditions.

Apart from their results, the two experiments described above differed in a number of important respects; for example, (a) one experiment employed a respiration rate that was one half the normal resting rate, whereas the other employed the normal resting rate; (b) respiration was controlled with different techniques; and (c) one experiment found the predicted effects only with the skin resistance measure, whereas that measure was not available in the other. Because of these differences, it did not seem possible to resolve the conflict in the results of the two experiments, and thus it was not possible to draw a firm conclusion concerning the effectiveness of controlled respiration for reducing arousal under threatening conditions.

The present experiment was conducted in an attempt to resolve the conflicts between the two previous experiments. Because it was felt that *slowed* respiration might be the crucial difference between the earlier experiments, in the present experiment subjects were threatened with painful electrical shocks while (a) their respiration rates were regulated at one half of the normal resting rate (*slow-breathing* condition); (b) their respiration rates were regulated at the normal resting rate (*normal-breathing* condition); or (c) their respiration rates were not regulated (*nonregulated-breathing* condition). It was predicted that subjects who had their breathing regulated at a slow rate would evidence less physiological arousal and report less anxiety than subjects who had their breathing regulated at a normal rate or subjects who did not have their breathing regulated. In addition, it was expected that breathing at a normal rate would not reduce arousal or anxiety relative to breathing at a nonregulated rate.

Although respiration *rate* appears to represent the chief difference between Harris et al. (1976) and Holmes et al. (1978), it is conceivable that the respiration manipulations used in those experiments also influenced subjects' respiration *depth*. Unfortunately, data concerning respiration depth were not reported in either of those earlier studies. This issue is important because of the results of several studies that have examined the influence of paced respiration in nonthreatening situations. In one of those studies it was concluded that respiration rate per se influenced heart rate (Levenson, 1976), but other investigations suggested either that rate has no effects (Engel & Chism, 1967; Epstein & Webster, 1975) or that it is the combination of changes in rate and depth that influences measures of arousal (Deane, 1965; Stroufe, 1971). Thus, even though the experimental manipulation employed in the present investigation was focused on respiration rate, both rate and depth were monitored.

In addition to examining respiration control, this experiment also investigated the possible effects of the expectations subjects hold about the coping strategy of respiration control. Because expectations concerning the effects of the breathing manipulations might have an influence on arousal (i.e., a placebo-like effect), half of the subjects were given explicit expectations that what they were doing would aid them in reducing arousal (*high-expectations* condition), whereas the other half were not given those expectations (*no-expectations* condition).

So that the influence of the threat, respiration, and expectation manipulations could be assessed relative to a nonthreatened group whose respiration was not regulated, a control condition was employed in which subjects were not threatened, breathing was not regulated, and expectations were not provided (*no-threat* control condition).

Method

Design

This experiment employed a 3 (slow breathing, normal breathing, nonregulated breathing) × 2 (high expectations, no expectations) factorial design in which all subjects were exposed to a threat, plus a one-cell control condition (nonregulated breathing, no expectations) in which subjects were not exposed to a threat.

Subjects and Laboratory Facilities

One hundred and ten male undergraduates at the University of Kansas participated individually in the experiment as part of their research requirement for a general psychology course.[1] The subjects were assigned randomly to the seven conditions formed by the 3 × 2 + 1 design. One subject's data were discarded because he did not understand the procedures, and 4 of the 93 subjects assigned to the threat conditions elected not to continue in the experiment. Therefore, the results are based on 105 subjects and cell *n*s range from 14 to 16.

While participating in the experiment, the subject sat at a desk in a straight-backed chair in a sound-attenuated room. A video monitor with a 30-cm screen (measured diagonally) was mounted at eye level 90 cm in front of the subject. The screen was vertically divided in half by a red stripe 2 cm wide. A control room that directly adjoined the experimental room contained a Beckman polygraph, a Sony video system, and a tape recorder that were used to record physiological responses, present visual materials, and present the instructions.

Procedure

Upon arrival, the subject was told that the experiment involved studying physiological responses to various types of stimulation. After the subject agreed to participate by signing an informed-consent statement, he was seated, and the physiological sensors were attached. Specifically, a Beckman photoelectric plethysmographic finger transducer was attached to the middle finger of the subject's nondominant hand; 20 × 25 mm monopolar Lafayette chrome skin resistance electrodes were attached to the second and fourth fingers of the same hand; and a Beckman strain gauge was placed around the subject's chest.[2] In addition, a large electrode was strapped to the subject's left wrist. That electrode was later described as a shock electrode to the subjects in the threat conditions.

The subject was then asked to sit quietly and relax for 5 min. After 5 min., the subject was instructed to complete an 18-item anxiety checklist that used adjectives taken from the Affect Adjective Checklist (Zuckerman, 1960). For each adjective, the subject checked a 4-point scale (1 = "not at all"; 4 = "a great deal") that reflected the degree to which that adjective described how he felt at that time. Responses to this checklist were used as a measure of the subject's initial level of anxiety.

Threat manipulation. After the subject completed the anxiety checklist, the experimenter reentered the room and gave the subject a second informed-consent statement. The form given to subjects in the threat conditions indicated that the stimulation being studied was "painful electric shock" and then asked the subject to rate on a 61-point scale (1 = "not at all"; 61 = "extremely") how concerned he was about receiving shocks. After the subject agreed to continue participating and indicated his concern about receiving shocks, the experimenter left the room and started the audiotape that described the shock procedure. The description indicated that after the instructions were completed, a green signal light would come on, indicating the beginning of a 90-sec "waiting (anticipation) period" to be followed by the onset of a red signal light indicating the beginning of a 90-sec "shock (threat) period," during which up to three shocks would be administered on a random schedule through the electrode on the subject's left wrist.

---

[1] Male subjects were used because of their availability in the subject pool. It might be noted that in the previous, similar experiment (Holmes et al., 1978), no differences between males and females were found on either physiological or self-report measures. Therefore, there was little reason to expect differences in this experiment.

[2] Skin resistance was recorded using constant voltage (see Venables & Christie, 1973).

In contrast, the second informed-consent statement given to each subject in the no-threat condition indicated that the stimulation being studied was "visual stimulation." After the subject agreed to continue participating, the experimenter left the room and started an audiotape that described the procedure for receiving the visual stimulation. The description indicated that after the instructions were completed, a green signal light would come on, indicating the beginning of a 90-sec "waiting period" to be followed by the onset of a red signal light that would provide 90 sec of visual stimulation.

Breathing manipulation. After receiving the instructions for the shock or visual stimulation, the subject was directed to look at the video monitor mounted in front of him. In the slow-breathing and normal-breathing conditions, a Sony video recorder was used to project a sine curve that moved from right to left across the monitor. The subject was instructed to watch the curving line as it passed the red line that divided the screen and to breathe so that he inhaled as the line went up and exhaled as the line went down. In addition, the subject was admonished not to hold his breath or to take huge breaths, but to breathe slowly and smoothly throughout each curve.

In the slow-breathing condition, the videotape was recorded to regulate the subject's respiration at a rate of approximately 8 breaths per minute. This is one half the normal rate, and it is the rate that previous investigators have employed in studying the effects of slowed breathing on arousal (see Harris et al., 1976). In the normal-breathing condition, the videotape was recorded to regulate the subject's respiration at a rate of approximately 16 breaths per minute. This rate was selected because previous research indicated that this was the rate at which subjects in similar experiments breathed when not under stress (see Harris et al., 1976; Holmes et al., 1978). Each subject in the nonregulated-breathing conditions was simply instructed to watch a test pattern on the video monitor. In all conditions, the subject was instructed to continue the task assigned to him until the end of the experiment.

Expectation manipulation. Half of the subjects in the threat condition were assigned to the high-expectations condition. After practicing the breathing task or watching the test pattern for approximately 2 min., they were told that their task was effective for reducing stress; that is, the subject was told either that regulated breathing or attending to a video monitor was an exercise designed to help people relax and stay calm. Furthermore, each of these subjects was informed that the exercise had been found to be effective for reducing stress in many previous experiments.

The other half of the subjects in the threat condition, as well as subjects in the no-threat condition, were assigned to the no-expectations condition. They were told either that regulating breathing or that attending to the video monitor was desirable in experiments using physiological measures because it kept their breathing or attention constant while allowing other physiological indexes like heart rate and skin resistance to vary. They were thus given no particular expectation that the tasks would be useful for reducing stress. After the expectation manipulation was completed, each subject in the threat condition was briefly interrupted and asked to report on a single-item questionnaire how relaxed he expected to be during the shock period. Responses to this item served as a check on the expectation manipulation.

When those procedures were completed, each subject was reminded to continue performing his specific task during the wait and during the shock/stimulation periods. The lights then came on as announced, but no shocks were actually administered. Following the 90-sec shock/stimulation period, each subject was asked to complete another anxiety checklist describing how he felt during the shock/stimulation period. Each subject in the threat condition was also asked to describe his thoughts and feelings while waiting for the shock and to report what he thought about to keep calm during the shock period. The session then concluded, and the subject was completely debriefed concerning the procedures, deception, and purpose of the experiment.

Results and Discussion

Scoring and Preparation of Physiological Data

The skin resistance data were scored for seven 30-sec periods: The highest level of resistance (lowest arousal) was determined for the last 30-sec period of the baseline period that preceded any of the experimental manipulations, and the lowest level of resistance (highest level of arousal) was determined for each of the three 30-sec periods in the anticipation period and each of the 30-sec periods in the threat period (or "visual stimulation" period, in the case of subjects in the no-threat condition).[3] To eliminate the influence of initial levels of arousal on subsequent levels of arousal (i.e., the "law of initial values"; Lacey, 1956; Wilder, 1962), the scores from the anticipation and threat periods were residualized (Cronbach & Furby, 1970) using the scores from the baseline period as the measures of initial levels of

[3] The highest skin resistance level (lowest arousal) was chosen as the best initial measure of skin resistance, to avoid taking into account brief fluctuations in the final 30 sec of the baseline procedure. This procedure appears to have been effective, since the average pretest–posttest correlation for skin resistance level was greater than .90.

arousal. (A residualized score consists of the difference between the score obtained during a given period and the score predicted by linear regression from the baseline score.) The scores for the periods within the anticipation period were averaged to obtain one score for the anticipation period, and the same was done for the threat period.[4]

Heart rate and finger pulse volume were scored for the same seven 30-sec periods. Heart rate was defined as the number of beats occurring in the period. Finger pulse volume (in mm) was determined by averaging the largest volume (lowest arousal) and the smallest volume (highest arousal) during each period. Both the heart rate and finger pulse volume scores were then residualized and averaged like the skin resistance scores.

Respiration rate was scored for seven 30-sec periods: the 30-sec period after the breathing manipulation, the three 30-sec periods during the anticipation period, and the three 30-sec periods within the threat period. The scores within the anticipation and threat periods were then averaged. Respiration rate was simply defined as the total number of breaths taken during the 30 sec. Respiration depth, measured (in mm) from the bottom to the top of an inspiration, was scored for the breath approximately in the middle of each of the same seven periods, as well as the baseline period. These scores were then adjusted and averaged like the skin resistance scores.

Effectiveness of Experimental Manipulations

Threat manipulation. Comparisons of the skin resistance scores of the subjects in the six threat conditions with the skin resistance scores of the subjects in the no-threat condition revealed that across the anticipation and threat periods, subjects in the threat conditions showed lower resistance (higher arousal) than subjects in the no-threat condition, $F(1, 92) = 7.07$, $MS_e = 8057.73$, $p < .01$.[5] Comparable comparisons conducted on the heart rate data revealed that across periods, subjects in the threat conditions evidenced higher heart rates than subjects in the no-threat condition, $F(1, 97) = 6.38$, $MS_e = 18.28$, $p = .01$. Analyses of the finger pulse volume data showed similar results, $F(1, 97) = 47.25$, $MS_e = 22.40$, $p < .01$. Finally, it was found that subjects in the six threat conditions reported significantly higher anxiety than did subjects in the no-threat condition, $F(1, 96) = 20.89$, $MS_e = 68.00$, $p < .01$. These results consistently indicate that the threat manipulation was effective in increasing subjects' physiological arousal and reported anxiety.

Respiration manipulation. A 3 (respiration) × 2 (expectations) × 3 (periods) analysis of variance with repeated measures conducted on the respiration rate scores revealed that only the respiration effect was statistically significant, $F(2, 80) = 114.71$, $MS_e = 3.46$, $p < .01$. Subsequent comparisons indicated that subjects in the slow-breathing condition showed a slower respiration rate than subjects in either the normal-breathing or nonregulated-breathing conditions, $ts(120) = 9.92$ and 10.83, respectively, $ps < .01$. There was no difference in respiration rate between the latter conditions ($Ms = 4.31$, 7.75, and 8.17, respectively).

A 3 (respiration) × 2 (expectations) × 3 (periods) analysis of variance with repeated measures conducted on the respiration depth scores revealed that only the respiration effect was statistically significant, $F(2, 82) = 11.02$, $MS_e = 204.73$, $p < .01$. Subsequent comparisons (Duncan's) indicated that subjects in the slow-breathing condition breathed more

---

[4] Initial analyses indicated that there were no pretest differences among conditions on any of the measures used as covariates. Other analyses indicated that the consideration of 30-sec periods within the anticipation and stress periods contributed no new information; therefore, for both heart rate and skin resistance, the factor representing 30-sec periods was eliminated to make the analyses less complex and to reduce the possibility of chance findings.

[5] Degrees of freedom for the analyses to be presented differ slightly for two reasons. First, six subjects were missing skin resistance scores, and one subject was missing heart rate and finger pulse volume scores. Second, in analyzing the self-report measures, a covariate (initial fear of shock) was used that was not gathered for no-threat subjects. Thus, the overall error term derived from all subjects could not be used for self-report measures. Comparisons between experimental and control subjects on the anxiety checklist were performed using only one covariate: initial anxiety checklist scores.

deeply than subjects in the other respiration conditions ($ps < .05$). Respiration depth for subjects in the normal and nonregulated respiration conditions did not differ ($Ms = 25.63$, 19.94, and 16.39, respectively).

The findings concerning respiration indicate that the respiration manipulation was effective for slowing subjects' respiration rate. At the same time, however, the manipulation affected respiration depth; subjects who were induced to breathe slowly took deeper breaths. Furthermore, since none of the other effects generated by these analyses was statistically significant, it appears that respiration was not influenced by the other manipulations and that subjects maintained their respiration rate and depth over time.

Expectations manipulation. A 3 (respiration) × 2 (expectations) analysis of covariance was conducted on the scores indicating how relaxed subjects expected to be during the threat period; subjects' initial levels on the fear of shock and the anxiety checklist measures were used as covariates. This analysis revealed that only the expectations effect was statistically significant, $F(1, 81) = 5.29$, $MS_e = 99.27$, $p = .02$. Inspection of the adjusted means indicated that subjects in the high-expectations condition reported expecting to be more relaxed than subjects in the no-expectations condition ($Ms = 29.31$ and 24.33, respectively). The fact that none of the other effects generated by the analysis was statistically significant indicates that subjects' expectations were not influenced by the respiration manipulation.

Summary. The results clearly indicate that (a) the threat manipulation was effective for increasing subjects' physiological arousal and reported anxiety, (b) the respiration manipulation was effective for slowing subjects' respiration rate, and (c) the expectation manipulation was effective for increasing the degree to which subjects thought they would be able to relax while waiting for shocks. In view of the effectiveness of the various experimental manipulations, it is appropriate to consider whether slowed respiration and expectations were effective in reducing the arousal and anxiety of subjects undergoing threat.

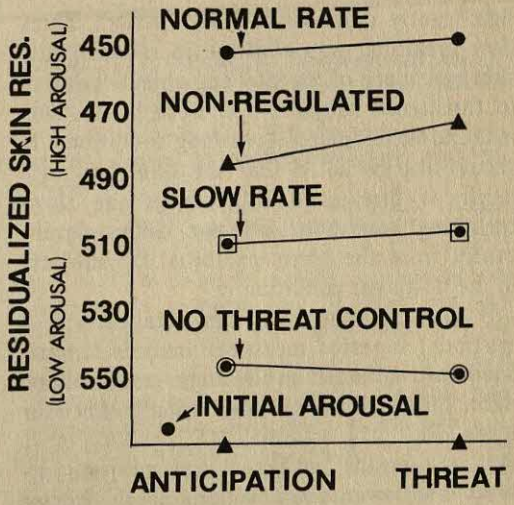


Figure 1. Mean residualized skin resistance scores for the breathing conditions and the no-threat control condition during the anticipation and threat periods. (Cell means are based on the following number of subjects: no-threat control, $n = 14$; slow breathing, $n = 29$; nonregulated breathing, $n = 28$; normal breathing, $n = 28$.)

Effects of Slowed Respiration

Physiological responses. A 3 (slow breathing, normal breathing, nonregulated breathing) × 2 (high expectations, no expectations) × 2 (anticipation period, threat period) analysis of variance with repeated measures conducted on the skin resistance data revealed a respiration effect that closely approached significance, $F(2, 92) = 2.87$, $MS_e = 15827.44$, $p = .06$, and a Respiration × Periods interaction that was statistically significant, $F(2, 92) = 3.48$, $MS_e = 288.02$, $p = .03$. The means contributing to this interaction are presented graphically in Figure 1.

To test the prediction that slowed respiration would be effective for reducing physiological arousal, a contrast was performed comparing the skin resistance level of subjects in the slow-breathing condition to the skin resistance level of subjects in the combined normal and nonregulated conditions. This contrast was significant for both the anticipation period, $F(1, 92) = 3.95$, $MS_e = 8057.73$, $p = .05$, and the threat period, $F(1, 92) = 5.42$, $MS_e = 8057.73$, $p = .02$. The skin resistance levels of subjects in the normal and nonregulated conditions were not

significantly different in either period. Thus, slow breathing was effective in reducing the skin resistance of threatened subjects relative to threatened subjects who were in the normal and nonregulated breathing conditions. It should also be noted that the skin resistance levels of threatened subjects in the slow-breathing condition did not differ significantly from the levels exhibited by subjects in the no-threat condition.

A 3 (respiration) × 2 (expectations) × 2 (periods) repeated measures analysis of variance was conducted on the finger pulse volume data. This analysis revealed a main effect for trials, $F(1, 97) = 8.65$, $MS_e = 3.42$, $p < .01$; the means indicated that subjects exhibited a lower pulse volume level (higher arousal) during the threat than during the anticipation period. The analysis also revealed a significant effect for respiration, $F(2, 97) = 3.30$, $MS_e = 41.39$, $p = .04$.

To test the prediction that slowed respiration would be effective for reducing physiological arousal, a contrast was performed comparing the finger pulse volume level of subjects in the slow-breathing condition to the finger pulse volume level of subjects in the combined normal and nonregulated conditions. Across periods this contrast was significant, $F(1, 97) = 5.43$, $MS_e = 22.40$, $p = .02$. The finger pulse volume levels of subjects in the normal and nonregulated conditions were not significantly different (Ms = 9.22, 6.22, and 7.23, respectively). Thus, slow breathing was effective in reducing the finger pulse volume of threatened subjects relative to threatened subjects who were in the normal and nonregulated breathing conditions. It should be noted, however, that subjects in the slow-breathing condition exhibited more arousal than subjects who were not threatened, $F(1, 97) = 25.81$, $MS_e = 22.40$, $p < .01$.

Analyses conducted on the heart rate data did not reveal any statistically significant main effects or interactions involving respiration.

Self-report responses. A 3 (respiration) × 2 (expectations) analysis of covariance was conducted on the anxiety scores. Subjects' scores on both the self-report measure of initial fear of shock and on the first anxiety checklist were used as covariates. The analysis of covariance revealed only a respiration effect that approached statistical significance, $F(2, 81) = 2.54$, $MS_e = 65.76$, $p = .08$. To test the prediction that slowed respiration would be effective for reducing psychological anxiety, a contrast was performed comparing the self-reported anxiety of subjects in the slow-breathing condition to the anxiety reported by subjects in the combined normal and nonregulated breathing conditions. This contrast was statistically significant, $F(1, 81) = 4.89$, $MS_e = 65.76$, $p = .03$. A comparison of the anxiety reported by subjects in the normal-breathing and nonregulated-breathing conditions was not significant (Ms = .75, 5.37, and 4.26, respectively). These data suggest that slow breathing was a successful technique for reducing psychological stress. It should be noted, however, that subjects in the slow-breathing condition reported more anxiety than did subjects who were not threatened with shock, $F(1, 96) = 10.96$, $MS_e = 68.00$, $p < .01$.

Effects of Expectations

The factorial analyses of variance, described above, that were conducted on the skin resistance, finger pulse volume, heart rate, and self-report data did not reveal any main effects or interactions involving expectations that reached or approached statistical significance. The consistent absence of those effects and the absence of meaningful ordering of means clearly indicates that in this investigation, differential expectations concerning the effectiveness of the coping strategy subjects were employing did not influence the actual effectiveness of the strategies.

Conclusions and Implications

This experiment provides support for the conclusions of Harris et al. (1976) that paced respiration can be effective for reducing arousal under threatening conditions. The data also suggest that the conclusions of Holmes et al. (1978) that controlled respiration is ineffective for reducing arousal must be qualified. Specifically, when respiration was controlled by slowing it to a rate well below normal, as in Harris et al., subjects evidenced

higher skin resistance and higher finger pulse volume levels, and they reported lower levels of subjective anxiety than did subjects in the other conditions. In contrast, when respiration was controlled at a normal, resting rate, as in Holmes et al., no beneficial effects for controlled respiration were found. The similarity of these results across both physiological and self-report measures provides converging validity for the effects of the slow-breathing strategy.

Despite the strength of these results, it is important to recognize two points. First, on both the self-report measure of anxiety and on the finger pulse volume measure, subjects who were threatened and who used slow breathing exhibited greater arousal than non-threatened subjects. Clearly, by itself, the slowed-breathing strategy is not a panacea for problems of arousal.

Second, the data concerning respiration demonstrated that subjects who were induced to change their respiration rate also changed their respiration depth; thus, it is not clear whether it was the change in rate, depth, or some combination of rate and depth that resulted in the reduction of arousal. From a pragmatic standpoint (e.g., using the slow-breathing strategy to reduce anxiety), it is clear that the manipulation of respiration rate influences arousal, and the question of whether rate or depth or some combination mediates these effects may not be crucial. From a theoretical standpoint, however, this question is important if we are to understand the phenomenon, and it may deserve further attention.

In summary, the present experiment emphasizes the importance of examining physiological strategies for controlling physiological and psychological stress. If social–personality psychologists are going to make a maximal contribution to the process of stress control, their efforts should include physiological components along with the cognitive components they have already studied and found to be effective.

References

Cronbach, L. J., & Furby, L. How we should measure "change"—Or should we? *Psychological Bulletin,* 1970, *74,* 68–80.

Davies, C., & Neilson, J. Sinus arrhythmia in man at rest. *Journal of Applied Physiology,* 1967, *22,* 947–955.

Deane, G. E. Cardiac rate as a function of changes in respiration. *Psychological Reports,* 1965, *16,* 41–42.

Engel, B. T., & Chism, R. A. Effect of increases and decreases in breathing rate on heart rate and finger pulse volume. *Psychophysiology,* 1967, *4,* 83–89.

Epstein, L. H., & Webster, J. S. Instructional, pacing, and feedback control of respiratory behavior. *Perceptual and Motor Skills,* 1975, *41,* 895–900.

Harris, V. A., Katkin, E. S., Lick, J. R., & Habberfield, T. Paced respiration as a technique for the modification of autonomic response to stress. *Psychophysiology,* 1976, *13,* 386–391.

Holmes, D. S., McCaul, K. D., & Solomon, S. Control of respiration as a means of controlling responses to threat. *Journal of Personality and Social Psychology,* 1978, *36,* 198–204.

Lacey, J. The evaluation of autonomic responses: Toward a general solution. *Annals of the New York Academy of Science,* 1956, *67,* 123–164.

Levenson, R. W. Feedback effects and respiratory involvement in voluntary control of heart rate. *Psychophysiology,* 1976, *13,* 108–114.

Obrist, P. A., Webb, R. A., & Sutterer, J. R. Heart rate and somatic changes during aversive conditioning and a simple reaction time task. *Psychophysiology,* 1969, *5,* 696–723.

Schachter, S. The interaction of cognitive and physiological determinants of emotional state. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 1). New York: Academic Press, 1964.

Stroufe, L. A. Effects of depth and rate of breathing on heart rate and heart rate variability. *Psychophysiology,* 1971, *8,* 648–655.

Venables, P. H., & Christie, M. J. Mechanisms, instrumentation, recording techniques, and quantification of responses. In William F. Prokasky & David C. Raskin (Eds.), *Electrodermal activity in psychological research.* New York: Academic Press, 1973.

Wilder, J. Basimetric approach (law of initial value) to biological rhythms. *Annals of the New York Academy of Science,* 1962, *98,* 1211–1220.

Zuckerman, M. The development of an affect adjective checklist for the measurement of anxiety. *Journal of Consulting Psychology,* 1960, *24,* 457–462.

Received March 20, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 4, 572–579

Interpersonal Attraction in Aversive Environments: A Problem for the Classical Conditioning Paradigm?

Douglas T. Kenrick and Gregory A. Johnson
Montana State University

Several studies have found decreased attraction for a stranger rated under aversive conditions, and these studies have been used to support a model that views attraction as a function of generalization of affect. A number of studies, however, have failed to support such a model, not only failing to show decreases in attraction as a function of negative circumstances but often showing the reverse, that is, enhanced attraction under aversive conditions. It is suggested that studies that have supported the affective generalization model may be limited in ecological validity as a function of their use of a "simulated stranger" paradigm and that results from studies in which "real" strangers are rated can be understood within a negative reinforcement model. Original data from a study in which female subjects rated both a fellow subject and a bogus stranger under conditions of either aversive noise or low-level noise are presented to support such a resolution.

Research stimulated by the reinforcement model of attraction (Byrne, 1971; Clore & Byrne, 1974; Lott & Lott, 1974) has produced an abundance of evidence supporting the idea that persons will come to like others associated with positive affect and dislike others associated with negative affect. With regard to the latter phrase of this proposition, a number of studies done to test a classical conditioning model of attraction have in fact shown that arousal of aversive affect leads to decreased interpersonal attraction (Gouaux, 1971; Griffitt, 1970; Griffitt & Veitch, 1971; Veitch & Griffitt, 1976). There is at the same time, however, a large body of seemingly contradictory findings. For instance, anecdotal evidence suggesting that individuals often become highly attracted to one another under aversive circumstances (Byrne, Allgeier, Winslow, & Buckman, 1975; Rubin, 1973, pp. 5–7) is supported by similar correlational findings (Driscoll, Davis, & Lipetz, 1972; Rubin, 1973, pp. 228–230). Likewise, a number of experimental studies of attraction under aversive circumstances (Bell, 1978; Bell & Baron, 1974, 1976; Byrne et al., 1975; Dutton & Aron, 1974; Kenrick, Cialdini, & Linder, 1979; Latané, Eckman, & Joy, 1966; Morris et al., 1976; Brehm, Gatz, Goethals, McCrimmon, & Ward, Note 1) have failed to show a classical conditioning effect and often show the opposite pattern (i.e., aversive circumstances increasing attraction).

Negative Reinforcement or Generalization of Negative Affect?

Although it is possible to interpret the latter findings as inconsistent with reinforcement principles (Berscheid & Walster, 1974), Kenrick and Cialdini (1977) suggested that these data could be parsimoniously explained as consistent with a *negative reinforcement* model. Briefly, since the presence of others has been found to reduce aversive arousal arising from such situations (Back & Bogdonoff, 1964; Bovard, 1959; Conger, Sawrey, & Turrell, 1958; Wrightsman, 1960), subjects in

The authors wish to thank Robert B. Cialdini for his comments on an earlier version of this manuscript. We would also like to express our appreciation to Robert A. Baron, whose help extended well beyond the call of duty of a topic editor.

Requests for reprints should be sent to Douglas T. Kenrick, Department of Psychology, Montana State University, Bozeman, Montana 59717.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3704-572$00.75

these experiments would be expected to find another's presence reinforcing, and hence reported attraction would be quite consistent with reinforcement principles. In the Dutton and Aron (1974) study, for instance, there was evidence that the target person did in fact reduce subjects' anxiety levels. According to this formulation, the reinforcement model of attraction is not incorrect at all. In this case, it must simply be kept in mind that positive affect is aroused not only by the onset of a rewarding stimulus (positive reinforcement) but also by the offset of a punishing stimulus (negative reinforcement).[1]

Although such an account is consistent with the reinforcement formulation of attraction (Byrne, 1971; Clore & Byrne, 1974; Lott & Lott, 1974), there remains nevertheless a disturbing inconsistency in the literature, with one group of studies suggesting that we will come to dislike others we meet under aversive circumstances (e.g., Griffitt, 1970; Griffitt & Veitch, 1971) and another group of studies contradicting this suggestion (e.g., Driscoll et al., 1972; Dutton & Aron, 1974). We would thus seem to have a situation in which reinforcement theory has little predictive utility. However, if one compares those studies supporting the *simple* affective generalization model with those which fail to support the model, one very consistent difference emerges. In those studies lending support to such a model, the target person was not physically present. In the remaining studies, on the other hand, which support an expanded negative reinforcement version of the model, the target person was actually present contiguous with the aversive stimuli.

Influence of Aversive Affect on Ratings of a "Bogus Stranger"

In the Griffitt and Veitch (1971) experiment, subjects were exposed to conditions varying from "hot and crowded" to roomy and comfortable. As predicted, ratings of strangers were negatively influenced by both uncomfortable heat and by crowding. The strangers rated, however, were not the other subjects present in the experimental chamber but were simulated strangers (Byrne, 1971). The simulated stranger in the Griffitt & Veitch (1971) experiment was described as a student of the same sex as the subject, who had participated in another study during the previous semester. The subject was shown a copy of an attitude survey ostensibly filled out by this person (in actuality, the attitude ratings were filled out by the experimenter so as to appear to be either similar or dissimilar to the subject's own attitudes). Similarly, subjects in the Griffitt (1970), Gouaux (1971), and Veitch and Griffitt (1976) studies rated such an artificially simulated stranger. Hence, when a target person is not physically present, there is reason to believe that negative affect generated by unpleasant circumstances will generalize to that other, leading to a negative impression.

Aversive Affect and Ratings of Another who is Physically Present

The Dutton and Aron (1974) and Brehm et al. (Note 1) studies, on the other hand, had subjects (who were or were not expecting painful electric shock) rate a physically present confederate who posed as a fellow subject. In these experiments, there was increased attraction for the confederate under the relatively more aversive circumstances. Byrne et al. (1975) likewise found increased attraction for same-sex others under relatively more aversive circumstances and suggested that these results "create potential problems within the larger context of the reinforcement–affect model of attraction" (Byrne et al., 1975, p. 11). Although this study also used a bogus attitude profile, however, as did studies by Bell (1978) and Bell and Baron (1974, 1976), subjects in these studies were led to believe that the profile had been filled out either by a confederate or by another subject who actually participated in the same experiment along with the sub-

[1] This term is often mistakenly interchanged with the term *punishment*. As used by contemporary learning theorists (e.g., Rachlin, 1970), however, negative reinforcement refers to the termination of an aversive stimulus, whereas punishment refers to its onset, which of course has more or less opposite consequences.

ject.[2] Similarly, Latané et al. (1966) reported highest attraction for same-sex others who were *present* under relatively more fear-producing conditions, whereas in correlational studies of the influence of adverse conditions on attraction (Driscoll et al., 1972; Rubin, 1973), subjects reported their attraction toward actual dating partners. Here again, results seem contradictory to a simple classical conditioning model, indicating increased attraction under relatively more adverse circumstances. It would seem, then, that results obtained using the bogus stranger method might not be readily generalizable to most situations involving actual *interpersonal* judgment under aversive conditions.

Validity of the Bogus Stranger Paradigm

If our reasoning up to this point is correct, the inconsistent results of those studies investigating the influence of negative situations on attraction pose a problem not for the classical conditioning model as a whole but only for the generality of one derivation from that model, which has received support and general acceptance because of the use of a particular limited methodology.

At this point it should be noted that the validity of the bogus stranger technique has been the subject of a great deal of debate in a different area of interpersonal attraction research, namely, in studying the relationship between similarity and attraction. Critics have suggested that results from such bogus stranger studies might not always generalize beyond the confines of the paradigm (Levinger, 1972; Murstein, 1971; Wright, 1971). In brief, the content of these arguments has been that results from observations of actual relationships will likely show a much weaker connection between similarity and attraction than that obtained using the laboratory paradigm. With regard to the connection between negative affect and interpersonal attraction, however, it would seem that the influence of negative environmental stimuli is not only weaker than one would expect on the basis of studies done within the bogus stranger paradigm but would actually tend to be *opposite* in its influence on actual interpersonal attraction, *even* within the confines of a short first encounter in a laboratory setting. In effect, while it may generally be profitable to view a target person as a neutral stimulus to which reinforcement value may be conditioned, that same other individual, *even if she or he is a complete stranger,* may actually take on a positive valence in threatening or aversive situations.

It is our contention that the data presented thus far seem, in themselves, to strongly support the resolution we have outlined. In order to control for the possible operation of other, perhaps less obvious, factors, however, we conducted an experiment in which subjects evaluated both bogus strangers and actual strangers under identical aversive conditions. It was predicted that contiguously presented aversive stimuli would have the effect of decreasing attraction toward the bogus stranger while increasing attraction toward a stranger who had been physically present with the subject.

Method

Subjects

Subjects were 66 undergraduate females enrolled in an introductory psychology course at Montana State University. They received credit toward their course grade for their participation. Subjects participated in groups of two. There was no theoretical reason for limiting the subjects to females. Given the necessity of having both subjects show up, females were chosen, because in our experience here they had been more reliable. Each member of a dyad came from a separate large introductory section. Four subjects (two dyads) were dropped from the experiment because they had prior acquaintance with their experimental partner, and two were dropped because of suspicion about the rated profiles. The remaining 60 subjects were equally divided into the high and low noise conditions (described below).

Procedure

Subjects arrived for an experiment on "environment and judgment." The two subjects arrived at

---

[2] Consistent with our own reasoning, Bell and Baron (1976) suggest that their failure to replicate the findings of Griffitt (1970) and Griffitt and Veitch (1971) may have been due to the fact that "any negative affect of high ambient temperature may have been neutralized by a positive feeling of 'shared suffering' " (pp. 28–29).

separate locations for any given experimental session and were brought (individually) to the experiment room. Here the experimenter introduced them, asked if they had prior acquaintance, and then explained:

> As the experimental title indicated, we're interested in the effects of environmental stimuli on judgment. We're particularly interested in the effects of environmental noise. You two will be in the high (low) noise condition.
>
> The first task you'll be working on is simply an attitude survey. We ask that you not discuss this with one another, since people are often strongly influenced by others on these things. We'll also be returning to a similar task later—so again, please do not discuss these with one another (all subjects complied with this request).

Noise manipulation. The experimenter turned on a noise tape at this point. Subjects in the aversive noise condition heard a tape that consisted of intermittent, unpredictable bursts of very loud noise—peaks were 95 dB (A)—determined to be highly aversive in preexperimental testing.[3] Subjects in the low noise condition heard the same tape played at a very low level (peaks were 32 dB).

Attraction ratings. After completing the 26-item attitude survey questionnaire, subjects were given a short filler task while the experimenter constructed bogus attitude surveys based on each subject's responses. Using a method based on Byrne's (1971) "method of constant discrepancy," a moderately similar profile was constructed for each subject (average discrepancy = 2 points on a 6-point scale). The experimenter then returned and explained:

> We'll be separating the two of you for the remainder of the experiment. I need one of you to come with me—it doesn't really matter which of you, since there's another speaker in the other room playing the same noise tape.

The experimenter always chose the subject to his right and in actuality took her to an adjoining room where the same condition noise tape was being played. After separating the subjects, the experimenter explained to each:

> This part of the experiment involves interpersonal judgment when only a small amount of information is available. You'll be rating a girl who was in a completely different experiment earlier this year and filled out an attitude survey similar to the one you've just completed. Please read the person's answers carefully and try to form an opinion about her.

The subject was given an attitude profile sheet with the name "Jean Haverland" on it. The other subject was given the same instructions, but was told that "you'll be rating (name of co-subject), the other girl in the experiment with you." Subjects were told that their ratings would remain "completely confidential" and were asked to give a "frank and honest judgment." Subjects who first rated the bogus stranger were then asked to rate the "other subject" and vice versa. In each case the subject received another profile of exactly the same degree of discrepancy, except that item directionality was varied. For instance, if the subject's responses to Items 8, 9, 10, and 11 were 3, 4, 4, and 3, respectively, one profile might read 1, 2, 6, and 5 on those items, whereas the other might read 5, 6, 2, and 1 (on sum, the mean discrepancy of each profile was identical).

For half the subjects, the "other subject" had indicated that she would shortly be leaving the state (in response to an item tapping attitudes about differential funding for the major rival branch of the state university system), whereas half of the subjects saw the same comment on the profile attributed to the bogus stranger. Order was randomized. Similar statements have been included in earlier research using this technique (Byrne, 1971) to lead the subject not to expect future interaction with the partner.

The noise tape (either aversive or control) was continued during the interpersonal judgment measures. The ratings were made on the standard Interpersonal Judgment Scale (Byrne, 1971). The "liking" and "desire to work" items were, in the typical fashion, summed to yield the major dependent variable. Following the ratings, the noise was terminated, and subjects were fully debriefed and thanked for their participation.

Predictions. It was expected that the results for the attraction measure would show an interaction of noise condition and type of stranger rated, indicating relatively more positive ratings of the other subject in the high noise condition, but an opposite effect for ratings of the simulated stranger.

Results

An initial $2 \times 2 \times 2$ analysis indicated no effect of the target person's intention to leave

---

[3] Thirty-two female subjects from the same population used in the experiment proper were exposed to either the high or low noise conditions while performing the initial task engaged in by experimental subjects (filling out the attitude questionnaire). They then responded to two questions designed to assess the aversiveness of the noise: "How annoying do you find this noise?" (0 = "not at all annoying"; 100 = "extremely annoying") and "How pleasant do you find this environment to be?" (−3 = "not at all pleasant"; +3 = "extremely pleasant"). Mean ratings on the first item were 67.87 and 31.67 for the high and low noise conditions, respectively, $F(1, 30) = 17.30$, $p < .001$. Similarly, the second item yielded mean ratings of −2.29 for the high noise and −.72 for the low noise conditions, $F(1, 30) = 14.27$, $p < .001$.

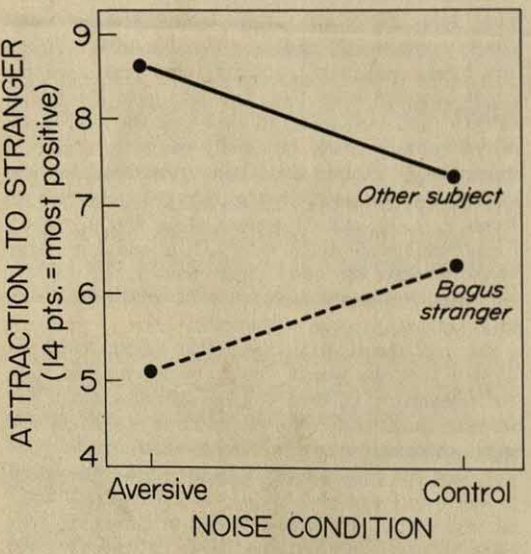


Figure 1. Attraction ratings as a function of noise condition and type of stranger rated.

the state nor any interaction of this factor with the central manipulation in the present study (all Fs < 1), and this variable will not be discussed further. Results are displayed in Figure 1 according to a 2 (between-groups : noise level) $\times$ 2 (within-groups : type of stranger) analysis of variance and indicate that our hypotheses were clearly supported, $F(1, 114) = 7.02$, $p < .01$, for this interaction. Ratings of the fictional stranger were, as predicted, significantly more negative in the aversive noise condition than in the low noise condition, $F(1, 56) = 4.80$, $p < .05$. Ratings of the other subject indicated that these were affected in the opposite direction by the noise manipulation, $F(1, 56) = 6.21$, $p < .03$. In addition, the overall main effect of the type of stranger indicated that other subjects were generally rated more positively than bogus strangers, $F(1, 114) = 62.11$, $p < .001$. The overall main effect of noise level was not significant, $F(1, 114) < 1$.

Discussion

Our results are clear in supporting the hypothesis that, under negative conditions, the bogus stranger method yields results that are contradictory to those obtained using "real" others as target persons, even in a short-term laboratory encounter. The results we obtained for bogus strangers are directly in line with those obtained in prior studies using this method (e.g., Griffitt, 1970; Griffitt & Veitch, 1971), while our results for ratings of contiguously present others are consistent with the general findings of previous studies using this method (e.g., Byrne et al., 1975; Dutton & Aron, 1974; Latané et al., 1966).

A recent paper by Rotton, Barry, Frey, & Soler (1978) reported results that are compatible with those of the present investigation. In their first experiment, these authors failed to find the predicted decreases in attraction for a stranger as a function of unpleasant odors. In this case, however, the attitude profile rated by the subject was presented as having been filled out by another subject in an adjoining room (and presumably "sharing stress" with the subject). In Rotton et al.'s (1978) second experiment, however, it was made clear to the subjects that they were alone in the situation and that the person being rated was not a fellow participant. In this study, unpleasant odor led to significant decreases in attraction for the rated other. These authors concluded that "negative affect generalized onto liking of other persons in the environment" (p. 67) but added the qualification that "air pollution reduces attraction only when a person is isolated" (p. 68). We would argue, of course, that the results of Rotton et al.'s first study are more likely to generalize to actual interpersonal encounters, since attraction would seem to be rarely developed when one is in isolation.

Thus, the use of the bogus stranger technique in the present line of investigation is even more strongly subject to criticisms of its validity than is the case in similarity–attraction research (Levinger, 1972; Murstein, 1971; Wright, 1971), where the general accuracy of the relationship, if not its intensity, seems to be ecologically valid (Byrne, Ervin, & Lamberth, 1970; Clore & Byrne, 1974).

Theoretical Considerations

While our results oppose one specific deduction from the classical conditioning model—that persons come to dislike strangers whom they meet under aversive circumstances—they

can in fact be seen to be in essential agreement with the overall reinforcement model. As we stated earlier, the failure of the simple classical conditioning model in this instance may stem from the fact that other persons function as *generalized reinforcers* under aversive circumstances (Back & Bogdonoff, 1964; Kenrick & Cialdini, 1977; Schachter, 1959; Wrightsman, 1960), and their negative reinforcement value (as potential reducers of aversive arousal) can generally be expected to neutralize or outweigh any immediate generalization of negative affect. Thus, we remain in agreement with the basic assumptions of the Clore and Byrne (1974) model, while calling into question certain lower level deductions from that model.

The present data suggest an important conceptual qualification on the simple affective generalization model, however. Certainly mere contiguity with a negatively toned context is not sufficient, in itself, to produce disliking for another. Instead, it seems necessary to consider the subject's cognitions to explain at least part of the present experiment's pattern of results. Subjects' ratings of their fellow subject showed a pattern that did not follow from simple contiguity assumptions, but instead seemed to indicate an active and discriminating attributional analysis of the situation. Such a consideration of cognitive information processing is, of course, compatible with recent statements of the reinforcement–affect model of attraction (Clore & Byrne, 1974).

Although the suggested negative reinforcement model is consistent with prior literature and seems a plausible explanation of the present results, it should be noted that we do not directly demonstrate the reduction of aversive arousal in our data. It might be the case, for instance, that attraction for the physically present other in the aversive condition was not increased because of the reduction (or even anticipated reduction) of aversive arousal, but instead because the experience of shared stress may have led to an increased perception of similarity or "unit relationship" with the fellow subject. Since the subjects had a great deal of information relating to the target person's attitudes on controversial positions, it seems to us unlikely that this one additional cognitive input would greatly alter the extent to which the target was seen as similar/dissimilar, but such an account cannot be ruled out, particularly since "shared stress" was certainly a salient issue in this case. Further research investigating the mediator of these effects would be worthwhile.

Future research is also needed to determine which components of the other subject's presence mediate the demonstrated increases in attraction under aversive circumstances. For instance, subjects in the aversive condition may have engaged in higher levels of mutual eye contact than those in the low noise condition. Research by Morris et al. (1976) supports the suggestion that aversive conditions do produce differential patterns of interaction between subjects, although the research of Rotton et al. (1978), discussed above, suggests that the simple knowledge that the other is sharing a stressful experience with oneself may sometimes be sufficient to produce attraction.

When Will Negative Affect Decrease Attraction?

The major point of these data is to suggest the severe ecological limitations of those earlier findings (e.g., Griffitt, 1970; Griffitt & Veitch, 1971), which social psychologists have widely accepted to mean that "any event that arouses a negative emotional state may create a temporary dislike for whoever is nearby" (Tedeschi & Lindskold, 1976, pp. 455–456). In assessing the effect of simple negative arousal on interpersonal attraction, results from studies using the bogus stranger paradigm would seem to explain only a rather limited range of situations (i.e., those in which one is evaluating another person who is not physically present). Whenever the object person is present, negative conditions would be expected to result in rather different effects on attraction.

It should be noted here that we are dealing only with the case of impersonal aversive or threatening environmental stimuli, from which it has been determined that individuals do

in fact seek solace in the company of others (Burlingame & Freud, 1943; Marshall, 1951; McIntyre, 1971; Sarnoff and Zimbardo, 1961; Schachter, 1959; Wrightsman, 1960). We would not, for instance, expect the present results to hold in situations involving potential interpersonal embarrassment, where the presence of others has punishment, as opposed to reinforcement, value. For instance, Sarnoff and Zimbardo (1961) found that whereas 95% of their subjects who expected painful electric shocks wished to wait in the company of others, the majority of those expecting to perform potentially embarrassing behaviors expressed a desire to wait alone. Consistently, Griffitt and Guay (1969) found that their subjects showed decreased attraction for others who were present when the subject received esteem-damaging feedback (negative creativity ratings), even when that other was not personally responsible for the feedback. On the other hand, Jacobs, Berscheid, & Walster (1971) found that females were more attracted to a male they met after receiving self-esteem-lowering feedback. In this case, however, the male was not present during the receipt of the feedback, but served to repair the subject's damaged self-esteem by expressing romantic interest in her.

Finally, it would be expected that other persons who are themselves *directly responsible* for inducing negative affect in an individual will come to be disliked, and there is an abundance of data to support this supposition (e.g., Burnstein & Worchel, 1962; Byrne & Rhamey, 1965; Griffitt & Guay, 1969; Hendrick & Taylor, 1971; Sigall & Aronson, 1969).

Reference Note

1. Brehm, J., Gatz, M., Goethals, G., McCrimmon, J., & Ward, L. *Psychological arousal and interpersonal attraction.* Unpublished manuscript, Duke University, 1967.

References

Back, K., & Bogdonoff, M. Plasma lipid responses to leadership, conformity, and deviation. In P. Leiderman & D. Shapiro (Eds.), *Psychobiological approaches to social behavior.* Stanford, Calif.: Stanford University Press, 1964.

Bell, P. A. Affective state, attraction, and affiliation: Misery loves happy company too. *Personality and Social Psychology Bulletin,* 1978, *4,* 616–619.

Bell, P. A., & Baron, R. A. Environmental influences on attraction: Effects of heat, attitude, similarity, and personal evaluations. *Bulletin of the Psychonomic Society,* 1974, *4,* 470–481.

Bell, P. A., & Baron, R. A. Aggression and heat: The mediating role of negative affect. *Journal of Applied Social Psychology,* 1976, *6,* 18–30.

Berscheid, E., & Walster, E. A little bit about love. In T. Huston (Ed.), *Foundations of interpersonal attraction.* New York: Academic Press, 1974.

Bovard, E. The effects of social stimuli on the response to stress. *Psychological Review,* 1959, *66,* 267–277.

Burlingame, D., & Freud, A. *Young children in wartime.* London: Allen & Unwin, 1943.

Burnstein, E., & Worchel, P. Arbitrariness of frustration and its consequences for aggression in a social situation. *Journal of Personality,* 1962, *30,* 528–540.

Byrne, D. *The attraction paradigm.* New York: Academic Press, 1971.

Byrne, D., Allgeier, A., Winslow, L., & Buckman, J. The situational facilitation of interpersonal attraction: A three-factor hypothesis. *Journal of Applied Social Psychology,* 1975, *5,* 1–15.

Byrne, D., Ervin, C. R., & Lamberth, J. Continuity between the experimental study of attraction and real-life computer dating. *Journal of Personality and Social Psychology,* 1970, *16,* 157–165.

Byrne, D., & Rhamey, R. Magnitude of positive and negative reinforcements as a determinant of attraction. *Journal of Personality and Social Psychology,* 1965, *2,* 884–889.

Conger, J., Sawrey, W., & Turrell, E. The role of social experience in the production of gastric ulcers in hooded rats placed in a conflict situation. *Journal of Abnormal and Social Psychology,* 1958, *57,* 214–219.

Clore, G. L., & Byrne, D. A reinforcement-affect model of attraction. In T. L. Huston (Ed.), *Foundations of interpersonal attraction.* New York: Academic Press, 1974.

Driscoll, R., Davis, K., & Lipetz, M. Parental interference and romantic love: The Romeo and Juliet effect. *Journal of Personality and Social Psychology,* 1972, *24,* 1–10.

Dutton, D., & Aron, A. Some evidence for heightened sexual attraction under conditions of high anxiety. *Journal of Personality and Social Psychology,* 1974, *30,* 510–517.

Gouaux, C. Induced affective states and interpersonal attraction. *Journal of Personality and Social Psychology,* 1971, *20,* 37–43.

Griffitt, W. Environmental effects on interpersonal affective behavior: Ambient effective temperature and attraction. *Journal of Personality and Social Psychology,* 1970, *15,* 240–244.

Griffitt, W., & Guay, P. "Object" evaluation and conditioned affect. *Journal of Experimental Research in Personality,* 1969, *4,* 1–8.

Griffitt, W., & Veitch, R. Hot and crowded: Influence of population density and temperature on interpersonal affective behavior. *Journal of Personality and Social Psychology,* 1971, *17,* 92–98.

Hendrick, C., & Taylor, S. P. Effects of belief similarity and aggression on attraction and counter-aggression. *Journal of Personality and Social Psychology,* 1971, *17,* 342–349.

Jacobs, L., Berscheid, E., & Walster, E. Self-esteem and attraction. *Journal of Personality and Social Psychology,* 1971, *17,* 84–91.

Kenrick, D. T., & Cialdini, R. B. Romantic attraction: Misattribution versus reinforcement explanations. *Journal of Personality and Social Psychology,* 1977, *35,* 381–391.

Kenrick, D. T., Cialdini, R. B., & Linder, D. L. Heterosexual attraction and attributional processes in fear-producing situations. In M. Cook & G. Wilson (Eds.), *Love and attraction.* London: Pergamon Press, 1979.

Latané, B., Eckman, J., & Joy, V. Shared stress and interpersonal attraction. *Journal of Experimental Social Psychology,* Supplement 1, 1966, 95–102.

Levinger, G. Little sand box and big quarry: Comment on Byrne's paradigmatic spade for research on interpersonal attraction. *Representative Research in Social Psychology,* 1972, *3,* 3–19.

Lott, A., & Lott, B. The role of reward in the formation of positive interpersonal attitudes. In T. Huston (Ed.), *Foundations of interpersonal attraction.* New York: Academic Press, 1974.

Marshall, S. *Men against fire.* Washington, D.C.: Combat Forces Press, 1951.

McIntyre, A. Social comparison and dependency as motives for affiliation in an anxiety-arousing situation. (Doctoral dissertation, Yale University, 1970). *Dissertation Abstracts International,* 1971, *32,* (1-B), 542. (University Microfilms No. 71-17,032)

Morris, W. N., Worchel, S., Bois, S. L., Pearson, J. A., Rountree, C. A., Samaha, G. M., Wachtler, J., & Wright, S. L. Collective coping with stress: Group reactions to fear, anxiety, and ambiguity. *Journal of Personality and Social Psychology,* 1976, *33,* 674–679.

Murstein, B. I. Critique of models of dyadic attraction. In B. I. Murstein (Ed.), *Theories of attraction and love.* New York: Springer, 1971.

Rachlin, M. *Introduction to modern behaviorism.* San Francisco: Freeman, 1970.

Rotton, J., Barry, T., Frey, J., & Soler, E. Air pollution and interpersonal attraction. *Journal of Applied Social Psychology,* 1978, *8,* 57–71.

Rubin, Z. *Liking and loving: An invitation to social psychology.* New York: Holt, Rinehart & Winston, 1973.

Sarnoff, I., & Zimbardo, P. G. Anxiety, fear, and social affiliation. *Journal of Abnormal and Social Psychology,* 1961, *62,* 356–363.

Schachter, S. *The psychology of affiliation.* Stanford, Calif.: Stanford University Press, 1959.

Sigall, H., & Aronson, E. Liking for an evaluator as a function of her physical attractiveness and nature of evaluations. *Journal of Experimental Social Psychology,* 1969, *5,* 93–100.

Tedeschi, J. T., & Lindskold, S. *Social psychology: Interdependence, interaction, and influence.* New York: Wiley, 1976.

Veitch, R., & Griffitt, W. Good news–bad news: Affective and interpersonal effects. *Journal of Applied Social Psychology,* 1976, *6,* 69–75.

Wright, P. H. Byrne's paradigmatic approach to the study of interpersonal attraction: Misgivings and alternatives. *Representative Research in Social Psychology,* 1971, *2,* 66–70.

Wrightsman, L. Effects of waiting with others on changes in levels of felt anxiety. *Journal of Abnormal and Social Psychology,* 1960, *61,* 216–222.

Received April 10, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 4, 580-590

Blood Donation and the Foot-in-the-Door Technique: A Limiting Case

Robert D. Foss
Department of Sociology and Anthropology
Western Carolina University

Carolyn B. Dempsey
University of Tennessee, Knoxville

Three experiments testing the effectiveness of the foot-in-the-door technique for recruiting blood donors consistently failed to demonstrate that this procedure influences either verbal or behavioral compliance, suggesting that the generality of the foot-in-the-door phenomenon is limited. Experiment 1 attempted to demonstrate that an earlier failure of this technique was due to poor operationalization rather than to the magnitude of the critical request or to the invalidity of the phenomenon, but it failed to do so. Experiment 2, designed to more closely resemble other foot-in-the-door studies by using telephone contacts and an initial request for persons to answer questions, was conducted to examine other possible explanations for the two previous failures. This experiment also failed to show any foot-in-the-door effect. Experiment 3 was a conceptual replication of Experiment 2 but used personal contacts. One apparent foot-in-the-door effect emerged in this case, but it was more likely due to a factor other than the experimental treatment. It is concluded that although the foot-in-the-door procedure may indeed influence verbal compliance with requests for minimal forms of aid, it probably will not significantly affect people's willingness to comply with more substantial requests involving behaviors that are psychologically costly to perform.

A number of recent investigations of the effects of compliance with a small request on subsequent compliance with a more substantial request (the "foot-in-the-door" phenomenon) suggest the possibility of a direct and exciting application of social psychological knowledge to practical matters. It has bee shown repeatedly that persons who are i duced to comply with a small request a much more likely to comply with a subs quent (critical) request of greater magnitu than would be the case if there had been preliminary request (Cann, Sherman, & Elke 1975; Freedman & Fraser, 1966; Pliner, Har Kohl, & Saari, 1974; Seligman, Bush, Kirsch, 1976; Snyder & Cunningham, 1975 The apparent reason for this effect is th initial compliance alters a person's self-pe ception, and as a result, the person, seei himself/herself as more of a helper than pr viously, is more likely to help a stranger whe asked (Snyder & Cunningham, 1975) merely when the situation presents its (Uranowitz, 1975).

One clear implication of these findings that various service organizations such as t American Cancer Society, the American Hea Association, the American Red Cross ought be able to make good use of this technique

The authors wish to express their appreciation to Morris White of the Asheville Regional Blood Center of the American Red Cross and Thomas Brinson of the MEDIC organization for their cooperation and assistance on this project and to Gail Whisenant for assisting with data collection in Experiment 1. A special note of appreciation is also due Anthony Greenwald for his careful consideration and comments on an earlier version of this article.

Carolyn B. Dempsey is now at the Special Projects Office, Eastern Band of the Cherokee Indian.

A report of Experiment 1 was presented at the annual meeting of the Southeastern Psychological Association, Hollywood, Florida, May 1977.

Requests for reprints should be sent to Robert D. Foss, Department of Sociology and Anthropology, Western Carolina University, Cullowhee, North Carolina 28723.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3704-580$00.75

induce compliance with their requests for aid in their respective tasks. In fact, Pliner et al. (1974) were able to nearly double the number of contributors to the Cancer Fund in a suburban neighborhood of Toronto, Canada, by using this technique. However, with this exception and one other, all studies of the foot-in-the-door phenomenon have used rather trivial requests, and a large number have employed a single research paradigm, which is to make requests over the telephone for persons to answer varying numbers of questions.

When the foot-in-the-door technique was used in a different and more realistic field setting in an attempt to increase blood donations, it failed (Cialdini & Ascani, 1976). In that study, the second request was made immediately after the first, which may not have allowed time for a person's self-perception to be sufficiently altered by compliance with the initial request. However, Cann et al. (1975) have shown, albeit in the somewhat artificial telephone interview paradigm, that timing of the second request apparently is not crucial for the foot-in-the-door effect to emerge. Therefore, it may simply be that moving from a rather trivial request such as answering questions in a telephone interview or donating money to a charity [1] to a more substantial one such as asking for a blood donation provides more of an obstacle than the foot-in-the-door technique can overcome.

Given that several studies of helping have documented a substantial rate of attrition between volunteering to do something and actually doing it (cf. Gross, Wallston, & Piliavin, 1975; Kazdin & Bryan, 1971), the generalizability of the foot-in-the-door phenomenon to settings in which such a technique might have some practical utility is further questioned; only Pliner et al. (1974) and Cialdini and Ascani (1976) have studied actual behavioral compliance with the critical request.

The present studies were designed to clarify some issues that are of importance for the practical application of the foot-in-the-door phenomenon. In particular, it would be useful to know (a) whether the effect does reliably occur with respect to overt behavior as well as verbal behavior, (b) whether the effect extends to other than question-answering behavior, and (c) how strong the effect is. Since Cialdini and Ascani's (1976) negative results amount to an acceptance of the null hypothesis, a replication of this result would be desirable before much confidence is placed in it. Therefore, the present series of experiments also used the foot-in-the-door technique in an attempt to increase blood donations during the visit of a bloodmobile to a university campus. Experiment 1 is a modified replication of Cialdini and Ascani's experiment in which timing of the critical request was manipulated to see whether it was the timing or the magnitude of the critical request that was more likely responsible for the failure of the foot-in-the-door effect to materialize in that study.

Experiment 1

Method

Research participants were 76 dormitory residents at a small, rural university. For each hall in each of the 10 dormitories on campus, a room number was randomly selected. Two female experimenters individually approached selected rooms between 6 p.m. and 10 p.m. and made the following initial request to the person answering the door:

> Hi, my name is _____, and I am working as a volunteer for the Red Cross. Our bloodmobile is to be here next week [on Tuesday] [tomorrow] and Wednesday from 12:00 to 6:00 p.m. Would you help us advertise its visit by putting this poster on your door?

Persons who agreed were given a poster that stated the time, place, and date of the bloodmobile visit and that contained the American Red Cross logo. The experimenter then took the subject's name and recorded his/her sex. All persons were thanked for their help, or if they did not cooperate, they were thanked just for listening. If no one answered the door of a preselected room, another room on the same hall was randomly selected and approached.

Timing between initial and critical requests was manipulated by making the initial request on the Wednesday (5-day delay) or Friday (3-day delay) of the week preceding the bloodmobile visit or on

---

[1] Forty-eight percent donated with no prior request in the Pliner, Hart, Kohl, and Saari (1974) study, indicating that donating small amounts of money ($M < \$1$) to a good cause is a behavior people are not terribly hesitant to engage in.

Table 1
Percentage of Verbal and Behavioral Compliance by Experimental Condition—Experiment 1

Compliance	Condition (%) 5-day delay	3-day delay	No delay	Control	χ^2	p
Verbal compliance						
Recontacted sample	53 (9 of 17)	54 (7 of 13)	55 (11 of 20)	53 (10 of 19)	.046	>.80
Behavioral compliance						
Recontacted sample	18 (3 of 17)	38 (5 of 13)	25 (5 of 20)	26 (5 of 19)	1.672	>.50
Entire sample	14 (3 of 22)	33 (5 of 15)	25 (5 of 20)	26 (5 of 19)	2.095	>.50

Note. *N*s are given in parentheses. Data for verbal compliance are for 69 persons who were successfully contacted for both the initial and critical request (recontacted sample). Data for behavioral compliance are reported for these persons (recontacted sample) and for these persons plus 7 persons who could not be contacted for the critical request but whose behavioral compliance could nevertheless be determined (entire sample).

the Monday evening (no delay) prior to the bloodmobile visit. The critical request was then made on the Monday evening preceding the bloodmobile visit.

In the 5-day and 3-day delay conditions, persons were recontacted by a different experimenter who also presented herself as a Red Cross volunteer and asked if the person would be willing to donate a pint of blood.

In the no-delay condition, after the person had agreed to the initial request to display the poster, the experimenter continued with the following request: "There is one other thing you can do for us. We are asking students to donate one pint of blood tomorrow and I wondered if you would volunteer to donate?" In the control condition, persons were simply contacted for the first time on the Monday evening preceding the bloodmobile visit and asked if they would donate. A list of the names of all persons who appeared at the donation center and either donated or were deferred for medical reasons was obtained from the American Red Cross. This list was used to determine how many persons in each experimental condition actually complied behaviorally with the critical request as well as how many of those in each condition were previous blood donors.

Results

There were no sex differences for verbal compliance, with 55% of males (18 of 33) and 53% of females (19 of 36) verbally agreeing to donate nor were there any differences for behavioral compliance, with 27% of males (9 of 33) and 25% of females (9 of 36) actually appearing at the donation center.[2] Therefore, male and female data were combined for the following analyses.

Verbal compliance. Table 1 indicates that there was practically no variation in compliance across conditions, and the differences did not approach statistical significance, $\chi^2(3) < 1$. Combining all experimental conditions, the overall verbal compliance was 54% (27 of 50), which is virtually identical to the 53% compliance rate in the control condition, $\chi^2(1) < 1$.

Behavioral compliance. It was expected that behavioral compliance would be substantially less than verbal compliance, and this turned out to be the case (see Table 1), with only about half of those who verbally complied actually showing up to donate. Again there were no significant differences in compliance rates between conditions, $\chi^2(3) = 1.67$, $p > .50$. The compliance rate in the combined experimental conditions was 26% (13 of 50), which is identical to the compliance rate in the control condition.

In an attempt to determine why behavioral compliance was so much lower than verbal compliance, information about the previous donation history of all persons who verbally complied was obtained from local Red Cross records. As Table 2 indicates, this one bit of information is vital in explaining differences in behavioral compliance among those who verbally complied. Of those who had donated

[2] Seven persons could not be recontacted for the critical request (five in the 5-day delay condition and two in the 3-day delay condition).

before, 82% (14 of 17) carried through on their verbal commitment, whereas only 21% (4 of 19) of those who had never donated before did so, $\chi^2(1) = 11.15$, $p < .001$ ($\phi = .61$).[3] Thirty-seven percent of the variation in behavioral compliance in this group was accounted for by donation history. Furthermore, it is evident from the table that this relationship between donation history and behavioral compliance holds up within both the experimental condition, $\chi^2(1) = 7.96$, $p < .005$ ($\phi = .63$) and the control condition (Fisher's exact test, $p = .083$, $\phi = .65$), suggesting that the experimental treatment had no effect on the relationship between these variables.

Table 2
Behavioral Compliance of Previous Donors and Previous Nondonors—Experiment 1

	Condition			
	Experimental[a]		Control[b]	
Compliance	Donor	Non-donor	Donor	Non-donor
Yes	9	4	5	0
No	1	12	2	3
Total	10	16	7	3

Note. This table includes only those persons who verbally complied with the request to donate blood plus two who failed to verbally comply but donated anyway.
[a] In the experimental condition, $\chi^2_{(1)} = 7.96$, $p < .005$, $\phi = .63$.
[b] In the control condition, Fisher exact, $p = .083$, $\phi = .65$.

Discussion

The present results call into question the external validity of the foot-in-the-door phenomenon. While it previously appeared that Cialdini and Ascani (1976) may not have obtained the usual effect because of an improper operationalization of the technique, the present study rules out this interpretation and leaves us with the suspicion that this phenomenon may not be robust enough to affect those behaviors that people are initially somewhat hesitant to engage in. It is well-known that most people, if they have never donated blood before, are quite reluctant to do so (American Red Cross, undated). On the other hand, examining the amount of compliance obtained in control conditions of previous foot-in-the-door studies, we find that anywhere from 17% (Freedman & Fraser, 1966) to 50% (Cann et al., 1975) of the people who were simply asked initially to engage in the criterion behavior were willing to do so. This suggests that none of these behaviors were particularly noxious in the first place. (Indeed, with the exception of the Freedman & Fraser experiments, the criterion behaviors were quite minimally inconveniencing or costly.)[4]

Another possible reason for the failure of the foot-in-the-door effect to materialize for Cialdini and Ascani (1976) as well as in the present experiment has to do with situational factors that may have intervened to prevent any alteration in self-perception as a result of compliance with the initial request. If there is strong external justification for a behavior, that behavior is likely to be dis-

[3] All chi-square values reported with 1 degree of freedom are corrected for continuity.

[4] It might be argued that the 32% verbal compliance rate obtained by Cialdini and Ascani (1976) and the 53% verbal compliance in the present study provide evidence that people are also fairly willing to donate blood; however, such a conclusion is not justified. The baseline compliance rate in blood donation studies is elevated by the presence in the sample of previous donors, who are much more likely to agree to give blood than persons who have never donated (Foss, Note 1). Consequently, this baseline rate is not an accurate indicator of the reluctance of people in general to engage in this particular criterion behavior.

In the present study, both the 53% verbal compliance and the 26% behavioral compliance are abnormally high and, in fact, are much higher than those reported by Cialdini and Ascani (1976). This can be accounted for the fact that there is a well-organized, continuing effort to obtain blood donations on the campus at which the present study was conducted, providing a climate of normative support for this activity (cf. Barton, 1969), and many students have donated previously. Indeed, of those in the present experiment who verbally complied, 47% were previous donors, and 78% of those who actually donated were previous donors. In a subsequent study (Experiment 2) conducted on another campus at which the normative climate is not so conducive to blood donation, verbal compliance was only 25% and behavioral compliance a mere 3%.

counted rather than taken as an indication of a person's character (Bem, 1972; Kelley, 1971). It is quite possible that the initial request (to help advertise the presence of a mobile blood-collection unit on campus) was for such a good cause that people perceived a great deal of external justification for doing so—"Sure I'll do that, anybody would."—and, hence, did not attribute "helper" characteristics to themselves subsequent to the initial behavior.

There is yet another possible explanation for the failure of the foot-in-the-door technique to influence blood donation. It may be that size of the initial and critical requests must be correlated such that for a larger critical request, one must use a more substantial initial request. Seligman et al. (1976) found this to be the case using the telephone interview paradigm, and it is plausible that one's self-perception will be more strongly influenced by an initial agreement to a moderate request than to a small request. Therefore, it may still be possible to develop practical applications of this technique in areas in which human aid or assistance are needed but in which people are reluctant to provide such services. However, by having to increase the magnitude of the initial request, one is of necessity decreasing the number of persons who can be subjected to the technique, because of the decreased likelihood of compliance with the initial request. One of the main appeals of the foot-in-the-door technique is the possibility that even the most trivial of initial requests may be sufficient to increase subsequent compliance. However, if this is not the case, and considering the substantial effort that would be necessary to carry out a recruitment project using this technique, the minimal payoffs may not be worth the effort. Nevertheless, in the interest of a better understanding of this phenomenon and to provide a more thorough specification of relevant parameters, two further experiments were conducted.

Experiment 2

A second study was conducted to examine the possibilities suggested earlier (a) that because of the compelling nature of the request used, altered self-perceptions might not be induced or (b) that the initial request was not large enough to alter self-perceptions sufficiently to influence compliance with so large a request as donating blood. Contacts were made by telephone, as has been the case in most other foot-in-the-door studies, and a condition similar to that typically used in such studies was added to approximate more closely the procedures used in studies in which this technique has been successful. This experiment was also conducted in connection with a 2-day bloodmobile visit to a university campus.

Method

Research participants were 135 dormitory residents at a large, urban university. Participants were randomly selected from a listing of dormitory residents and assigned to one of three experimental conditions or to a control group. Persons in the experimental conditions were then contacted by phone on either Wednesday or Thursday evening between 6:00 p.m. and 9:30 p.m. and were asked if they would be willing to do one of the following tasks, depending on the condition to which they had been assigned:

Routine initial request. Subjects were asked to answer a few questions about blood donation. If they agreed, five questions were asked, including whether they had ever donated, and they were thanked for their cooperation.

Poster request. Subjects were asked to allow the blood organization to put a poster on their dormitory room door to advertise the bloodmobile visits. If they agreed, they were asked the location of their room, their age, and whether they had ever donated blood. They were then thanked for their cooperation.

Large initial request. Subjects were asked to recruit four friends to donate blood as part of a new approach to donor recruitment being tried during the upcoming bloodmobile visit. If they agreed, they were simply told to contact four friends, inform them that the bloodmobile would be on campus the following Monday and Tuesday, and ask them to donate. Participants were then asked their age and whether they personally had ever donated. They were then thanked for their cooperation.

On the following Sunday and Monday nights (3–5 days later), all experimental subjects and all members of the control group were called by a different person, who informed them that the bloodmobile would be on campus for two days and who asked them if they would be willing to donate a unit of

Table 3
Percentage of Verbal and Behavioral Compliance by Experimental Condition—Experiment 2

Item	Condition (%)				χ^2	p
	Routine request	Poster request	Large request	Control		
Verbal compliance						
Recontacted sample	20 (6 of 30)	29 (6 of 21)	31 (9 of 29)	20 (6 of 30)	1.51	>.60
Previous nondonors	19 (5 of 26)	22 (4 of 18)	33 (9 of 27)	22 (5 of 23)	1.67	>.60
Behavioral compliance						
Entire sample	5 (2 of 41)	4 (1 of 28)	0 (0 of 36)	3 (1 of 30)	—	—

Note. *N*s are given in parentheses. Data for verbal compliance are for 110 persons who were successfully contacted for both the initial and critical request. Data for behavioral compliance are for these persons plus 25 others who could not be contacted for the critical request but whose behavioral compliance could nevertheless be determined (one of the donors in the routine condition was in this group).

blood.[5] Persons who agreed were told of the time and place and were reminded to eat a good meal before coming to donate.

To summarize, there were three experimental conditions ranging from a small initial request, which is the type normally used in foot-in-the-door research, to a moderate initial request that was essentially identical to that used in Experiment 1, to a large initial request that should have a greater effect on self-perceptions if persons agree to it. There was also a control condition with no initial request.

Results

Verbal compliance. Twenty-five persons could not be recontacted for the critical request (7 each in the poster and large request conditions, 11 in the routine condition), hence, for these people the amount of verbal compliance is unknown.[6] Persons who said they might donate were classified as refusals. Table 3 shows the amount of verbal compliance both before and after persons who reported they were previous donors were removed from the analysis. In neither case was the overall chi-square significant. In both instances the greatest amount of verbal compliance was obtained in the large initial request condition, but in neither case was this significantly different from the control condition. Finally, combining the experimental groups, the overall compliance rate was 26% (21 of 80), which does not differ from the 20% compliance obtained in the control condition, $\chi^2(1) < 1$. With previous donors removed, the combined compliance rate was 25% (18 of 71), which is not significantly different from the 22% compliance in the control condition, $\chi^2(1) < 1$.

Behavioral compliance. There was virtually no behavioral compliance, with only four persons donating, one of whom had not been successfully contacted for the critical request. This result underscores one conclusion of another study of blood donation that personal contact is a far more effective recruitment technique than telephone contact (Ford & Wallace, 1975). Unfortunately, the extremely small amount of behavioral compliance obtained in the present study (3%) precludes a definitive test of the possible effects of differing initial requests, although the failure of effects to materialize with verbal com-

[5] All callers were unaware of the hypothesis being tested. Although some persons in the large request condition may have failed to behaviorally comply with the initial request, only 2 persons (poster condition) would not agree to the initial request. Due to the time pressures generated by having to contact 135 persons within two fixed 3½-hour time periods, these 2 persons were not recontacted for the critical request but were included in the analyses where possible (i.e., behavioral compliance).

[6] This is a disturbingly large attrition rate; however, there is little reason to believe that these persons differ in any important way from those we could successfully recontact. Because of the logistics of contacting a large number of persons during a relatively short time period, only a limited number of call backs were possible, and these persons simply were not in when we called.

pliance suggests once again that the foot-in-the-door technique simply does not work with blood donation. Nevertheless, a third experiment was conducted in a further attempt to determine if it might yet work under the right circumstances.

Experiment 3

Method

This experiment was conducted on the same campus as Experiment 1, 9 months later, during a different academic year, and again in connection with a bloodmobile visit to campus. Procedures were similar to those in Experiment 2, but dormitory rooms rather than individuals were sampled to maximize the sample size with which we would ultimately be able to make two contacts. Five or 6 days prior to the bloodmobile visit, research participants were contacted in person by an experimenter who delivered one of three initial requests to whomever answered the door, depending on the condition to which the room had been assigned:

Routine initial request. Subjects were asked to answer six or seven questions about their knowledge of and experience with blood donation.

Poster request. Subjects were asked to put a poster on their door advertising the bloodmobile visit to campus.

Large initial request. Subjects were asked to recruit four of their friends to donate blood.

Persons who agreed to the request were asked the questions, given the poster, or told how to recruit friends as in Experiment 2. They were then asked their names and whether they had ever donated blood. For purposes of secondary analysis, a record was kept of all persons who reported conditions that would legitimately disqualify them from donating blood (chronic illness, underweight, temporary illness, etc.).

On the evening preceding the bloodmobile visit, all persons who had received an initial request and persons in the control group were contacted by a different experimenter (who feigned ignorance of the earlier contact by an experimenter)[7] and asked to donate blood the following day. Those who agreed to donate were reminded of the time and place and to eat a good meal before coming to donate. Everyone was thanked for his/her time, even if he/she had refused to donate.

Results

Of the 127 persons initially contacted, 95 were successfully recontacted and asked to donate blood. In addition, 36 persons in the control group were contacted for the first time and asked to donate. Since there were no significant differences between males and females for verbal compliance (36% vs. 39%) or for behavioral compliance (12% vs. 14%), data were combined for the following analyses.[8]

Verbal compliance. Although there were some moderate variations in verbal compliance across conditions (see Table 4), these did not approach statistical significance, $\chi^2(3) = 2.30$, $p > .50$. Compliance was greatest in the poster condition (44%), although it did not differ significantly from the control condition, $\chi^2(1) = 1.27$, $p > .20$. The 40% (38 of 95) compliance rate among combined experimental conditions also did not differ significantly from the control condition, $\chi^2(1) < 1$.

Behavioral compliance. Since 32 persons could not be recontacted for the critical request (8, 11, and 13 in the routine, poster, and large request conditions, respectively), two sets of analyses were conducted. Although these persons were not directly asked to donate, they were in experimental conditions that, in view of the processes apparently involved in the foot-in-the-door phenomenon, should have made them more likely to do so when the opportunity presented itself (cf. Uranowitz, 1975). On the campus at which this study was conducted, almost everyone knows when the bloodmobile is on campus and, therefore, is aware of the opportunity to donate.

First, in the recontacted sample (see Table 4), which provides the easier task for the foot-

---

[7] Both verbal and behavioral compliance rates were substantially higher in Experiment 1. This finding appears to have been due primarily to the fact that Experiment 3 was conducted during the winter, in the midst of a flu epidemic, whereas Experiment 1 was conducted in the spring.

[8] Although recruiters were initially blind to the respondent's experimental condition, those in the poster condition were easily recognized as such when the recruiter approached the room. This is unfortunate but is probably not serious, since recruiters were unaware of the specific nature of the hypotheses being tested. Furthermore, the crucial question became (see below) whether any or all experimental conditions differed from the control group rather than whether there were any differences between experimental conditions.

Table 4
Percentage of Verbal and Behavioral Compliance by Experimental Condition—Experiment 3

Item	Condition (%)				χ^2	p
	Routine request	Poster request	Large request	Control		
Verbal compliance						
Recontacted sample	43 (16 of 37)	44 (14 of 32)	31 (8 of 26)	31 (11 of 36)	2.30	>.50
Behavioral compliance						
Recontacted sample	16 (6 of 37)	31 (10 of 32)	12 (3 of 26)	6 (2 of 36)	8.83	<.05
Entire sample	13 (6 of 45)	23 (10 of 43)	8 (3 of 39)	6 (2 of 36)	6.79	<.08

Note. *Ns* are given in parentheses. Data for the recontacted sample are for 131 persons who were successfully contacted for both the initial and critical request. Data for the entire sample also include 32 persons who could not be contacted for the critical request but whose behavioral compliance could nevertheless be determined.

in-the-door technique, there was a significant overall effect, $\chi^2(3) = 8.83$, $p < .05$; this was accounted for by the poster condition, which differed significantly from the control condition, $\chi^2(1) = 6.03$, $p < .02$. Neither of the other experimental conditions differed significantly from the control group. The combined experimental conditions differed from the control group at a marginal significance level (20% vs. 6%), $\chi^2(1) = 3.05$, $p < .09$, but again this is due primarily to the contribution of the poster condition.

For the entire sample, behavioral compliance in the experimental conditions was again somewhat higher than in the control condition, and the overall effect approached conventional levels of statistical significance, $\chi^2(3) = 6.79$, $p < .08$. The poster condition accounted for this effect and was significantly different from the control condition (23% vs. 6%), $\chi^2(1) = 3.49$, $p < .06$, whereas neither of the other two experimental conditions differed significantly from the control group. It should be pointed out that none of the 32 persons who were not recontacted actually donated, and it is arguable that the only fair test of the foot-in-the-door effect is the previous one involving only those contacted for both the initial and critical requests.[9]

Donation history. Among those persons who reported having donated previously, 59% (23 of 39) agreed to donate and among those who had not donated previously 30% (25 of 83) agreed to donate; $\chi^2(1) = 8.09$, $p < .01$.[10] Actual donation was also more common among previous donors (24%, 13 of 54) than among persons who had never donated (8%, 8 of 100), $\chi^2(1) = 6.39$, $p < .02$.

In the present experiment, previous donation history did not play such a strong role in determining whether persons who agreed to donate would actually do so as it did in Experiment 1. Nearly half (48%) of those who agreed to donate were previous donors, and of those, 39% actually donated, whereas 28% of those who agreed to donate but were not previous donors actually did so, $\chi^2(1) < 1$. Still, donation history is an important determinant of behavioral compliance, with those who have donated before three times as likely to donate as those who have not previously done so.[11] In the present experiment,

[9] A third set of analyses was conducted after excluding the 44 persons who reported a legitimate medical excuse for not donating, as well as the 32 persons not recontacted. This altered the percentages, but the same pattern of results as in the above analyses emerged for both verbal and behavioral compliance.

[10] Because of an oversight, information about donation history was not obtained from nine persons in the control group.

[11] In this study, information on donation history was obtained from the persons themselves, whereas in Experiment 1, it was obtained from actual records. It is possible that some persons felt a need to deceive us about their donation history and, therefore, that the results of this analysis are less reliable than those from Experiment 1.

previous donors constituted only one third (35%) of the sample, yet nearly two thirds (62%) of the donations were made by persons who had previously donated.

General Discussion

Since the experiments reported here primarily provide confirmation of a null hypothesis, a brief discussion of methodological issues surrounding this procedure is in order. Traditionally, in social psychology, experimental results that confirm the null hypothesis have been viewed with more than a little skepticism—often with good reason (Aronson & Carlsmith, 1968). Even if a study is well designed and well executed, one plausible alternative explanation of experimental results that confirm a null hypothesis is that the statistical test used may not have been powerful enough to detect a true effect. Thus, conventional canons of research dictate that very large samples are necessary to effectively test a null hypothesis.

The samples in the present study clearly are not particularly large, averaging about 31 persons per condition, which leaves open the possibility that the foot-in-the-door technique does indeed have an effect on compliance with requests to donate blood, but the effect is too small to detect with such small sample sizes. However, the purpose of the present studies was not the traditional one of examining a theoretical proposition but rather to examine the utility of a phenomenon in practical application. In such a situation, we are not particularly interested in small effects, real though they may be. Even though a compliance-inducing procedure might have the (statistically reliable) effect of increasing compliance by, say, 25% over normal procedures (an effect that would not be detected with a moderate sample size), this might very well be too small an effect to be of value in practical applications.[12] In organizational donor recruitment programs, cost–benefit analyses must be conducted, and on balance, procedures that produce real but small effects are not feasible unless they involve essentially no extra effort or expense to put into practice. To this we would add that in the present series of experiments, there is little evidence to suggest that there is even a small foot-in-the-door effect that would have been detected with larger sample sizes. The fluctuation in compliance rates across experiments and conditions appears to be essentially random.

In the three experiments conducted, comprising a total of nine experimental groups evaluated on two dependent measures each, only one foot-in-the-door effect occurred (Experiment 3, poster condition, behavioral compliance). The fact that this effect occurred in a condition that had no discernible effect in several other comparisons in Experiments 1 and 2 and that a behavioral compliance effect occurred in a condition in which there was no verbal compliance effect was somewhat puzzling. Further examination of the data yielded an explanation for this result. It was discovered that this one condition contained a substantially larger proportion of previous donors than any other condition (47% vs. 28%, 29%, and 37% in the routine, large request, and control conditions, respectively). Given the much stronger inclination of previous donors to donate, this provides a more parsimonious explanation of the greater behavioral compliance in this condition than does the foot-in-the-door explanation.

Given the consistent failure across three experiments to find a foot-in-the-door effect, despite different procedures (both telephone and in-person requests), study of different populations (urban and rural universities), and testing some of the most plausible alternative explanations for the initial failure, the only reasonable conclusion we are left with is that the foot-in-the-door technique does not work with blood donation. It appears that this behavior is too costly for a slightly altered self-concept, which the foot-in-the-door procedure presumably produces, to influence a person's willingness to donate.

---

[12] Assuming a 12% rate of compliance with a straightforward request to donate blood (based on control groups of studies in which persons were personally asked), a 25% increase above that would amount to only 3% more donors, which would hardly justify the expense of doubling the number of personal contacts in order to bring about the foot-in-the-door effect.

The present series of experiments provides evidence that the power of the foot-in-the-door technique for practical applications may have been overstated. Two of the most comprehensive and popular recent social psychology textbooks maintain that the effectiveness of this technique is well established (Baron & Byrne, 1977; Wrightsman, 1977). However, although the existence of the foot-in-the-door phenomenon is well documented, its utility in practical applications is not, with only a single study demonstrating an effect on behavioral compliance with a request for a meaningful form of aid. The present series of failures, in addition to at least two others (Cann, 1975; Cialdini & Ascani, 1976), combined with the shortcomings of other studies previously mentioned, suggest that this technique may be of only limited value to organizations seeking ways to increase compliance with their requests for aid.

We suggest that at least part of the reason that the misconception about the power and practical usefulness of this technique has developed is because of the characteristics of the research publication system, which discourage the submission and acceptance for publication of studies that report negative results, that is, those studies that support the null hypothesis (Greenwald, 1975). Such biases against accepting the null hypothesis are particularly serious when results of social psychological inquiry are likely to be rapidly applied. If only positive findings emerge, a very misleading conception of a phenomenon can develop, and this appears to be precisely what has occurred with the foot-in-the-door phenomenon. The technique does appear to work well in some fairly limited instances, but the limitations have not been spelled out. The present series of experiments is a first step in the direction of identifying the crucial parameters that influence if, when, and perhaps how this technique works. Clearly, much remains to be done in order that future reports on the foot-in-the-door technique can be accompanied by appropriate cautionary remarks detailing the limitations of this procedure.

One final point is worth mentioning. Although the present series of experiments demonstrates convincingly that the foot-in-the-door effect does not influence blood donation, some caution should be exerted in interpreting these findings. On the basis of this single series of experiments plus one other (Cialdini & Ascani, 1976), it would be risky to conclude that all costly behaviors are resistant to the foot-in-the-door phenomenon, although the implication is certainly there. In some ways, blood donation is perhaps an unusual form of helping. For example, previous donation is an extremely powerful factor in present behavior, which is primarily a reflection of the irrational fears people have about donating blood (cf. Foss, Note 1). On the other hand, it would be equally unwise to attempt to explain away the present results as being due to the peculiar nature of blood donation. Other behaviors such as donation of organs, bone marrow, and large amounts of time or money have equally high costs (of varying sorts) and are also behaviors that a given proportion of any population simply cannot engage in because of physical or other unalterable limitations. Finally, if blood donation were also resistant to other, thoroughly documented, forms of social influence, we would be justified in concluding that the current results tell us more about blood donation than about the foot-in-the-door phenomenon. However, blood donation is quite susceptible to various other forms of social influence (cf. Cialdini & Ascani, 1976; Condie, Warner, & Gillman, 1976; Ford & Wallace, 1975; Foss, Note 1). Therefore, we may conclude with some confidence that the peculiar nature of blood donation alone does not explain our consistent negative findings.

Reference Note

1. Foss, R. D. *The role of social influence in blood donation.* Paper presented at annual meeting of American Psychological Association, Toronto, August 1978.

References

American National Red Cross. *Selected readings in donor motivation and recruitment.* Washington, D.C.: Author, undated.

Aronson, E., & Carlsmith, J. M. Experimentation in social psychology. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (Vol. 2, 2nd ed.). Reading, Mass.: Addison-Wesley, 1968.

Baron, R. A., & Byrne, D. *Social psychology: Understanding human interaction.* Boston: Allyn & Bacon, 1977.

Barton, A. H. *Communities in disaster.* New York: Doubleday, 1969.

Bem, D. J. Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6). New York: Academic Press, 1972.

Cann, A. A. *The effects of performing a small consonant behavior on subsequent attitudes and behaviors.* (Doctoral dissertation, Indiana University, 1975). *Dissertation Abstracts International,* 1976, *36,* 5859B. (University Microfilms No. 76-11, 359)

Cann, A. A., Sherman, S. J., & Elkes, R. Effects of initial request size and timing of second request on compliance: The foot in the door and the door in the face. *Journal of Personality and Social Psychology,* 1975, *32,* 774–882.

Cialdini, R. B., & Ascani, K. Test of a concession procedure for inducing verbal, behavioral and further compliance with a request to give blood. *Journal of Applied Psychology,* 1976, *61,* 295–300.

Condie, S. J., Warner, W. K., & Gillman, D. C. Getting blood from collective turnips: Volunteer donation in mass blood drives. *Journal of Applied Psychology,* 1976, *61,* 290–294.

Ford, G. T., & Wallace, E. L. Effects of donor recruitment methods on population responses. *Transfusion,* 1975, *15,* 159–164.

Freedman, J. L., & Fraser, S. C. Compliance without pressure: The foot-in-the-door technique. *Journal of Personality and Social Psychology,* 1966, *4,* 195–202.

Greenwald, A. G. Consequences of prejudice against the null hypothesis. *Psychological Bulletin,* 1975, *82,* 1–20.

Gross, A. E., Wallston, B. S., & Piliavin, I. M. Beneficiary attractiveness and cost as determinants of responses to routine requests for help. *Sociometry,* 1975, *38,* 131–140.

Kazdin, A. E., & Bryan, J. H. Competence and volunteering. *Journal of Experimental Social Psychology,* 1971, *7,* 87–97.

Kelley, H. H. *Attribution in social interaction.* Morristown, N.J.: General Learning Press, 1971.

Pliner, P., Hart, H., Kohl, J., & Saari, D. Compliance without pressure: Some further data on the foot-in-the-door technique. *Journal of Experimental Social Psychology,* 1974, *10,* 17–22.

Seligman, C., Bush, M., & Kirsch, K. Relationship between compliance in the foot-in-the-door paradigm and size of first request. *Journal of Personality and Social Psychology,* 1976, *33,* 517–520.

Snyder, M., & Cunningham, M. R. To comply or not comply: Testing the self-perception explanation of the foot-in-the-door phenomenon. *Journal of Personality and Social Psychology,* 1975, *31,* 64–67.

Uranowitz, S. W. Helping and self-attributions: A field experiment. *Journal of Personality and Social Psychology,* 1975, *31,* 852–854.

Wrightsman, L. *Social Psychology* (2nd ed.). Monterey, Calif.: Brooks/Cole, 1977.

Received April 21, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 4, 591-601

Erotica and Aggression: The Influence of Sexual Arousal, Positive Affect, and Negative Affect on Aggressive Behavior

Leonard A. White
Purdue University

Male college students were first either angered or not angered by a confederate of the experimenter and were then ostensibly given an opportunity to aggress against the confederate by means of electric shock. Prior to aggressing, subjects were shown one of four sets of stimuli chosen to effect a factorial variation in the intensity of positive sexual arousal (high, low) and negative affect (high, low) elicited by exposure to such material. In addition, one group of angered subjects (no-exposure control) was included who did not view any of the four sets of stimuli prior to being given an opportunity to aggress. Results indicated that exposure to affectively positive erotic stimuli significantly reduced retaliatory behavior by angered males to a level below that exhibited by subjects exposed to neutral stimuli and by those in the no-exposure control group. In contrast, relative to baseline controls, subjects' exposure to erotic stimuli that were reported to be disgusting and unpleasant slightly enhanced subsequent aggressive behavior. A number of possible mechanisms (e.g., attentional shifts, incompatible responses, cognitive labeling) are discussed in relation to these results.

During recent years, it has become apparent that exposure to erotic stimuli may have a variety of effects on subsequent aggressive behavior. In past studies, exposure to sexual imagery has been found at times to decrease attacks against others (Baron, 1974a, 1974b; Baron & Bell, 1977), at other times to increase subsequent aggression (Zillmann, 1971), and at still other times to have seemingly little effect on such behavior. Experimental investigations (e.g., Baron & Bell, 1977) aimed at delineating more precisely the types of erotic material that inhibit or facilitate subsequent aggression have implicated the sexual arousal potential of such stimuli. Specifically, mildly arousing stimuli have been found to decrease subsequent aggression, whereas exposure to more highly arousing sexual imagery tends to facilitate such behavior. In addition, a close examination of past research reveals that affective responses to erotic imagery may also need to be taken into account.

Affective Response to Erotic Stimuli and Aggression

Some suggestion as to the way affective responses may relate to aggression is provided by the results of a recent study by Baron (in press). In the study by Baron, angered females showed increases in aggression following exposure to explicit erotic scenes, which had been found in a previous study (Baron & Bell, 1977) involving males to inhibit such behavior. To reconcile these seemingly discrepant findings, Baron pointed out that females reported finding these depictions of explicit sexual activity to be disgusting and unpleasant, whereas for males, sexual imagery

The author would like to express appreciation to Donn Byrne, Robert A. Baron, Carl Castore, Elizabeth Capaldi, Jeffrey D. Fisher, and William A. Fisher for their valuable comments and assistance in this study. Thanks also go to Donald Amoroso, Kathy Kelley, and the Institute for Sex Research, Bloomington, Indiana, for their assistance in obtaining the experimental stimuli. Finally, the assistance of the Social Science Data Center at the University of Connecticut is gratefully acknowledged.

Requests for reprints should be sent to Leonard A. White, who is now at the Department of Medical Psychology, Uniformed Services University of the Health Sciences, 4301 Jones Bridge Road, Bethesda, Maryland 20014.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3704-591$00.75

was regarded as positively exciting and pleasant. These contrasting affective responses, in turn, were hypothesized to have led to different levels of aggression by modifying subjects' attributions concerning the source of their arousal (cf. Baron, 1977; Rule & Nesdale, 1976). Specifically, for females, increments in sexual arousal that were regarded as unpleasant may have been more readily interpreted as anger and consequently increased subsequent aggression. Alternatively, it was suggested that males, who responded with greater positive affect, may have been willing to label their arousal more positively (e.g., as due to the sexual stimuli, rather than as anger), with the result that subsequent retaliatory behavior was reduced.

At first glance, it may appear to follow directly from the studies described above that exposure to erotic materials that are labeled and experienced as unpleasant and disgusting will enhance subsequent aggression, whereas exposure to more positively valenced sexual stimuli effectively inhibits such behavior. Upon closer examination, however, it is apparent that the results described above must be regarded as tentative and clearly demand further experimental scrutiny. For example, in the studies by Baron (Baron, in press; Baron & Bell, 1977) outlined above, while depictions of explicit sexual activity may have evoked contrasting affective responses by males and females, these feelings were not measured directly in the study involving males, but inferred from subjects' comments during debriefing. In addition, since baseline control groups have not been included in past studies of erotica and aggression, it is possible that what has been interpreted as an aggression-inhibiting effect of positively valenced sexual imagery was actually due to an aggression-facilitating effect of more boring and less arousing control stimuli (cf. Zillmann & Sapolsky, 1977). In addition, no evidence is available concerning the impact of more negatively valenced sexual imagery on subsequent aggression by males. Thus, the present study was designed to provide data relevant to answering some of the questions surrounding the past research described above, through an examination of the influence of erotica-produced affect and arousal responses on aggression.

Dimensions of Affective Response

One final concern surrounding the preparation of specific sets of stimuli involved the measurement of affective responses to erotica and their relationship to sexual arousal. Toward this end, the present study incorporated the results of recent work (Byrne, Fisher, Lamberth, & Mitchell, 1974) that delineates these relationships more precisely. The results of this research indicate that sexual arousal and affective responses to erotica may be conceptualized as varying along two orthogonal dimensions—one dimension that I will refer to as *positive sexual arousal,* composed of subjective sexual arousal and positive affective responses related to sexual arousal, and an orthogonal dimension of *negative affect,* characterized by responses of disgust, nausea, and guilt. Within this two-dimensional framework, four distinct patterns of response to erotic stimuli can be identified: Exposure to erotic imagery may lead to increases in sexual arousal that are labeled and experienced as pleasant and entertaining (*positive response*), evoke a combination of positive affect and negative feelings (*ambivalent response*), lead primarily to negative affect (*negative response*), or elicit feelings of relative indifference (*neutral response*).

Experimental Design and Hypotheses

Procedures in the present investigation followed closely those used by Baron (e.g., Baron & Bell, 1977) in past research relating exposure to erotica and aggression. Male college students were either angered or not angered by a confederate and were then provided with an opportunity to aggress against the confederate by means of electric shock. Prior to aggressing, participants were exposed to one of four sets of stimuli chosen to effect a factorial variation in the intensity of positive sexual arousal (high, low) and negative affect (high, low) associated with an erotic stimulus. Each set of stimuli was chosen to elicit one of the four categories of response to

rotic imagery outlined above—positive, negtive, ambivalent, and neutral.

Based on the discussion above, an interction was anticipated such that aggressive esponding would be (a) inhibited by expoure to erotic stimuli that evoked primarily ositive affect (positive response), (b) enanced by exposure to arousing erotic imagery nat evoked a combination of positive and egative affect (ambivalent response), and c) neither reduced nor facilitated by exosure to neutral stimuli or erotic imagery hat led primarily to heightened negative eelings. Simple increases in negative affect esulting from exposure to sexual stimuli negative response) were not expected to ncrease subsequent aggression, since the failitation of aggression by erotica presumably equires heightened levels of sexual arousal e.g., Baron & Bell, 1977). Although this redicted pattern of results was expected to e more clearly evident for angered males relative to nonangered), this prediction must e regarded with caution, since exposure to rotic stimuli has been found (e.g., Baron & Bell, 1977) to alter aggression by nonangered ersons.

Finally, to determine if the anticipated eduction in aggression by exposure to erotic magery that evokes primarily positive affect nay actually represent an aggression-facilitating effect of neutral control stimuli, a group of provoked subjects was included who lid not view any of the four sets of stimuli rior to being given an opportunity to aggress.

Pilot Study

Method

Prior to the main experiment, a pilot study was conducted to determine if the chosen sets of stimuli were suitable for inducing a factorial variation in positive sexual arousal and negative affect.

Thematic Content of the Experimental Stimuli

Each of the four sets of experimental stimuli was composed of 10 color slides that depicted the following themes: (a) positive response—mutual genital fondling, sexual intercourse, and mildly explicit fellatio; (b) ambivalent response—moderately explicit mutual oral–genital contact, fellatio, and cunnilingus; (c) negative response—same-sex masturbation and highly explicit cunnilingus; (d) neutral response—clothed and partially clothed males and females standing alone or mutually involved in what seemed to be a nonsexual activity.

Subjects and Procedure

After providing informed consent, 48 male undergraduates were randomly assigned to view one of the four sets of stimuli until a total of 12 subjects had rated each set. Following presentation of the stimuli, each participant completed a Feelings Scale (Byrne et al., 1974) that required a rating of sexual arousal and self-ratings of affect along 10 dimensions: disgusted, entertained, anxious, bored, angry, nauseated, depressed, guilty, excited, and disappointed with oneself. Males then responded to items pertaining to sexual-physiological reactions (Schmidt & Sigusch, 1970) and to six semantic-differential scales.

Results and Discussion

As an initial step in the analysis, responses to the Feelings Scale (for all four sets of stimuli combined) were factor analyzed and orthogonally rotated using a varimax criterion (Kaiser, 1958). In line with expectations based on past research (e.g., Byrne et al., 1974), two orthogonal dimensions of response were found. The first factor, accounting for 35% of the variance, was labeled as "positive sexual arousal" and consisted of the rating of sexual arousal and four positive affective responses related to sexual arousal (excited, anxious, entertained, and not bored).[1] The second factor, accounting for 30% of the variance, was labeled "negative affect" and consisted of responses to the items angry, disgusted, nauseated, disappointed with oneself, depressed, and guilty.

Subsequent analyses sought to determine if the chosen sets of stimuli produced the intended differentiations in response along these two dimensions. Multivariate analysis of variance applied to responses along these two dimensions yielded the intended differentiations, showing only a main effect of positive sexual arousal, multivariate $F(5, 40) =$

[1] Anxiety as a positive response would appear to denote feelings of positive anticipation and excitement that preceded the onset of each stimulus.

Table 1
Mean Affect and Arousal Responses to Experimental Stimuli in the Pilot Study

Condition	Positive sexual arousal		Negative affect	
	Sexual arousal	Positive affect related to sexual arousal	Disgust	Negative affect (with disgust)
High sexual arousal/high negative affect (ambivalent response)	4.17_a	4.18_a	3.17_a	2.39_a
High sexual arousal/low negative affect (positive response)	4.17_a	4.10_a	1.25_b	1.23_b
Low sexual arousal/high negative affect (negative response)	1.25_b	1.95_b	4.33_a	3.37_a
Low sexual arousal/low negative affect (neutral response)	1.67_b	2.12_b	1.00_b	1.18_b

Note. $n = 12$ within each cell. A higher number (on a 7-point scale) indicates more intense affect and arousal responses. Within each dependent measure, means that do not share a common subscript differ significantly at the .05 level by Duncan's multiple-range test.

10.02, $p < .001$, and a main effect of negative affect, multivariate $F(6, 39) = 12.09$, $p < .001$.[2] As summarized in Table 1, males exposed to erotic imagery chosen to elicit high (as opposed to low) negative affect reported much greater disgust ($p < .0001$), as well as more intense anger, nausea, guilt (for each item, $p < .001$), and self-disappointment ($p < .02$). In addition, and as intended, sets of stimuli chosen to elicit high positive sexual excitement (as opposed to low) were reported to be more sexually arousing, more exciting, more entertaining, less boring, and more anxiety provoking (for each item, $p < .001$).

Corroborating the above findings, males more frequently reported either a partial or full erection, $F(1, 40) = 23.33$, $p < .001$, and greater physical excitement, $F(1, 40) = 27.97$, $p < .001$, in response to those stimuli chosen to elicit high (as opposed to low) levels of positive sexual arousal. In addition, males reported feeling more negative, unpleasant, uncomfortable, sad, and bad (for each item, $p < .005$) following exposure to those stimuli selected to elicit more intense negative affect.

Taken collectively, the findings described above indicate that the four sets of stimuli can be used effectively to create the intended factorial variation in positive sexual arousal and negative affect.[3]

Main Experiment

Method

Subjects

Participants were 95 males who were at least 18 years of age and were attending a major midwestern university. Fifty-one males who had not taken a course in psychology and had not participated in a previous psychology experiment were recruited from classes in introductory communication. Each student received extra credit in his communication class in exchange for his participation. Since only a limited number of students were available via communication classes, an additional 44 males were recruited by means of advertisements in the student news-

[2] All of the multivariate analyses of variance reported in the pilot study and in the main experiment used Pillai's trace criterion, described by Morrison (1967), as the multivariate test statistic.

[3] As a part of the pilot study, subjects' dispositional affective orientation to sexuality was measured by the Sexual Opinion Survey (White, Fisher, Byrne, & Kingma, Note 1). Validation studies have indicated that subjects who score high on this scale (erotophiles) report more intense sexual arousal, greater positive affect, and less negative affect following exposure to erotica, relative to those subjects who score low (erotophobes). Analysis of responses to the Feelings Scale, which included this measure, revealed that the desired factorial variation in positive sexual arousal and negative affect was evident for both erotophiles and erotophobes. Thus, the experimental manipulations appear to be generally effective across individual differences in affective orientation to sexuality.

paper, in which it was stipulated that volunteers should not have taken a course in psychology and that each volunteer would be paid $2.00.[4]

Design and Apparatus

The main experiment was arranged as a $2 \times 2 \times 2$ between-subjects factorial with factors corresponding to two levels of provocation (angry, not angry) and the intended characteristics of the experimental stimuli, positive sexual arousal (high, low) and negative affect (high, low). One group of angered subjects was also included who were not shown any of the four sets of stimuli prior to the opportunity to aggress. Participants were randomly assigned to each of the nine cells with the stipulation that each group ($n = 10$) contain an approximately equal number of paid volunteers and students from communication classes.

The apparatus consisted of a modified Buss (1961) "aggression machine," lists of 40%–60% Glaze association nonsense syllables, and a bogus intercom system. Ten push-button switches numbered from 1 to 10 were arranged in a row across the front of the aggression machine and were ostensibly to be used to deliver shocks of varying intensity to another person. A Lafayette stop clock (Model 54015) was used to record shock duration.

Procedure

The procedure followed closely that described by Baron and Bell (1977) and as a result will only be briefly outlined here.

After obtaining informed consent, the male experimenter escorted both the subject and the confederate to the experimental room, where it was explained that they would work on an impression-formation task. For purposes of this task, both individuals began by writing a brief summary of the central aspects of their own personalities, then examined each other's descriptions, and finally exchanged their mutual impressions of each other.

Anger manipulation. As in previous studies (Baron, 1978; Baron & Bell, 1977), the confederate's rating of the subject was prepared in advance by the experimenter in order to manipulate the subject's anger toward the confederate. In the *angry* condition, the personality rating ostensibly provided by the confederate was very unfavorable and quite insulting. In the *nonangry* condition, the rating by the confederate was more favorable. To enhance the anger manipulation still further, the confederate described himself as being "rather intolerant of others" in the angry condition and, in a more neutral manner, as "an average sort of guy" in the nonangry condition.

Aggression phase. Following the impression-formation task, the subject and the confederate were seated in a second room where the aggression machine was located. It was then explained that the second portion of the experiment would now begin, and was to be concerned with the effects of exposure to unpleasant stimuli on physiological responses. During this part of the experiment, one person would serve as a *responder* and receive a series of electric shocks delivered by the *stimulator.* Through the use of a rigged drawing, the confederate was designated to be the responder and the subject to be the stimulator.

The responder was then taken to an adjacent room, where the experimenter attached two sets of electrodes to the confederate's hand and wrist in full view of the subject. The electrodes were ostensibly to be used to record the responder's physiological responses and for the delivery of finger shock. The recording electrodes, in turn, were attached to a polygraph, which the subject could see.

The experimenter then proceeded to explain to the subject his role as the stimulator. It was initially made clear that the responder would be trying to memorize a list of eight paired-associate trigrams, which the stimulator would administer to him using the intercom located next to the aggression machine. Presumably, the use of this memory task would allow the experimenter to obtain more accurate evidence concerning the effects of unpleasant stimuli on physiological responses by effectively distracting the responder, and thereby preventing him from trying to "get ready" for each shock. Following these remarks, the experimenter informed the subject that at certain times a red signal light would be illuminated, at which time he (the stimulator) was to deliver an electric shock to the responder by depressing one of the 10 buttons on the aggression machine. It was made clear that the occurrence of the red signal light was random and was in no way related to the responder's performance on the paired-associates task. In order to familiarize the subject with the levels of shock supposedly to be used, sample shocks from buttons 4 and 5 were administered to the stimulator.

Presentation of the experimental stimuli. Following the sample shocks, the subject was told that there would be a short delay before continuing the experiment, in order to insure that the responder's physiological responses had returned to baseline levels. To fill this waiting time, the experimenter indicated that he had been asking subjects to rate some slides that he was planning to use in a future study. All subjects agreed to view and rate the stimuli, and were then shown one of the four sets

---

[4] Responses from five subjects were not included in the data analysis. Four subjects (two angered and two nonangered) were suspicious of some aspect of the experimental manipulations, and one subject was dismissed who failed to follow instructions properly.

In addition, since the results of a preliminary analysis revealed that the type of incentive used to recruit subjects (i.e., paid vs. course credit) had no significant effect on either the intensity or duration of shocks delivered to the confederate, this classification variable will not be mentioned further.

of 10 color slides described in the pilot study. Each slide was automatically projected once onto a screen for 20 seconds. Two seconds elapsed between the offset of one slide and the onset of the next slide. Following presentation of the stimuli, subjects completed a Feelings Scale (cf. Pilot Study).

While the subject was viewing the stimuli and reporting his reactions, the experimenter waited in the next room, where the confederate was located. Two minutes after the slides were presented, the experimenter informed the subject that the responder was ready. The experimental trials were then initiated. The red shock signal was given 20 times during the session. Both shock intensity and shock duration were recorded on each of the 20 trials.

No-exposure control. Ten angered subjects did not view or rate any of the four sets of stimuli. Individuals in this group were told that the experimenter would use the available free time while the responder's responses were stabilizing to adjust the polygraph, and no mention was made of the slides. The experimenter then returned to the room where the confederate was seated, while the subject sat and waited. Three minutes later the experimenter returned, informed the subject that the responder was ready, and then initiated the experimental trials.

Postexperimental questionnaire and debriefing. Following the last shock trial, subjects completed a questionnaire on which they reported (on 10-point scales) if they thought that there was a certain number of high (and low) intensity shocks that they were supposed to give (definitely yes/definitely not); if they tried to deliver shock as they thought others might have done (definitely yes/definitely not); their level of anger toward the confederate (angry/not angry); their feelings toward the confederate (comfortable/uncomfortable, pleasant/unpleasant, positive/negative); and how anxious they were for the experiment to end (not at all anxious/very anxious). A brief verbal suspicion check was then administered, followed by an extensive debriefing.

Results

Affect and Arousal Responses to the Experimental Stimuli

Statistical procedures used to determine if the experimental stimuli produced the intended differentiations in affect and arousal response followed closely those used for this purpose in the pilot study. As in the pilot study, subjects' ratings of themselves as sexually aroused, not bored, entertained, excited, and anxious, loaded on a common factor, and were combined to form a single multivariate F value, which was used to assess the effectiveness of the manipulation of positive sexual arousal. Analysis yielded the required differences, showing only a main effect of positive sexual arousal, multivariate $F(5, 68) = 9.37$, $p < .001$. Participants reported being more sexually aroused, entertained, excited, less bored (all $p < .001$), and more anxious ($p < .02$) following exposure to sets of stimuli chosen to elicit high (as opposed to low) positive arousal: $M = 3.44$ for high sexual arousal, and $M = 1.85$ for low sexual arousal, on a 7-point scale.

Next, subjects' responses to items relating to disgust, anger, nausea, guilt, depression, and self-disappointment were combined to form a single multivariate F value that was used to determine the effectiveness of the manipulation of negative affect. Responses to these six items each shared high loadings on a single factor, as in the pilot study. Analysis revealed only the desired main effect of negative affect, multivariate $F(6, 67) = 6.41$, $p < .001$. Responses to those stimuli chosen to elicit high (rather than low) negative affect were most strongly differentiated with respect to feelings of disgust ($p < .0001$): $M = 3.90$ for high negative affect, and $M = 1.70$ for low negative affect, on a 7-point scale. In addition, subjects reported greater nausea ($p < .002$), depression ($p < .002$), guilt ($p < .01$), anger ($p < .02$), and self-disappointment ($p < .10$) in response to those stimuli chosen to evoke more intense negative feelings.

Taken as a whole, the results presented above, in conjunction with those of the pilot study, strongly indicate that the experimental manipulations were effective.

Aggressive Behavior

Subjects' shock intensity, shock duration, and a transformed Shock Intensity × Shock Duration score (composite measure) were combined to form a single multivariate F value, which was interpreted as an index of aggressive behavior toward the confederate.[5] Table 2 presents the mean level of aggression (for all three measures) in each of the nine

[5] The transformed Shock Intensity × Shock Duration measure has been used in past studies of human aggression (e.g., Baron, 1978; Baron & Bell, 1977). This composite measure was calculated from subjects' shock intensity and shock duration scores using a simple transformation, $X = (x + 1)^{\frac{1}{4}} + (x)^{\frac{1}{4}}$, where x equals the product of shock intensity and shock duration.

Table 2
Aggression as a Function of Exposure to Erotic Stimuli and Prior Provocation

Condition	Shock intensity		Shock duration (sec)		Composite measure	
	Nonangry	Angry	Nonangry	Angry	Nonangry	Angry
High sexual arousal/ high negative affect (ambivalent response)	4.66_{ab}	5.29_{ac}	$.46_{bc}$	$.72_{a}$	3.23_{bc}	4.06_{a}
High sexual arousal/ low negative affect (positive response)	3.95_{bc}	3.86_{b}	$.36_{b}$	$.44_{bc}$	2.68_{b}	2.84_{bc}
Low sexual arousal/ high negative affect (negative response)	3.22_{b}	5.28_{ac}	$.29_{b}$	$.64_{ac}$	2.37_{b}	3.86_{ac}
Low sexual arousal/ low negative affect (neutral response)	4.43_{ab}	5.57_{a}	$.36_{b}$	$.51_{ab}$	2.84_{bc}	3.61_{ac}
No-exposure control[a]		5.11		.44		3.26

Note. $n = 10$ within each cell. A higher number indicates more intense shock (on a 10-point scale) and shock of longer duration (range = .01 to 2.68 sec). For each dependent measure within the eight basic experimental groups, means that do not share a common subscript differ significantly at the .05 level by Duncan's multiple-range test.

[a] For shock intensity, high sexual arousal/low negative affect versus no-exposure control, $F(1, 27) = 2.26$, $p < .05$ (one-tailed), by Dunnett's test.

experimental groups. Multivariate analysis of variance applied to aggression scores yielded a significant main effect of provocation, multivariate $F(3, 70) = 4.67$, $p < .005$, and a main effect due to negative affect, multivariate $F(3, 70) = 3.02$, $p < .05$, which was further qualified by a significant interaction between negative affect and positive sexual arousal, multivariate $F(3, 70) = 2.82$, $p < .05$.

As an initial step in further probing these significant findings, the main effect due to provocation was examined more closely. Univariate tests revealed that individuals who received a negative evaluation from the confederate (relative to nonangered subjects) evidenced stronger aggression on each of the three measures of such behavior: $F(1, 72) = 8.24$, $p < .01$, for shock intensity; $F(1, 72) = 11.61$, $p < .001$, for shock duration; and $F(1, 72) = 14.37$, $p < .001$, for the composite measure. Thus, it appears that the manipulation of anger was generally effective.

Turning next to the effect of exposure to the various types of erotic stimuli on aggression, univariate tests revealed that the main effect due to negative affect failed to attain significance on any of the three dependent measures of aggression: Fs$(1, 72) = .21$, 3.10, and 3.24, all ps $> .05$, for shock intensity, shock duration, and the composite measure, respectively. A closer examination of the significant multivariate interaction between positive sexual arousal and negative affect, however, yielded a parallel univariate effect for shock intensity, $F(1, 72) = 6.86$, $p < .05$, and the composite measure, $F(1, 72) = 5.51$, $p < .05$, but not for shock duration, $F(1, 72) = 1.71$, $p > .20$. Simple effects tests were then used to probe differences among the four group means comprising these significant univariate interactions. Analysis revealed that aggressive behavior toward the confederate was reduced by exposure to erotic stimuli that evoked primarily positive responses, relative to the effect of exposure to neutral material; $F(1, 72) = 4.93$, $p < .05$, for shock intensity: and relative to ambivalent stimuli; $F(1, 72) = 4.74$, $p < .05$, for shock intensity; and $F(1, 72) = 8.60$, $p < .01$, for the composite measure.

Because further inspection of the means presented in Table 2 suggested that a more

complex pattern of results might underlie the interactive effects presented above, the means of all eight groups were compared (for all three measures) using Duncan's multiple range tests. As can be seen in Table 2, the results of these comparisons revealed that exposure to erotic stimuli that evoked primarily positive affect reduced retaliatory behavior by angered individuals ($p < .05$), but had only weak or negligible effects on aggressive behavior by nonangry persons. Thus, the effect of exposure to the various types of erotic stimuli on subsequent aggressive behavior toward the confederate was most clearly evident for angered males.

Finally, Dunnett's test (Dunnett, 1955) was used to compare each of the four group means within the angry condition against the no-exposure control. The results of this analysis indicated that the intensity of aggressive behavior was inhibited, $F(1, 27) = 2.26$, $p < .05$ (one-tailed test), by exposure to erotic stimuli that led primarily to increases in positive affect. No further comparisons involving the no-exposure control attained significance.

In summary, aggressive behavior toward the confederate was inhibited by exposure to sexually arousing erotic stimuli that evoked primarily positive affect and slightly enhanced by increases in sexual arousal that were responded to with ambivalence.

Postexperimental Questionnaire

Analysis of responses to the postexperimental questionnaire indicated that males who received a negative evaluation from the confederate (as opposed to a neutral one) rated themselves as more angry, $F(1, 72) = 93.17$, $p < .001$; reported being more negative, $F(1, 72) = 32.91$, $p < .001$, uncomfortable, $F(1, 72) = 104.16$, $p < .001$, and unpleasant, $F(1, 72) = 98.19$, $p < .001$; and found their experimental participation to be a less enjoyable experience, $F(1, 72) = 5.91$, $p < .05$. These findings amplify the results described above showing greater aggression by angered males than by nonangry individuals.

Further examination of responses to the postexperimental inventory revealed that participants reportedly felt there was not a correct number of high intensity ($M = 7.15$) or low intensity ($M = 7.21$) shocks that they were supposed to administer to the confederate and reported no obvious intention ($M = 5.71$) to deliver shock in a manner that mimicked the behavior of other participants. Taken as a whole, subjects' mean responses to these items and the failure to find any significant effects for these items due to the independent variables, suggest that the experimental manipulations did not operate simply to evoke an expectation that one pattern of shock was more correct or desirable than another.

Discussion

It appears that exposure to sexually arousing erotic stimuli may sometimes inhibit aggression, and at other times slightly increase such behavior, depending on how exposure to such material makes one feel. In the present study, there was a slight tendency for aggressive behavior to be enhanced by exposure to arousing erotic stimuli that were reported to be disgusting and unpleasant. In contrast with this group, subjects' exposure to sexual imagery reported to be entertaining and positively exciting decreased retaliatory behavior by angered individuals to a level that fell slightly below that of subjects who were never provoked. Further comparison with a baseline control group strongly suggested that exposure to such erotica acted to inhibit aggressive behavior by angry males.

The inhibition of motivated aggression by exposure to positively arousing erotic stimuli observed in the present study may be viewed as corroborating and extending earlier findings (e.g., Baron, 1974a, 1974b; Baron & Bell, 1977; Zillmann & Sapolsky, 1977). Moreover, since only those stimuli which were labeled and experienced as entertaining and exciting led to reductions in retaliatory behavior, it appears that subjects' affective response to such material is a critical factor leading to the inhibition of aggression by erotica.

Alternative Explanations for the Major Findings

Attentional shift. The pattern of results described above may be viewed as generally

failing to support the hypothesis (e.g., Donnerstein, Donnerstein, & Evans, 1975) that exposure to erotic stimuli inhibits aggressive behavior by shifting subjects' attention away from the source of their anger. Based on the arguments presented by Donnerstein et al., it would seem reasonable to expect that exposure to virtually any type of mildly arousing erotic stimuli would be more distracting than sexually arousing and, therefore, would be effective in reducing subsequent aggression. In the present study, however, exposure to seemingly attention absorbing erotic imagery that led to heightened negative affect not only failed to reduce aggression but tended slightly to enhance subsequent aggressive behavior.

Cognitive labeling. One explanation for the results of the present study may be that subjects' contrasting affective responses to arousing erotic imagery led to different levels of aggression by modifying the type of label males applied to any arousal they experienced (Baron, 1977; in press). This interpretation is in keeping with other work (e.g., Geen, Rakosky, & Pigg, 1972; Zillmann & Bryant, 1974) suggesting that whether or not a given source of arousal will act to facilitate or to inhibit subsequent aggression depends on a person's cognitions regarding the source of his or her arousal. Thus, in the present study, following exposure to erotic stimuli that evoked primarily positive responses, males may have been more inclined to label any arousal they experienced in a positive manner (e.g., as sexual excitement or entertainment, rather than as anger), with the result that subsequent retaliatory behavior was reduced. In contrast, following exposure to sexual material that was reported to be unpleasant and disgusting, subjects may have been more willing to interpret any arousal they experienced as anger (rather than resulting from exposure to the erotic stimuli), with the result that aggressive behavior toward the confederate was slightly facilitated, or at least was not inhibited.

The cognitive interpretation presented above assumes that the facilitation of aggression by erotica requires an increase in arousal deriving from exposure to such stimuli and the adoption of a cognitive label appropriate to the expression of aggression (cf. Zillmann & Sapolsky, 1977). In line with this view, simple increases in negative affect resulting from exposure to erotica (negative response) were not found to enhance subsequent aggression in the present study. Continuing this line of reasoning, it is less clear, however, why exposure to reportedly ambivalent material that led to increases in both sexual arousal and negative affect failed to increase subsequent aggressive behavior more significantly. While a number of reasons may be offered to explain this result (e.g., dissipation of negative affect after rating the stimuli), it seems reasonable to suggest that the ambivalent erotic imagery used in the present study simply failed to produce a sufficient degree of sexual arousal (cf. Baron & Bell, 1977). As such, exposure to this set of stimuli may not have generated a sufficiently intense excitatory reaction (i.e., arousal response) whose residue could identify subsequent aggressive behavior (cf. Zillmann, 1971). In support of this interpretation, a closer examination of the present results revealed that angered individuals who reported the highest levels of sexual arousal following exposure to ambivalent stimuli were subsequently found to be the most aggressive: $t(8) = 3.20$, $p < .02$, for shock intensity. In general, however, subjects reported only moderate levels of sexual excitement following exposure to such material ($M = 3.25$, on a 7-point scale) and did not exhibit heightened aggressive attacks against the confederate.

Incompatible responses. Although ambivalent erotic depictions may not have been sufficiently arousing to enhance subsequent aggression, exposure to affectively positive and arousing erotic stimuli was found to inhibit aggression as was predicted. As discussed above, while exposure to such erotica may have effectively reduced aggressive behavior by affording subjects more positive cognitive labels, other, though not necessarily incompatible, explanations are also possible. In particular, positive responses of pleasure, entertainment, and excitement deriving from exposure to sexual stimuli may be hedonically incompatible with feelings of anger and/or the performance of aggressive acts and con-

sequently may inhibit such behavior directly (e.g., Baron, 1974a, 1974b, 1976, 1978; Zillmann & Sapolsky, 1977). That is, once an angry individual becomes immersed in pleasant and entertaining erotic stimuli, it may be difficult to maintain heightened feelings of annoyance or to engage in intense retaliatory behavior. Following the implications of this analysis, it may be that exposure to virtually any pleasant stimulus (e.g., art, scenery) is capable of reducing aggressive attacks by angry individuals. In line with this interpretation, exposure to enjoyable and entertaining cartoons has been found (Baron, 1978; Baron & Ball, 1974) to reduce aggressive behavior by angry persons.

Conclusion

In summary, the present study serves to extend previous research by demonstrating that the effect of exposure to erotic stimuli on subsequent aggression depends, to an important extent, on the type of affective response elicited by exposure to such stimuli.[6] One useful direction for extending the results of the present study and related research, would appear to be to explore further the efficacy of training individuals who have difficulty controlling their anger to emit pleasant images and other incompatible behaviors in response to provocation, as a means of inhibiting anger and reducing the intensity of retaliatory behavior (cf. Baron, 1977; Novaco, 1977; Smith, 1973).

[6] Out of convenience, male subjects and male confederates were used in the present study, which leaves open the question of whether exposure to the various types of erotic stimuli used in the present study would have similar effects on aggressive behavior toward males and females and/or by females toward these target persons. Although empirical research that examines the combinations of aggressors and victims described above may be required to answer these questions definitively for specific types of stimuli, a number of studies are available that may provide at least a partial answer to these concerns. With respect to the sex of the victim, exposure to erotic material has been found (Donnerstein & Barrett, 1978; Jaffe, Malamuth, Feingold, & Feshbach, 1974) to have similar effects on aggressive behavior toward males and females. Further, it has been argued (Baron, in press) that exposure to sexual imagery that evokes similar patterns of affect and arousal responses in males and females will lead to similar levels of subsequent aggression. Thus, on the basis of the evidence available, it may be tentatively assumed that the conclusions of the present study would apply independently of the sex of the aggressor or the sex of the victim. It should be emphasized that the specific characteristics of those stimuli (e.g., thematic content) which evoke positive, negative, ambivalent, and neutral responses may differ greatly for males and females (e.g., Schmidt & Sigusch, 1970).

Reference Note

1. White, L. A., Fisher, W. A., Byrne, D., & Kingma, R. *Development and validation of a measure of affective orientation to erotic stimuli: The Sexual Opinion Survey.* Paper presented at the meeting of the Midwestern Psychological Association, Chicago, May 1977.

References

Baron, R. A. The aggression-inhibiting influence of heightened sexual arousal. *Journal of Personality and Social Psychology,* 1974, *30,* 318–322. (a)

Baron, R. A. Sexual arousal and physical aggression: The inhibiting influence of "cheesecake" and nudes. *Bulletin of the Psychonomic Society,* 1974, *3,* 337–339. (b)

Baron, R. A. The reduction of human aggression: A field study of the influence of incompatible reactions. *Journal of Applied Social Psychology,* 1976, *6,* 260–274.

Baron, R. A. *Human aggression.* New York: Plenum Press, 1977.

Baron, R. A. Aggression-inhibiting influence of sexual humor. *Journal of Personality and Social Psychology,* 1978, *36,* 189–198.

Baron, R. A. Heightened sexual arousal and physical aggression: An extension to females. *Journal of Research in Personality,* in press.

Baron, R. A., & Ball, R. L. The aggression-inhibiting influence of nonhostile humor. *Journal of Experimental Social Psychology,* 1974, *10,* 23–33.

Baron, R. A., & Bell, P. A. Sexual arousal and aggression by males: Effects of type of erotic stimuli and prior provocation. *Journal of Personality and Social Psychology,* 1977, *35,* 79–87.

Buss, A. H. *The psychology of aggression.* New York: Wiley, 1961.

Byrne, D., Fisher, J. D., Lamberth, J., & Mitchell, H. E. Evaluations of erotica: Facts or feelings? *Journal of Personality and Social Psychology,* 1974, *29,* 111–116.

Donnerstein, E., & Barrett, G. Effects of erotic stimuli on male aggression toward females. *Journal of Personality and Social Psychology,* 1978, *36,* 180–189.

Donnerstein, E., Donnerstein, M., & Evans, R. Erotic stimuli and aggression: Facilitation or inhibition.

Journal of Personality and Social Psychology, 1975, *32*, 237–244.

Dunnett, C. W. A multiple comparison procedure for comparing several treatment means with a control. *Journal of the American Statistical Association*, 1955, *50*, 1096–1121.

Geen, R. G., Rakosky, J. J., & Pigg, R. Awareness of arousal and its relation to aggression. *British Journal of Social and Clinical Psychology*, 1972, *11*, 115–121.

Jaffe, Y., Malamuth, N., Feingold, J., & Feshbach, S. Sexual arousal and behavioral aggression. *Journal of Personality and Social Psychology*, 1974, *30*, 759–764.

Kaiser, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 1958, *22*, 187–200.

Morrison, D. F. *Multivariate statistical methods*. New York: McGraw-Hill, 1967.

Novaco, R. W. Stress inoculation: A cognitive therapy for anger and its application to a case of depression. *Journal of Consulting and Clinical Psychology*, 1977, *45*, 600–608.

Rule, B. G., & Nesdale, A. R. Emotional arousal and aggressive behavior. *Psychological Bulletin*, 1976, *83*, 851–863.

Schmidt, G., & Sigusch, V. Sex differences in responses to psychosexual stimulation by films and slides. *Journal of Sex Research*, 1970, *6*, 268–283.

Smith, R. E. The use of humor in the counterconditioning of anger responses: A case study. *Behavior Therapy*, 1973, *4*, 576–580.

Zillmann, D. Excitation transfer in communication-mediated aggressive behavior. *Journal of Experimental Social Psychology*, 1971, *7*, 419–434.

Zillmann, D., & Bryant, J. Effects of residual excitation on the emotional response to provocation and delayed aggressive behavior. *Journal of Personality and Social Psychology*, 1974, *30*, 782–791.

Zillmann, D., & Sapolsky, B. S. What mediates the effect of mild erotica on hostile behavior in males? *Journal of Personality and Social Psychology*, 1977, *35*, 587–596.

Received April 24, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 4, 602-607

Predictions of Others' Responses in a Mixed-Motive Game: Self-Justification or False Consensus?

Lawrence A. Messé and John M. Sivacek
Michigan State University

This study examined speculations by Dawes, McTavish, and Shaklee concerning the extent to which persons use their own responses in mixed-motive situations as a basis for predicting the behavior of others. Female subjects played a one-trial Prisoner's Dilemma game and predicted the responses of their partner and a person in another dyad. As predicted, subjects, irrespective of the particular choice they had made, tended to attribute their own response to others. In many subjects, this attribution appeared to be self-justifying, in that it was specific to the subjects' partner. However, it was even more frequently the case that subjects attributed their own response to both partner and nonpartner (i.e., a false consensus), although they also tended to be more confident about their prediction of their partner's choice.

Dawes, McTavish, and Shaklee (1977) recently speculated that behavior in some mixed-motive situations might not be a consequence of an attribution about the other(s), but, rather, that the reverse relationship might be operating—that is, the attribution about the other's behavior is a consequence of the subject's own behavior. Dawes et al. used a multiperson one-trial dilemma to examine the validity of their position. They compared the attributions to others that were made by participants in a commons dilemma to the attributions that were made by observers who did not participate in the situation but merely predicted the behavior of those who did. They expected that participants would be more variable in their attributions than would observers, given that the former should base their judgments of others on their own behavior. Their findings did support this behavior-to-attribution hypothesis.

Unfortunately, the methodology that Dawes et al. employed did not permit the examination of more specific components of their argument. For example, they speculated that the attributions of persons are influenced by their own actions, irrespective of whether they compete or cooperate. While Dawes et al.'s results suggest that this was the case in their study, their design and analysis did not permit separate examination of the data of the subjects who cooperated and of those who competed. Moreover, prior studies that have examined this hypothesis with encounters of short duration (e.g., Kanouse & Wiest, 1967; Terhune, 1968) have yielded somewhat inconsistent findings, although, on the whole, they support the notion that there is a great deal of correspondence between a subject's choice and the behavior he or she attributes to others.

Dawes et al. (1977) also suggested—but did not examine—the notion that one or both of two mechanisms might account for the hypothesized tendency in subjects in mixed-motive situations to base their judgments about others on their own behaviors. They suggested that subjects might engage in this attributional process in order to justify their responses; subjects who compete might feel the need to justify what could be considered ex-

This research was conducted while the second author was a National Institute of Mental Health trainee in social psychology. We are grateful to David Lovell, John Lovell, and Steven Fox for their assistance in carrying out this research.

Requests for reprints should be sent to Lawrence A. Messé, Department of Psychology, Snyder Hall, Michigan State University, East Lansing, Michigan 48824.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3704-602$00.75

ploitative and antisocial behavior, and subjects who cooperate might attribute a tendency toward similar behavior to others in order to "avoid feeling duped." In addition, they suggested that subjects might use their own behavior as a guide for guessing "what everybody does." This latter explanation has been termed the "false consensus" phenomenon (see Ross, Green, & House, 1977).

It should be noted that Dawes et al. employed the concept of self-justification somewhat narrowly. Ross et al. (1977), in their discussion of false consensus judgments, suggested that such behavior could be based on a number of underlying processes, including the need to justify oneself. Clearly, if self-justification is a salient motive, it is most efficiently expressed by attributing one's own behavior only to one's partner. But, in a larger sense, this narrow attribution might not be sufficient, perhaps because one needs to believe that one's partner (who often is a stranger) is just like everyone else, or perhaps because there is greater comfort in thinking that one is not atypical in general. On the other hand, other explanations for the false consensus phenomenon also are plausible—for example, that one uses his or her own behavior as the best evidence (perhaps as the only evidence) of what people in general do. In any event, it seemed of interest to attempt to examine the extent to which persons in short-duration mixed-motive situations appear to use their predictions of others to self-justify in the narrow sense suggested by Dawes et al. (1977) or to make more general predictions about the behavior of others, for whatever reason.

Thus, the present study examined the behavior and predictions of subjects in a one-trial dilemma to explore some of the issues proposed but left unresolved by Dawes et al. First, we examined the extent to which both persons who competed and persons who cooperated attributed their own responses to other participants. Second, we used a procedure that permitted differentiation between the false consensus and (narrow) self-justification hypotheses. In the present study, all subjects served as both participants and observers. Thus, they responded in a Prisoner's Dilemma and made predictions about the responses of both their partner and a person with whom they were not paired. This within-subjects procedure permitted examination of both the self-justification and false consensus explanations. To the extent that self-justification is a valid explanation, we would expect that subjects would attribute to their partner a response similiar to their own, but their response would not be related to their prediction about the person with whom they were not paired. On the other hand, to the extent that false consensus is a relevant phenomenon, we would expect that subjects would be likely to attribute responses that were similar to their own to both their own partner and the individual with whom they were not paired.

Method

Subjects

One hundred seventy-two female students from introductory psychology classes at Michigan State University volunteered to participate in the research for extra course credit. In addition, as explained below, they also received up to 80¢ each, depending upon the decision that they and their partners made when playing a one-trial Prisoner's Dilemma (PD) game.

Design

In addition to making a "cooperative" or "competitive" choice in the PD game, subjects also made predictions about their partner's (P) choice and the choice of another person (O) with whom they were not paired. The order in which they made these decisions was varied such that half the subjects made their own choice first, whereas the remainder first made a prediction about the decision of one of the other people. Also, the order of the predictions was varied such that half the subjects predicted their partner's choice first, whereas the other half predicted the nonpartner's choice first. This procedure generated a design whose dimensions were 2 (subjects' choice: cooperate or compete) × 2 (order of subjects' choice: before or after their first prediction) × 2 (order of predictions: about P first or about O first). The latter two factors were included primarily for reasons of experimental control. In actuality, we were primarily interested in (a) the subjects' own choice and (b) the extent to which they attributed the same choice to P and/or O.

Procedure

Subjects were examined in groups of four. When the four subjects in a session had assembled, each was given 30¢ and was paired at random with a male experimenter. One of the experimenters shuffled

Table 1
Outcome Values (in cents) in Prisoner's Dilemma Game

Person A's choice	Person B's choice	
	Cooperate	Compete
Cooperate	25, 25	−25, 50
Compete	50, −25	−10, −10

Note. The first value in each cell is the outcome to Person A.

a deck of four cards (numbered 1, 2, 3, and 4) and dealt one face down to each subject. Then, each experimenter escorted the subject assigned to him to one of four cubicles. From this point on, subjects had no further contact with each other.

Once in the cubicle, the experimenter seated the subject at a small table and presented her with a "typical" 2 × 2 PD matrix. (Table 1 presents the values that were contained in the matrix.) The experimenter explained in detail the nature of the one-trial PD game—there are two players, each has two choices, and so on. At the end of his explanation, which was a script that he had memorized, the experimenter made sure that the subject understood the nature of the game by soliciting questions and reiterating points about which the subject appeared uncertain. The experimenter then informed the subject that the four of them had been divided at random into two pairs and that she had been paired with the person who had drawn "Card Number *X*" (where *X* was a number on one of the cards that had been dealt to the other subjects).

At this point, depending upon the condition to which she had been randomly assigned, the subject was asked (a) to make a choice of PD alternatives—with the understanding that her choice, coupled with that of her partner, actually determined payment to both—or (b) to predict her partner's PD choice, or (c) to predict the PD choice of one of the persons in the other pair. In addition, whenever a subject made a prediction about her partner's or the other person's choice, she was asked immediately afterwards to indicate on a scale of 50 to 100 how confident she was in her prediction. She was told to mark a 50 (indicating a "50–50 guess") if she was merely guessing, a 100 if she was completely confident, or an appropriate number in between to indicate an intermediate degree of confidence. Depending on the nature of the first task, the subject then was asked either to make her own choice and the remaining prediction (and confidence rating) or to make the two predictions (and confidence ratings) in the order dictated by the experimental design.

It should be noted that the subject was not informed that she was to make a series of decisions; that is, she was told about the second task only after she had completed the first, and about the third task only after she had completed the second. It also should be noted that the subject made her judgments in private. The experimenter did not observe her while she was making her decision or her predictions and confidence ratings. In each instance, she was told to fold the card on which she had marked her response(s) before returning it to the experimenter.

When the subject had completed the three tasks, the experimenter excused himself for a moment to consult with the experimenter who had been assigned to her partner. Together they determined the outcome of the one-trial PD game for the pair. Each then returned to his respective subject and adjusted her payment accordingly. Every subject was debriefed, sworn to secrecy, and thanked for her participation. A subject was allowed to leave in such a way that she did not come into contact with the other participants in her session.

Results

The subjects' own PD choice and their predictions about the choices of P and O were examined via chi-square analyses. Similarly, their confidence ratings were examined through analysis of variance (ANOVA). Preliminary analyses were performed to explore the possibility that the order in which subjects performed the three tasks influenced their responses. Then, the data were cast and analyzed in a manner that most directly addressed Dawes et al.'s (1977) speculations.

Examination of Possible Order Effects

Three chi-square tests were calculated to explore the possibility that the four combinations of order of own choice and order of predictions about others' choices influenced the subjects' responses in the PD game, their predictions about their partners' choices, or their predictions about the choice of one person in the other dyad. None of the obtained chi-squares reached significance, χ^2s(3) = 1.03, 1.08, 4.88, respectively, ps > .20. Likewise, a preliminary ANOVA indicated that subjects' relative confidence in a prediction was not affected by order of PD decision, $F(1, 157) = .12$, $p > .70$; order of predictions, $F(1, 157) = .70$, $p > .40$; or their interaction, $F(1, 157) = 0$, $p > .99$.[1] Thus, order of responses was

[1] Note that 7 subjects had missing data, in that they did not give one or both confidence ratings and/or experimenters failed to record their order conditions. Thus, the total sample size for these data was 165.

disregarded in subsequent analyses. It is likely that response orders had little effect, because subjects probably rehearsed their own choice even before they were explicitly told that they were to make one.

Tests of Dawes et al.'s Speculations

As noted above, Dawes et al. (1977) speculated that subjects predict that others will act in a PD situation as the subjects themselves did, irrespective of the specific choice that the subjects made (i.e., irrespective of whether they competed or cooperated). In addition, they also speculated that the tendency to attribute one's own behavior to others might be due to one or both of two mechanisms, self-justification and false consensus. The data and analyses presented below bear directly on these speculations.

Predictions of subjects who cooperated or competed. Table 2 presents the proportions of the 73 subjects who chose cooperatively and the 99 subjects who chose competitively who attributed similar or dissimilar responses to their partner and to the person in the other dyad.[2] These data provide strong support for Dawes et al.'s (1977) speculation, since there was little difference in the frequencies with which persons who cooperated and persons who competed attributed their response to their partner, $\chi^2(1) = .39$, or to the other person, $\chi^2(1) = .92$. In addition, there was a strong tendency for subjects to attribute the same response as their own to their partner, $\chi^2(1) = 38.82$, $p < .001$, and, to a somewhat lesser extent, to the other person, $\chi^2(1) = 6.44$, $p < .025$.

Table 2
Proportions of Competitive and Cooperative Subjects Who Predicted That Others Would Make the Same or a Different Response as Their Own

Subject's response	Predicted response: Same as own	Predicted response: Different from own
Prediction about partner		
Competition	.72	.28
Cooperation	.76	.24
Total sample	.74	.26
Prediction about person in other dyad		
Competition	.63	.37
Cooperation	.56	.44
Total sample	.60	.40

Table 3
Proportions of Subjects Who Made Similar or Dissimilar Predictions About Their Partner's and the Other's Responses

Prediction about other's response	Prediction about partner's response: Same as subject	Prediction about partner's response: Different from subject
Same as subject	.44	.16
Different from subject	.30	.10

Examination of the self-justification and false consensus predictions. Dawes et al. (1977) speculated that the attribution of one's own behavior to one's partner probably was based on one or both of two mechanisms: an attempt to justify one's decision and the tendency to use one's own behavior as the basis for generalizing about everyone's behavior. To examine the validity of these predictions, data were organized by tabulating the number of subjects who attributed their own behavior to both their partner and the other person, the number who attributed such behavior only to their partner, the number who attributed their own response only to the other, and the number who attributed their response to neither party. (Since there was no difference in the attributions of similarity by competitors and cooperators, the data were collapsed across this variable.) Table 3 presents these data (in the form of proportions).

A chi-square test of this array (McNemar's test for correlated proportions; see Hays,

[2] Note that data are presented in the tables in the form of proportions, to make the comparison of the data of subjects who cooperated and of those who competed more straightforward, given that competitive responses were made somewhat more frequently, $\chi^2(1) = 3.70$, $p < .075$. Of course, in every case the chi-square analyses were performed on the raw frequencies.

Table 4
Mean Confidence Ratings of Subject's Predictions of Partner's and Other's Responses

Prediction about partner	Confidence in prediction of	
	Partner	Other
Same response as subject (124)	69.38	63.93
Different response from subject (44)	62.92	64.94

Note. Scores could range from 50 (indicating no confidence) to 100 (indicating complete confidence). Numbers in parentheses indicate cell frequencies.

1963, p. 602) yielded a significant value, $\chi^2(1) = 7.29$, $p < .01$. Thus, there was a greater tendency to attribute similar behavior to one's partner than to a nonpartner. Although this finding provides support for the self-justification hypothesis, the false consensus hypothesis also appears to be supported to some extent as well, since the frequency with which subjects attributed their own behavior to both P and O was significantly greater than the frequency with which subjects attributed such behavior only to their partner, $\chi^2(1) = 4.50$, $p < .05$.

Analysis of Confidence Scores

Recall that subjects were asked to indicate how confident they were in each of their predictions. These data were analyzed via a 2 (prediction about partner) × 2 (prediction about other) × 2 (target of confidence rating: partner/other) ANOVA with repeated measures on the third factor.[3] This analysis yielded one significant effect: the Prediction About Partner × Target of Confidence Rating interaction, $F(1, 164) = 6.76$, $p < .01$. Table 4 presents the relevant cell means. Subsequent simple effects analyses (Winer, 1971, pp. 347–351) of these data revealed that subjects who predicted that their partners would behave as they did expressed more confidence in their prediction about their partner than in their prediction about the person in the other dyad, $F(1, 164) = 7.15$, $p < .01$, irrespective of whether they predicted that the nonpartner would respond similarly to, or differently from, them. However, subjects who predicted that their partner would choose differently from them were no more confident in their prediction about their partner than in their prediction about the other person, $F(1, 164) = .89$. Thus, these findings indicated that even those subjects who predicted that both P and O would choose as they did were more confident in their predictions about P than about O.

Discussion

The results provide strong support for Dawes et al.'s (1977) speculation that people predict that others will behave as they themselves did in one-trial PD situations, irrespective of whether their behavior was cooperative or competitive. Moreover, this attributional phenomenon appears to be due to both a need to justify one's response and a tendency to think of one's own behavior as typical of everyone. As noted above, however, Dawes et al. employed the concept of self-justification somewhat narrowly. It could be, in fact, that those subjects who attributed their own response to both partner and other did so in an attempt to justify their own behavior (e.g., "I'm just doing what my partner and everyone else in this situation would do"). The finding that these subjects had more confidence in their prediction of their partner's behavior than they had in their prediction of the other's behavior lends indirect support for this interpretation. This result suggests that subjects might have possessed a greater need to believe that their partner would choose as they did than to believe that someone whose behavior was irrelevant to them would do so.

Two related findings of the present research appear to merit special attention, since they suggest that, for whatever reasons, the judgments of subjects were not entirely rational. As noted above, subjects were more likely to predict that their partner would behave as they did than they were to attribute

[3] Preliminary analyses revealed no main effects or interactions for subjects' own choice (compete or cooperate). Thus, the data were collapsed on this factor. It should also be noted that 4 subjects failed to provide confidence scores for their predictions. This analysis, therefore, was based on data from 168 subjects.

this choice to the nonpartner; also, they had greater confidence in the attribution of a similar choice when they made this prediction about their partner. These results occurred in spite of the fact that the procedure used precluded the possibility that subjects could objectively differentiate their partner from the nonpartner. Recall that in a given session, all four subjects were led into separate cubicles, and they could know only that their partner and the other person about whom they made a prediction were two of the three people in the other cubicles. There was no way that they could identify these two people specifically. Thus, subjects had no objective basis for making differential predictions or confidence ratings; yet they tended to do so. These findings suggest that there is some within-subject process (or processes)—a need to justify one's own behavior, a need to feel in control (Langer, 1975), and so forth—that produce(s) the illusion in subjects that the two targets of their predictions are identifiably different.

The systematic study of the relationship between people's behaviors and their predictions of others' behaviors in mixed-motive situations began with Kelley and Stahelski's (1970) examination of the "triangle" hypothesis. They proposed that people who compete in mixed-motive situations do so because they have a biased perception that other people are competitive; on the other hand, cooperators are less likely to stereotype in this fashion, and thus, their behavior is influenced by the actual behavior of the other person. Dawes et al. (1977) proposed their explanation as a contrasting position to Kelley and Stahelski's triangle hypothesis. They concluded that their findings—and most likely they would conclude that our findings as well—tend to disconfirm Kelley and Stahelski's explanation. In fairness, however, it should be noted that this conclusion is somewhat extreme. While there is ample evidence (e.g., the present study; Dawes et al., 1977; Kanouse & Wiest, 1967; Miller & Holmes, 1975) that the triangle hypothesis is not valid in all mixed-motive situations, Kelley and Stahelski (1970) themselves noted that biased perceptions should affect responses only under limited circumstances. And there is evidence (e.g., their own study; Miller & Holmes, 1975) that the triangle hypothesis is valid in substance under the appropriate conditions (i.e., mixed-motive situations of extended duration in which contingent cooperation is both rational and congruent with social norms; see, for example, Luce & Raiffa, 1964, p. 101).

Thus, taken as a whole, the present and previous studies provide evidence that no single explanation can account for the link between behavior and attributions in all mixed-motive situations. Instead, the processes that underlie this linkage appear reasonably complex. Therefore, attempts to demonstrate the general utility of one perspective over another seem to be a less fruitful pursuit than are attempts to understand the conditions under which the different processes proposed by these perspectives operate.

References

Dawes, D. M., McTavish, J., & Shaklee, H. Behavior, communication, and assumptions about other people's behavior in a commons dilemma situation. *Journal of Personality and Social Psychology,* 1977, *35,* 1–11.

Hays, W. L. *Statistics for psychologists.* New York: Holt, Rinehart & Winston, 1963.

Kanouse, D. E., & Wiest, W. M. Some factors affecting choice in the Prisoner's Dilemma. *Journal of Conflict Resolution,* 1967, *11,* 206–213.

Kelley, H. H., & Stahelski, A. J. Social interaction basis of cooperators' and competitors' beliefs about others. *Journal of Personality and Social Psychology,* 1970, *16,* 66–91.

Langer, E. J. The illusion of control. *Journal of Personality and Social Psychology,* 1975, *32,* 311–328.

Luce, R. D., & Raiffa, H. *Games and decisions: Introduction and critical survey.* New York: Wiley, 1964.

Miller, D. T., & Holmes, J. G. The role of situational restrictiveness on self-fulfilling prophecies: A theoretical and empirical extension of Kelley and Stahelski's triangle hypothesis. *Journal of Personality and Social Psychology,* 1975, *31,* 661–673.

Ross, L., Green, D., & House, P. The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology,* 1977, *13,* 279–301.

Terhune, K. W. Motives, situation, and interpersonal conflict within Prisoner's Dilemma. *Journal of Personality and Social Psychology Monograph,* 1968, *8*(3, Pt. 2).

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1971.

Received May 26, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 4, 608-620

Information, Choice, and Reactions to Stress: A Field Experiment in a Blood Bank With Laboratory Analogue

Richard T. Mills
University of Southern California

David S. Krantz
Uniformed Services University of the Health Sciences

Two experiments examined the effects of various operations of personal control on reactions to stress. The first study incorporated two features into the blood-drawing procedure at a blood bank: providing donors with accurate information and allowing donors to choose the arm to be used. Measurement of nurses' actions to prevent donors from fainting and self-reports of discomfort revealed that the combination of choice and information was somewhat effective in reducing distress. However, providing either information or choice alone was more effective. In a second laboratory study using a cold pressor stimulus as stressor, subjects given a choice (the option to terminate the aversive stimulus and choice of hand used) showed a reduction of aftereffects on a measure of attention to detail. Subjects given information but not choice also showed this reduction. Combining information and choice was no different from either treatment alone. Taken together, the results of both studies indicate that moderate levels of choice and information are optimal for coping with stress. An explanation was suggested based on a contextually determined relationship among choice, information, and perceived control.

Laboratory and field studies have shown that negative effects of threatening stimuli are frequently lessened when subjects believe that they can escape or avoid them (e.g., Glass & Singer, 1972; Seligman, 1975) or when subjects are provided with information about the threatening event (e.g., Johnson, 1973; Staub & Kellet, 1972). In addition, providing subjects with options or choices may also affect behavioral and physiological stress reactions (e.g., Corah & Boffa, 1970; Langer & Rodin, 1976; Stotland & Blumenthal, 1964). These operations have been conceptualized as forms of personal control because they presumably allow subjects to alter or affect outcomes. Recent investigations have explored the applications of personal control to naturalistic health care situations that are inherently stressful. In this regard, information, participation, and choice have been implicated as important variables that may influence outcomes in both health treatment settings and the institutional environments (Krantz & Schulz, in press; Langer & Rodin, 1976; Leventhal & Everhart, 1978).

Situational Context and Control Interventions

A recent review, however, concluded that various operations of personal control are related to stress in a complex fashion (Averill, 1973). The particular situational context and

Experiment 1 was conducted as part of the first author's second year project at the University of Southern California under the direction of the second author. During this time, Richard Mills was also employed as research associate in nursing research at City of Hope National Medical Center. This research was supported by National Institutes of Health Grants NU00508 to Geraldine Padilla and HL21390 to David Krantz.

We wish to thank Carroll Spurling and the Pomona Blood Chapter of the American Red Cross for their cooperation and Joyce Watson and Carol Miles for assistance with data collection. Helpful comments on earlier drafts of this article were made by Gary Evans, Sheldon Cohen, Rick Gibbons, Geraldine Padilla, and Marcia Grant.

Requests for reprints should be sent to David S. Krantz, Department of Medical Psychology, Uniformed Services University of the Health Sciences, 4301 Jones Bridge Road, Bethesda, Maryland 20014.

In the public domain

the meaning of the control response to the individual determines whether it will be effective in reducing stress. This conclusion is supported by studies demonstrating an interaction among some operations of personal control. For example, Corah and Boffa (1970) manipulated the ability to escape (terminate) or not escape a loud noise. Half of the subjects were also given instructions that heightened their sense of choice about what to do, and the remainder were not given a choice. Results indicated that ability to escape was stress reducing. Choice reduced stress for the no-escape subjects, but was *not* effective (and even increased rated discomfort) for subjects given an escape response. Other research indicates that under certain conditions, making responses available to subjects or providing information may actually be stress inducing (Averill & Rosenn, 1972; Epstein, 1973). One conclusion that can be drawn from this research is that the situational context affects the outcome of interventions. Therefore, to determine the specific effects of real-life operations of control, it is necessary to conduct studies in naturally occurring settings.

The first study in this article examines the effects of operations of personal control on donor stress reactions to giving blood (phlebotomy). Blood donors were selected as the target population because giving blood is a real-life stress situation that lends itself easily to experimental interventions. We manipulated information and choice because both these variables are of theoretical interest in medical settings.

Effects of Information and Choice

Information has been conceptualized as a form of cognitive control because it often results in the interpretation of an aversive event so that the threat is lessened. Research suggests that accurate information about sensations is effective in reducing distress (cf. Johnson, 1973), presumably because it increases ability to prepare for and validate experience with aversive events. Sensory information has been shown to reduce distress associated with noxious medical examinations and unpleasant procedures (Johnson & Leventhal, 1974; Fuller, Endress, & Johnson, Note 1) as well as to facilitate recovery from surgery (Johnson, Rice, Fuller, & Endress, Note 2).

Participation and choice often lead to an increase in perceived control. Choice is seen as a form of control because it may provide subjects with the perception (correct or not) that they can have an effect on outcomes.

Both choice and participation have also been described as effective means of facilitating favorable health outcomes and reducing distress associated with various medical procedures (Cromwell, Butterfield, Brayfield, & Curry, 1977; Krantz & Schulz, in press; Taylor & Levin, 1976). Langer and Rodin (1976), in a study of institutionalized elderly people, found that providing elderly residents with the freedom to make choices about daily matters and enhancing their sense of responsibility for rather routine events resulted in heightened activity, happiness, and well-being. Moreover, a long-term follow-up found that the choice and responsibility interventions had sustained positive effects on measures of health and mortality (Rodin & Langer, 1977).

Rationale and Hypotheses

Although recent discussions of behavioral approaches to health care encourage the active involvement of clients in their own treatment (e.g., Cromwell et al., 1977; Taylor & Levin, 1976), there has been relatively little systematic investigation of the effects of various types of patient control. In a blood-bank setting, allowing donors to choose the arm to be used was a means of increasing donor involvement that could easily be incorporated into the ongoing procedure. Providing information about sensations and procedures is another means of enhancing control that could be easily incorporated into the procedure. Since little is known about how information and choice might operate together, research combining these operations can provide insight into possible positive or negative effects of providing for patient control over some aspects of treatment.

Experiment 1

The present study is designed to evaluate the effectiveness of information and choice in a real-life setting. Accordingly, we manipulated both information about the blood-drawing procedures and participation in the process by allowing donors to choose the arm to be used. Assuming that enhancing personal control would reduce distress in blood donors, it was expected that information and choice manipulations would each reduce distress involved in the phlebotomy procedure. We also expected that the combination of information and choice would reduce distress and perhaps be more effective than either treatment alone.

Method

Overview

Forty blood donors were assigned to one of four conditions in a 2×2 design. The two manipulated factors included information (high/low) and choice (high/low). High-information subjects heard taped communications describing the upcoming procedures and discussing physical and psychological sensations donors frequently experience. Low-information subjects heard a Red Cross documentary tape of comparable length revealing little about procedures. High-choice subjects were allowed to select the arm to be used for phlebotomy, whereas low-choice subjects were merely instructed that the nondominant arm would be used. After the manipulations, a nurse drew blood according to the standard American Red Cross procedure. Donor stress reactions were indexed by observing the necessary health maintenance actions taken by a trained nurse and also through self-reports of discomfort, pain, and anxiety experienced at several points in the procedure. The Stroop color-word test, a response-competition performance measure, was administered during the donor recovery period.

Subjects and Recruitment

Subjects were 40 first-time donors who passed American Red Cross screening at a local blood bank in Southern California. Twelve additional donors were approached, but they declined to participate in the study. There were 19 males and 21 females, ranging in age from 17 to 50 years, with a mean age of 26.7. Subjects were randomly assigned to conditions with the qualification that groups contain an approximately equal number of males and females.

Procedure

Subjects were tested individually. After initial screening, each subject was escorted to the testing area by a nurse. The potential subject was informed by the first experimenter of the general nature of the study, which was "to find ways to make donors more comfortable while giving blood." After agreeing to participate and signing a consent form, subjects completed brief self-report scales including premanipulation ratings of pain, discomfort, and anxiety on "visual analogue" scales (Ohnhaus & Adler, 1975). These scales contained printed instructions and were anchored with the words *none* at one end and *severe* at the other. Subjects were then assigned randomly to one of four conditions that were factorial combinations of the two experimental treatments. The manipulation for each subject was introduced by an appropriate tape-recorded message.

Information manipulation. Half of the subjects (high information) listened to a 2-minute high-information tape that described the step-by-step procedures involved in donating blood and that also discussed some of the physiological and psychological sensations commonly experienced by donors. Excerpts of the message used are as follows:

> There are three steps involved in donating blood: the preparation of the arm, the donation itself, and a short rest period afterwards. . . . We will clean your arm with soap and an antiseptic, then one of the nurses will place a blood pressure cuff on your arm, inflate it, select the largest vein from the area that has been cleaned, and insert a needle. Some people feel a slight sting initially, due to the medication that's in the needle. Once the needle is in place, blood will flow into the collection bag. After about 10 minutes the nurse will remove the cuff and the needle. If you should feel a slight numbness or tingling in your arm while the blood is being drawn, it will go away once the cuff is removed. Then you will relax. Should you feel a little faint, there's no reason to be concerned—you need a few moments to adjust.

The remaining subjects (low information) heard a taped message of comparable length that discussed the Red Crosss blood donor program but revealed little about the upcoming procedures. That message, containing statistical and very general documentary information, was excerpted from a recent report on donor recruitment (Sherer, 1976).

Choice manipulation. The tape continued with an additional 2-minute message introducing the appropriate choice manipulation. Half of the subjects (high choice) heard a message designed to heighten their feelings of active participation in the procedure by telling them that they would be allowed to choose the arm from which blood would be drawn. The message was designed to lead the donors to use a standard arm (nondominant arm), although they were actually given the opportunity to choose

whichever arm they preferred (18 of 20 subjects chose the nondominant arm). Excerpts of the message used are as follows:

> It is now time for you to choose the arm that we will use. . . . Other donating centers have their donors use the nondominant arm. That's so that if there's any soreness due to the needle afterwards, it will be on the nondominant arm and will not interfere with their normal routine. But remember, this study is interested in your role in the blood-drawing process. Therefore, regardless of what other centers have decided, here, the choice about your arm is completely up to you. Which arm do you choose? [Subjects then chose the arm that they preferred to use for the phlebotomy.]

The remaining half of the subjects (low choice) were merely instructed that the nondominant arm would be used in the blood-drawing procedure. The justification for use of this arm was to "prevent soreness from interfering with normal routine." Subjects were given no choice in selection of the arm.

Dependent Measures

Following the taped messages, subjects were escorted to the blood-drawing area where an experienced nurse from the blood center (who was blind to the subject's experimental condition) drew blood according to the standard Red Cross procedure.

Measurement of nursing interventions. During the blood-drawing procedure, the nurse performing the phlebotomy closely observed the subjects for signs of faintness and distress. The nurse was instructed to intervene as she judged necessary and to respond to her assessment of the intensity of donor reactions in a standard, graded fashion. The type and severity of health maintenance action between nurse and donor could therefore be used as a measure of donor distress. The severity of nursing interventions was coded by a second observer who was also blind to the subject's experimental condition. Earlier pilot work as well as data collected during the study revealed that nurse–donor interactions could be standardized in a way that reflected intensity of action required by donors. Nurse interventions were coded on a 4-point scale as follows: 1 = brief verbal communication to insure that the donor was not having an adverse reaction; 2 = extended verbal communication to reassure the donor (and check on his/her status); 3 = physical contact (touch); and 4 = the use of cold towels and/or ammonia capsules to revive the patient. Use of this procedure made it possible to systematically measure nursing interventions and at the same time to provide necessary care to donors. In addition, both nurse and observer independently responded yes or no to whether they had observed adverse physical reactions in donors during the phlebotomy. Physical reactions were defined as overt signs of perspiration, tension, dizziness, or fainting.

Postsession self-ratings and performance measures. After the blood drawing was concluded, the second observer administered scales containing instructions that asked subjects to retrospectively rate the degree of pain, anxiety, and discomfort they experienced during the time blood was being drawn. Subjects also rated their postsession reactions, indicating how they felt after the procedures were over. Next, according to usual Red Cross procedures to facilitate recovery, subjects were given juice and a snack. After these refreshments were offered, subjects were given the Stroop color-word test. This timed response-competition task, given in three parts, has been used as a measure of behavioral aftereffects of stress (Glass & Singer, 1972). Manipulation checks and debriefing concluded the procedure.

Results [1]

Checks on Experimental Manipulations

The manipulations were effective in inducing both differential perceptions of choice and differential perceptions of information received from the taped communications. An item on the postexperimental questionnaire asked subjects to indicate where they received the most information about the blood-drawing procedure. Seventy-nine percent of subjects in the high-information conditions rated the taped message as their primary source of information, whereas 37% of those in the low-information conditions did the same, $\chi^2(1) = 5.29$, $p < .03$. Low-information subjects indicated that other sources had provided them with the most information (e.g., companions, nursing staff).

Subjects responded to the question, "To what extent did you have a choice in selecting which arm would be used for blood drawing?" on a 5-point scale. A 2 (choice) × 2 (information) analysis of variance of this item revealed the expected main effect for choice, $F(1, 34) = 110.90$, $p < .001$. There was also a main effect for information, $F(1, 34) = 4.16$, $p < .05$, with high-information groups rating more perceived choice. There were no other significant effects. Examination of the means for each condition indicated that the high-information–high-choice and low-information–high-choice groups

[1] Data were recoverable for only 38 subjects on manipulation checks and for 39 subjects on self-rating measures of distress.

Table 1
Number of Subjects Requiring Each Type of Nurse Intervention and Mean Intervention Score

Condition	Severity of intervention				
	None	Verbal status check	Touch	Cold towels	*M* intervention score[a]
Low information–low choice	1	2	2	5	1.9
Low information–high choice	7	2	0	1	1.3
High information–low choice	8	1	0	1	1.2
High information–high choice	4	2	1	3	1.6

Note. There were 10 subjects in each condition.
[a] This score was computed by dichotomizing nurse actions as none versus intervention (verbal, touch, or towels). Higher scores on this scale reflect greater severity.

rated the most choice ($M = 4.9$ and 4.7, respectively) with ratings for high-information–low-choice ($M = 2.0$) and low-information–low-choice groups ($M = 1.0$) significantly lower ($p < .05$ by Newman–Keuls test). High-information–low-choice subjects also apparently perceived somewhat more choice than their low-information–low-choice counterparts ($p < .05$).

Nurse Interventions During Blood Drawing

The principal measure of donor distress during the phlebotomy procedure consisted of observer ratings of nurse interventions coded on a 4-point scale of intensity (no adverse reaction to the application of cold towels). In addition, both the nurse and second observer recorded whether an adverse reaction (overt weakness, dizziness, or faintness) was present or absent. Evidence for the reliability of nurse intervention intensity as a measure of distress was provided by the fact that everyone (10 subjects) who received cold towels was rated by the nurse as having had a reaction. Across the range of interventions there was 80% agreement between independent reports of reactions made by the nurse and the second observer.

It was hypothesized that donors who were singly given information or choice would show reduced stress responses and that the combination of these factors would be even more effective in reducing stress. The data for the 4-point measure of nurse interventions are presented in Table 1. These data were not normally distributed and, strictly speaking, were not appropriate for analysis of variance. To overcome this distributional problem, the dependent measure was first dichotomized as no intervention versus intervention (verbal communication, touch, or towels) and then subjected to a 2 (choice) × 2 (information) analysis of variance.[2] This analysis revealed no significant main effects ($p > .15$). Instead, there was a significant interaction between choice and information treatments, $F(1, 36) = 12.86$, $p < .002$. This effect indicated that although the high-information–high-choice ($M = 1.6$) condition reduced the severity of interventions relative to the low-information–low-choice condition ($M = 1.9$, $p < .05$, by Newman-Keuls test), both conditions required significantly greater nurse interventions ($p < .05$) than either the high-information–low-choice ($M = 1.2$) or low-information–high-choice groups ($M = 1.3$). In short, these results indicated that although combining information and choice was effective in reducing stress reactions, information or choice given alone was more effective. This conclusion is further supported by a multivariate analysis of variance conducted on the cluster of measures of distress during blood drawing (namely, nurse intervention intensity and rated discomfort, pain, and anxiety dur-

[2] Analysis of variance on the original 4-point nurse intervention data in Table 1 revealed the identical pattern of results as did post hoc contrasts.

Table 2
Mean Self-Ratings of Three Sections of Experiment

Experimental condition	Self-rating measure of distress	Section of experiment: Pre-manipulation	Blood drawing[a]	Postsession[a]
Low information–low choice[b]	Pain	1.2	4.3	1.4
	Discomfort	2.5	3.8	2.1
	Anxiety	4.6	4.2	1.6
Low information–high choice[b]	Pain	1.3	3.4	1.2
	Discomfort	1.9	1.5	1.1
	Anxiety	3.8	2.5	1.1
High information–low choice[c]	Pain	1.2	1.8	1.6
	Discomfort	1.9	1.1	1.7
	Anxiety	3.3	3.8	2.6
High information–high choice[b]	Pain	1.7	3.5	2.5
	Discomfort	2.8	4.0	2.3
	Anxiety	5.1	4.2	2.1

Note. These ratings were taken on visual scales ranging from 1 = none to 14 = severe.
[a] Means for this measure are adjusted for premanipulation score.
[b] $n = 10$.
[c] $n = 9$.

ing the procedure). The analysis revealed a significant interaction between information and choice treatments, $F(4, 32) = 4.55$, $p < .01$, and no main effects. Relevant univariate effects for these variables are presented elsewhere in the results section.

Self-Ratings

Subjects were asked to rate degree of perceived discomfort, pain, and anxiety on 14-point scales for three separate time periods, with high scores indicating greater distress. The first rating was conducted prior to delivery of the experimental manipulations. The next two ratings were taken after blood drawing. It should be noted at the outset that subjects tended to rate the measures of distress toward the low end of the scale. On premeasures of discomfort, pain, and anxiety, there were no initial group differences prior to receiving the manipulations. This was confirmed by separate 2 (high–low choice) × 2 (high–low information) analyses of variance on each premeasure of the three rated measures of distress that revealed no significant effects (all $ps > .20$).

In the postsession following blood drawing, two ratings were made. Subjects retrospectively rated discomfort, pain, and anxiety felt *during* phlebotomy and then rated their present (postphlebotomy) perceptions on each measure of distress scale. To determine how the manipulations affected self-ratings, repeated measures analyses of covariance[3] were conducted. For each choice of either discomfort, pain, or anxiety, respective premanipulation ratings served as the covariate and the two postmanipulation measures as the within-subject term. Adjusted means for these ratings are presented in Table 2.

Covariance analysis for discomfort revealed a significant trials effect, $F(1, 35) = 12.69$, $p < .002$, indicating greater discomfort during blood drawing than for the postmeasure. Simple effects were not significant; however, there was a reliable interaction between choice and information treatments, $F(1, 34) = 4.35$, $p < .05$. The interaction parallels the finding for nurse interventions and indicates that

[3] Although there were no significant between-conditions effects on premeasures, covariance analyses were used for postmanipulation ratings to correct for small initial differences. A repeated-measures analysis of variance on these measures of distress revealed the same but somewhat stronger pattern of results.

both the low-information–low-choice and high-information–high-choice groups rated more discomfort than low-information–high-choice and high-information–low-choice subjects. Moreover, a significant interaction between choice, information, and trials, $F(1, 35) = 4.60$, $p < .04$, reinforced the fact that between-conditions differences were greatest for rated discomfort during blood drawing. This conclusion was confirmed by separate 2×2 analyses conducted on ratings of each trial. For rated discomfort during blood drawing, the Choice × Information interaction was significant ($p < .03$). For discomfort in postsession, this effect was weaker ($p < .09$).

Repeated measures covariance analysis of pain ratings once again revealed a trial effect, $F(1, 35) = 11.81$, $p < .003$. The only other significant effect was an interaction between information treatments and trials, $F(1, 35) = 4.44$, $p < .05$. The nature of this interaction was revealed by separate 2×2 analyses for each trial. For ratings of pain during blood drawing, there were no significant effects ($ps > .17$), whereas for self-reported pain in postsession, a main effect for information treatments emerged, $F(1, 35) = 4.45$, $p < .05$. This effect indicated higher postsession pain ratings for the high-information groups. Newman-Keuls tests revealed that this effect was largely due to the high-information–high-choice group rating more pain than all the other groups in postsession ($p < .05$) and the low-information–high-choice group rating less pain than all the others ($p < .05$).

Finally, repeated measures covariance analysis of anxiety ratings revealed only a significant trials effect, $F(1, 36) = 7.67$, $p < .01$, reflecting a decrease in rated anxiety across trials.

Pearson correlation coefficients were computed between the 4-point nurse intervention measure and self-ratings of experience during blood drawing. Obtained coefficients for the 39 cases were .31 for pain, .30 for discomfort, and .46 for anxiety. These correlations of moderate magnitude were all significant ($ps < .03$). In sum, results for the self-rating data are generally supportive of the results obtained on the behavioral measure of nurse interventions. Single manipulations of information and choice were effective in reducing distress indexed by rated discomfort and, to a weaker extent, by a postmeasure of pain. There was some indication that ratings were highest for the group receiving the combination of information and choice. For rated anxiety, the treatments did not appear to have any differential effect.

Stroop Color-Word Interference Test

Analysis of this measure revealed no significant main effects or interactions.

Discussion

Results of this experiment suggest that information and choice manipulations, administered alone, are effective means of reducing distress during a phlebotomy procedure in a blood-bank setting. However, when these treatments are combined and given together, their effects are not additive. On a measure of nursing interventions, the data indicate that high levels of both information and choice are somewhat less effective than single manipulations of either variable. The pattern of results obtained on self-rating data are more complex, but still support the general conclusion that moderate degrees of personal control appear to be optimal for coping with stress in this setting.

Although relatively few studies have examined the effects of choice in conjunction with other operations of personal control, results of the present study resemble those obtained by Corah and Boffa (1970), who examined the effects of combining choice with the ability to escape aversive stimulation. They found no evidence that these combined treatments were more effective than single manipulations of either variable. In fact, Corah and Boffa found that subjects given both choice and ability to escape aversive noise actually rated the noise as producing as much discomfort as the group receiving neither treatment. This is the same effect for rated discomfort obtained in the present study.

Although much of stress research is characterized by a dissociation among behavioral,

self-report, and physiological measures (cf. Glass & Singer, 1972; Lazarus, 1966), a problematic feature of the present study is a degree of inconsistency among the various self-rating indexes of distress. Although rated discomfort and the postsession measure of pain generally support the conclusion that moderate levels of personal control reduce distress, neither self-rated anxiety nor rated pain during blood drawing were affected by the manipulations in this study. It should be noted that ratings tended to be rather low—indicating little stress. This may have been so because donors chose to give blood and to be subjects. Moreover, rating data were gathered during the postsession after nursing measures were taken to assist donors in coping with the physical distress of giving blood.

Reflecting on the original hypotheses of the study, we sought to determine if two treatments that could be easily incorporated into the ongoing blood-donating procedure would lessen donor distress. The surprising result on the nurse intervention measure indicated that in the high-information–high-choice group, there were actually more severe reactions than in the groups receiving only one manipulation. Fortunately, this condition still resulted in less severe reactions than those in the group given neither information nor choice. In a purely descriptive sense, this finding reinforces Averill's (1973) conclusion that the context and meaning of a "control response" determines whether it will be effective in reducing stress. However, we might speculate briefly about the cognitive mechanisms that could have produced these results. One possibility is that the combination of information and choice manipulations gave donors more of a role in the blood-drawing procedure than they would have preferred. That is, after being informed of the procedures and sensations, donors might rather have had the nurse be responsible for the rest of the procedure. This explanation raises the intriguing possibility that participation or choice might actually result in a decrease in "perceived control" (ability to affect outcomes) if the individual prefers not to have it or believes that it will not be effective (see Houston, 1972). Under those circumstances, increased choice or participation may even heighten stress. Although others have speculated that this possibility may apply to health care outcomes (Freidson, 1975; Rodin & Langer, 1977), this issue has not been extensively researched and deserves further investigation.

Finally, there were no reliable effects on Stroop test performance at the end of the study. An examination of the blood-bank procedure, however, revealed several reasons why group differences in performance on an aftereffects task may not have been observed in that setting. In addition to receiving cold towels and ammonia capsules, subjects with adverse reactions were reassured and reoriented; their stress was reduced. We suspected that the nursing interventions during phlebotomy and refreshments given to donors following the procedure precluded an adequate assessment of aftereffects in that setting. We therefore conducted a partial laboratory replication of the blood donor study to determine if information and choice can prevent aftereffects of exposure to aversive stimulation.

Experiment 2

Research (Glass & Singer, 1972) suggests that a variety of stressors—at least when they are unpredictable and uncontrollable—produce aftereffects that appear relatively soon after stimulation is terminated. Aftereffects have been observed as impairments in performance of a variety of tasks including measures of frustration tolerance and attention to detail. Moreover, Glass and Singer (1972) have shown that if subjects believe they have control over the onset or offset of aversive stimuli, these negative aftereffects of stress arousal are reduced. In several studies, this "perceived control" variable was manipulated by providing subjects with the option of terminating an aversive stimulus.[4] The ex-

[4] Glass and Singer (1972) have conceptualized operations similar to this manipulation as "perceived control," whereas others (e.g., Corah & Boffa, 1970) have termed it *choice*. Although we feel it is reasonable that choice and perceived control do not always vary together, we have evidence that the choice manipulation in Experiment 2 did heighten feelings of control (see Results section).

plicit relationship between providing information about a stressor and behavioral aftereffects has not been investigated previously.

To create a rough analogue to the blood-drawing procedure, a cold pressor stimulus was used in which subjects immersed their hand in ice water. Half of the subjects (high information) received information about the sensations of a cold pressor stimulus (coldness, aching, etc.), whereas the remaining half (low information) received no information. Crosscutting this treatment, half of the subjects (high choice) were allowed to choose the hand to be used and could elect to remove their hand from the cold water if they chose to do so. The remaining subjects (low choice) were not given these options. Dependent measures included performance on an aftereffects task and self-ratings. Based on the results of the previous study, it was predicted that single manipulations of information and choice would be effective in preventing aftereffects on a poststress task. The combined manipulations were not expected to be any more effective than either treatment alone and perhaps somewhat less effective as observed on some measures in the first study.

Method

Subjects

Subjects were 44 undergraduate psychology students at the University of Southern California who received class credit for participation. There were 13 males and 31 females in the sample who were randomly assigned to one of the four experimental conditions.

Procedure

Subjects were tested individually. After being greeted by the experimenter, each subject was seated at a table and was instructed that the purpose of the experiment was to "measure blood pressure under various conditions." During the study subjects were told that they would be required to place their hand in ice water and were given the option to withdraw from the study. (All subjects elected to continue.) Each subject then heard a taped message describing 7-point self-report adjective scales to be used during the study. The tape was stopped to make sure instructions were being understood, and subjects filled out a premeasure of Likert scales measuring their present state of discomfort, pain, and anxiety. The scales were anchored by 1 = none at all, 4 = moderate amount, and 7 = most ever experienced. After the subjects completed these premeasures, the tape was turned on again, and prerecorded messages presented the appropriate experimental manipulations.

Information manipulation. Half of the subjects (high information) listened to the high-information manipulation that described the physiological and psychological sensations commonly experienced by persons who placed their hands in ice water. The manipulation was as follows:

> We would like to let you know what you can anticipate from holding your hand in the ice water. Most people report a coldness, aching, tightness of the skin across the hand. Also, they experience a "pins and needles" sensation and numbness. Such reactions are normal, and after you remove your hand from the water, the reactions will go away.

The remaining half of the subjects (low information) received no information at this point and heard only the subsequent messages.

Choice manipulation. Half of the subjects (high choice) heard instructions that emphasized their ability to determine which hand would be inserted into the water. (Instructions encouraged choice of the nondominant hand.) In addition, although discouraged from doing so, they were told they would be able to withdraw their hand if they chose. Excerpts of the message are as follows:

> The next step is holding your hand in the ice water . . . You can place whichever hand you want into the water . . . the choice of hands is entirely up to you. It is important that you do not exchange hands after you have placed one of them in the water. Secondly, while we would like the subjects to keep their hand in the water for a predetermined length of time, you are free to withdraw your hand at any time you may wish. Now pick the hand you will place in the water. We find that most people pick the hand that they do *not* use for writing so that you can fill out the scales with the free hand. When I tell you, place your hand in the metal bowl, resting your palm flat on the bottom surface. Although a few people who come here don't keep their hand submerged the full time, most do. For the success of the experiment, I prefer you to keep your hand submerged until I tell you remove it, but that's entirely up to you.

The remaining half of the subjects (low choice) were told which hand they would insert into the water and were told not to remove their hand. The message was as follows:

> It is important that *I* control the length of time you keep your hand in the water. For this study to be successful, all the subjects must be exposed to the ice water for the same length of time. So

please keep your hand in the water until I tell you to remove it. Also, we have chosen the hand you will place in water. Place the hand that you do *not* use for writing in the water so that it will be easier for you to fill out the scales. When I tell you, I will want you to place your hand in the metal bowl, resting your palm flat on the bottom surface. Leave your hand there until I tell you to remove it.

At this point all subjects submerged their hand in a flat-bottomed metal pan containing ice water at 5 °C. The tape recorder remained on and after 2 minutes delivered a message instructing subjects to again complete the four self-rating scales. This measure reflected self-reports during exposure to the cold pressor stimulus. After a total of 4 minutes had elapsed from the time of submersion, the tape instructed subjects to remove their hand from the water. At this time the experimenter measured blood pressure with a cuff and administered postmeasures.

Postsession ratings. After subjects removed their hand from the water, they were asked to indicate once more how they had felt when their hand was submerged in the ice water and, finally, to indicate how they felt *at that moment*. As with the premeasure, these ratings were taken on scales measuring discomfort, pain, and anxiety.

Aftereffects task. The final task for the subject was to proofread a simplified version of a 10-page passage used extensively in previous research (e.g., Glass & Singer, 1972; Krantz & Stone, 1978) to measure attention to detail. This passage contained typographical errors, misspellings, and grammatical and punctuation mistakes. Subjects were instructed to proofread the passage, detect the errors, and underline them. Instructions emphasized both accuracy and speed. After instructions were given, the subject was left to work alone for 6 minutes, although nothing was said about time. In accord with previous research using this measure (e.g., Glass & Singer, 1972; Krantz & Stone, 1978) a percentage correct score was computed as a measure of performance. This score consisted of the percentage of correctly located errors in the portion of the passage completed by subjects when asked to stop working. This was a measure of performance accuracy correcting for amount read. "Total amount read" was also analyzed. After the proofreading task, manipulation checks were administered and subjects were debriefed.

Results [5]

Manipulation Checks

Data from postexperimental questions indicated that the manipulations were effective. Subjects were first asked to rate (on a 7-point scale) the degree of choice they had in selecting the hand to be placed in the cold water.

Table 3
Mean Percentage of Correctly Located Errors on Proofreading Task

Information	High choice	Low choice
High	71.2	67.8
Low	74.2	52.3

Note. $n = 11$ subjects per condition. Higher scores reflect better performance.

A 2 (high–low information) × 2 (high–low choice) analysis of variance revealed that high-choice subjects indeed rated more choice, $F(1, 40) = 89.30$, $p < .001$. There were no other significant effects. A similar analysis was conducted on an item asking subjects to rate the degree of choice they felt in determining when their hand would be removed from the water. Once again, only the appropriate main effect for choice emerged, $F(1, 40) = 71.40$, $p < .001$.

To check on the information manipulation, subjects were asked how much information they had received about sensations they would experience from the cold water. Analysis of these 7-point scale ratings revealed only an information main effect, $F(1, 40) = 62.57$, $p < .001$.

Dependent Measures

Aftereffects task. It was predicted that single manipulations of information and choice would prevent behavioral aftereffects of exposure to the experimental stressor. High levels of both information and choice were not expected to produce any additional remedial effect. A two-way analysis of variance was conducted on percentage of errors correctly identified on the proofreading task. Means for this measure are presented in Table 3. There was a significant main effect for choice, $F(1, 40) = 12.52$, $p < .001$, with

[5] Four subjects in the high-choice conditions removed their hand from the ice water before the full 4-minute submersion period. Data both excluding and including these individuals on all dependent measures were analyzed. Results on all measures were virtually identical both ways. Therefore, data for the full complement of 44 subjects are presented.

high-choice subjects ($M = 72.7$) being more accurate than low-choice subjects ($M = 60.1$). In addition, there was an unreliable tendency toward an information main effect ($p < .09$). A significant Choice × Information interaction also was evident, $F(1, 40) = 6.66$, $p < .02$. This interaction indicated that under low-choice conditions, providing information led to improved performance ($p < .05$ by Newman-Keuls test). Under high-choice conditions, providing information made little difference and even depressed performance slightly, though not reliably. Newman-Keuls tests revealed that the low-choice–low-information group was significantly less accurate than the other three groups ($p < .05$). The latter three groups, however, did not differ reliably from one another.

An analysis was also conducted on amount of the proofreading passage read (i.e., number of lines of text) during the 6-minute test period. This 2 × 2 variance analysis revealed only a significant Choice × Information interaction, $F(1, 40) = 4.00$, $p < .052$. This interaction indicated that low-information–high-choice subjects read the most ($M = 52.2$ lines), followed in decreasing order by high-information–low-choice ($M = 48.7$ lines), high-information–high-choice ($M = 43.5$ lines), and low-information–low-choice ($M = 39.8$ lines). These individual group mean differences did not reach reliable levels of significance ($ps > .05$ by Newman-Keuls); however, these data demonstrate that on a measure of performance speed as well as accuracy, the effects of information and choice are not purely additive in preventing aftereffects.

Self-Report Measures

Analyses of variance (2 × 2) were conducted on each premeasure of discomfort, pain, and anxiety. These analyses revealed no initial premanipulation rating differences between groups on any of the measures of distress. Next, to determine if postmanipulation self-reports were affected by the experimental treatments, repeated measures analyses of covariance were conducted on the three subsequent measures of distress. These analyses revealed no significant treatment effects or interactions for any of the measures. On 7-point scales, mean self-ratings during submersion ranged between 4.5 and 5.1 for discomfort, 4.0 and 4.4 for pain, and 3.1 and 3.9 for anxiety. For each measure of distress, however, there was a significant trials effect ($ps < .01$) reflecting heightened aversive experience during the cold pressor test and subsequent relief in postsession. Examination of the means revealed that relative to premeasures, experience during cold pressor was rated with heightened discomfort, pain, and anxiety by subjects in all conditions. In addition, Pearson correlation coefficients were computed between rated experience during the cold pressor test and the aftereffects measure of percentage correct on proofreading. These correlations were uniformly low and nonsignificant, indicating a lack of relationship between these measures.

Other Relevant Data

Analysis of blood pressure measured following the cold pressor test revealed no reliable effects. However, it is worth noting that there was considerable variance in this measure, and the low-information–low-choice and high-information–high-choice groups tended to show somewhat higher systolic and diastolic blood pressures. Additionally, in a postexperimental questionnaire, subjects rated the degree of control they felt over removing their hand from the water An analysis of variance of this item revealed only a choice main effect ($p < .01$), with high-choice subjects rating more control.

Discussion

Results on the behavioral aftereffects of measure of Experiment 2 are generally consistent with our hypotheses and with the results of the first experiment. The information and choice manipulations were effective in preventing performance deficits when each was given alone; combining the treatments added little to subsequent performance. The data on the proofreading measure also reveal that the choice manipulation was more potent than providing subjects with relevant in-

formation in preventing behavioral aftereffects.

The manipulations in Experiment 2 did not affect any of the self-reports of discomfort, pain, or anxiety; and correlations between aftereffects and self-ratings were nonsignificant, as in the Glass and Singer (1972) research. Although related studies (e.g., Glass & Singer, 1972, pp. 47–55; Sherrod, Hage, Halpern, & Moore, 1977) have reported that control manipulations did not affect rated aversiveness, other studies have found a tendency for predictable and controllable stimuli to be rated as less aversive (Corah & Boffa, 1970; Fuller et al., Note 1). We have no ready explanation for these apparent inconsistencies. Further investigation is necessary to clarify the conditions under which self-report measures are or are not affected by manipulations of personal control.

In our discussion of the results of Experiment 1, we speculated that too much patient control might not be desired or expected in a blood bank. In accord with this explanation, we found in the first study that both of the groups receiving only one manipulation experienced reduced stress reactions compared to the high-information–high-choice condition. Results of the cold pressor study (Experiment 2) indicate that groups receiving either combined or single manipulations evidenced reduced stress on an aftereffects task. We attribute this somewhat different pattern of results to characteristics of the respective settings in which the studies were conducted and also to the particular operations used to manipulate choice.

First, subjects in a laboratory study probably do not enter the situation with established expectancies about how much control they will be allowed. Therefore, in contrast to a clinical setting, there is little expectancy violation involved in being given control over a laboratory stressor. Second, the nature of the choice manipulation in the cold pressor study (i.e., the option to terminate the stimulus) made it explicitly associated with "perceived control" (Glass & Singer, 1972). Subjects had every reason to believe that this form of control would be effective in terminating the aversive stimulation, thereby promoting the stress-reducing properties of this treatment. (Note the choice main effect.) This logic argues that choice manipulations will reduce stress to the extent that they are perceived to be effective in regulating aversiveness.

The two studies presented here shed light on some of the complexities of the control–stress relationship. Based on these findings, we can conclude that information and choice can play a useful role in remediating adverse physical reactions to an aversive health-care procedure. Reasonably similar manipulations were shown to affect the magnitude of behavioral aftereffects induced by a laboratory stressor. In both studies, there was little or no gain in effectiveness from combining the treatments. In light of the economic and medical benefits of self-care and the growing application of behavioral science principles to health-care settings (e.g., Krantz & Schulz, in press; Leventhal & Everhart, 1978), it is important to investigate further the effects of personal control in its various operational forms. Moreover, subtle differences between the clinical and laboratory settings suggest that naturalistic study of this problem can be maximally beneficial.

Reference Notes

1. Fuller, S. S., Endress, M. P., & Johnson, J. E. *Control and coping with an aversive health examination*. Paper presented at the meeting of the American Psychological Association, San Francisco, August, 1977.
2. Johnson, J. E., Rice, V. H., Fuller, S. S., & Endress, M. P. *Sensory information, behavioral instruction, and recovery from surgery*. Paper presented at the meeting of the American Psychological Association, San Francisco, August 1977.

References

Averill, J. R. Personal control over aversive stimuli and its relationship to stress. *Psychological Bulletin*, 1973, *80*, 286–303.

Averill, J. R., & Rosenn, M. Vigilant and nonvigilant coping strategies and psychophysiological stress reactions during the anticipation of electric shock. *Journal of Personality and Social Psychology*, 1972, *23*, 128–141.

Corah, N. L., & Boffa, J. Perceived control, self-observation, and response to aversive stimulation.

Journal of Personality and Social Psychology, 1970, *16,* 1–4.

Cromwell, R. L., Butterfield, E. C., Brayfield, F. M., & Curry, J. J. *Acute myocardial infarction: Reaction and recovery.* St. Louis, Mo.: Mosby, 1977.

Epstein, S. Expectancy and magnitude of reaction to a noxious UCS. *Psychophysiology,* 1973, *10,* 100–107.

Freidson, E. Increasing choice within limits: Commentary on Donald Kennedy's model for change. In J. Howard & A. Strauss (Eds.), *Humanizing health care.* New York: Wiley, 1975.

Glass, D. C., & Singer, J. E. *Urban stress: Experiments on noise and social stressors.* New York: Academic Press, 1972.

Houston, B. K. Control over stress, locus of control, and response to stress. *Journal of Personality and Social Psychology,* 1972, *21,* 249–255.

Johnson, J. E. Effects of accurate expectations about sensations on the sensory and distress components of pain. *Journal of Personality and Social Psychology,* 1973, *27,* 261–275.

Johnson, J. E., & Leventhal, H. Effects of accurate expectations and behavioral instructions on reactions during a noxious medical exam. *Journal of Personality and Social Psychology,* 1974, *29,* 710–718.

Krantz, D. S., & Schulz, R. Life crisis, control, and health outcomes: A model applied to cardiac rehabilitation and relocation of the elderly. In A. Baum & J. E. Singer (Eds.), *Advances in Environmental Psychology (Vol. 2).* Hillsdale, N.J.: Erlbaum, 1979, in press.

Krantz, D. S., & Stone, V. Locus of control and the effects of success and failure in young and community-residing aged women. *Journal of Personality,* 1978, *46,* 536–551.

Langer, E., & Rodin, J. The effects of choice and enhanced personal responsibility for the aged: A field experiment in an institutional setting. *Journal of Personality and Social Psychology,* 1976, *34,* 191–198.

Lazarus, R. S. *Psychological stress and the coping process.* New York: McGraw-Hill, 1966.

Leventhal, H., & Everhart, D. Emotion, pain and physical illness. In C. E. Izard (Ed.), *Emotions and psychopathology.* New York: Plenum, 1978.

Ohnhaus, E. E., & Adler, R. Methodological problems in the measurement of pain: A comparison between the verbal rating scale and the visual analogue scale. *Pain,* 1975, *1,* 379–384.

Rodin, J., & Langer, E. Long-term effects of a control-relevant intervention with institutionalized aged. *Journal of Personality and Social Psychology,* 1977, *35,* 897–902.

Seligman, M. E. P. *Helplessness.* San Francisco: Freeman, 1975.

Sherer, P. B. Blood donor motivation: A background paper for National Heart and Lung Institute conference on behavioral science and medicine. In S. M. Weiss (Ed.), *Proceedings of the National Heart and Lung Institute Working Conference on Health Behavior* (DHEW Publication No. NIH 76-868). Bethesda, Md.: Department of Health, Education and Welfare, 1976.

Sherrod, D. R., Hage, J. N., Halpern, P. L., & Moore, B. S. Effects of personal causation and perceived control on responses to an aversive environment: The more control the better. *Journal of Experimental Social Psychology,* 1977, *13,* 14–27.

Staub, E. & Kellet, O. S. Increasing pain tolerance by information about aversive stimuli. *Journal of Personality and Social Psychology,* 1972, *21,* 198–203.

Stotland, E., & Blumenthal, A. The reduction of anxiety as a result of the expectation of making a choice. *Canadian Journal of Psychology,* 1964, *18,* 139–145.

Taylor, S. E., & Levin, S. *The psychological impact of breast cancer: Theory and research.* San Francisco: West Coast Cancer Foundation, 1976.

Received June 20, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 4, 621-634

Measuring Causal Attributions for Success and Failure

Timothy W. Elig and Irene Hanson Frieze
University of Pittsburgh

A great deal of research has been generated by the Weiner et al. attribution model of achievement behavior. Although generally supportive of the model, the literature is marred by a lack of concern with the reliability and validity of the measurements used. In this study, causal attributions for a manipulated success–failure event were collected from college students on five different measuring instruments. Results indicated that the open-ended response measure showed poorer interest correlation validity than did the structured measures. Rating scales showed a better fit to attribution conceptualizations than did the percentage method. Overall, scale measures seem to be the method of choice, although possible uses for open-ended measures of attributions are suggested.

Although a great deal of research based on Weiner et al.'s (1971) model of causal attributions for success and failure events has been published during the 1970s, relatively little attention has been given to the question of how causal attributions should best be measured (Deaux & Farris, Note 1; Smith, Note 2). Although a few articles (e.g., Elig & Frieze, 1975; Frieze, 1976; Weiner, 1974; McHugh, Note 3) have referred to the variety of measures used for assessing attributions, there has been no formal study of the implications of using one attribution measure over another, and researchers have tended to be unsystematic in their selection and use of the common techniques. This article explores the interrelationship of several measures of causal attributions to assess their validity and to make recommendations concerning the selection of instruments to be used in future studies of causal attributions for success and failure events.

As shown in Table 1, there are a number of commonly used techniques for assessing causal attributions: open-ended responses, independent ratings, ipsative ratings, choice of one major cause, and bipolar ratings. Each of these methods has its own advantages and disadvantages in terms of practical considerations.

Most attribution studies use structured ratings rather than open-ended data. An open-ended procedure involves asking subjects to state in their own words why a particular event has occurred. These verbal responses can then be classified by a skilled rater into any of a set of previously defined attributional categories. Such a coding procedure is described more fully by Elig and Frieze (1975). Other systems have been developed by Bar-Tal and Darom (in press) and by Cooper and Burger (Note 4). The necessity for training coders and the time-consuming nature of this type of causal assessment contributes to its rare use in attribution research. However, there are also problems with structured response measures, which confine subjects to a limited set of factors defined in advance by the experimenter as important for the situation. This set may not include the factors of importance for some subjects. Open-

Portions of this article were presented at the meeting of the American Psychological Association in 1978 and served as part of the requirements for a master's degree for the first author.

The authors would like to thank Grace French-Lazovik, Maureen McHugh, Diane Ruble, Eliot R. Smith, and Bernard Weiner for comments on earlier drafts of this article.

Requests for reprints should be sent to Irene Hanson Frieze, Department of Psychology, University of Pittsburgh, Pittsburgh, Pennsylvania 15260.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3704-621$00.75

Table 1
Methods of Assessing Causes of Success and Failure

Method	Method used by	Example
Unstructured		
Open-ended	Elig & Frieze (1974, 1975) Frieze (1976) McHugh (Note 5)	Why do you think you *succeeded* on this task?
Structured		
1. Independent		
Unipolar ratings	Feather & Simon (1971b, 1972) Valle & Frieze (1976)	Rate the extent to which these factors caused your *success:* 1. If the factor to no extent caused the outcome. 9. If the factor to an extremely high extent caused the outcome. Your high general intelligence _____ How easy this type of task is _____
2. Ipsative		
Percentage assessment	Meyer (1970)	To what extent was your *success* caused by your high general intelligence _____ how easy this type of task is _____
Choice of one cause	Bailey, Helm, & Gladstone (1975)	Which of the following contributed the most to the outcome? Ability Effort
Bipolar ratings	Feather (1969) Feather & Simon (1971a) Weiner, Nierenberg, & Goldstein (1976)	My outcome was mainly due to: Ability–Luck[a]
Paired comparisons	McMahan (1973)	Of each pair, circle which is more responsible for your outcome. Ability, luck Ability, effort

[a] Measured on a 5-point scale.

ended response measures avoid this problem as well as the cuing of subjects toward considering causal possibilities that they may not have spontaneously considered (Frieze, 1976; Smith, Note 2).

Weiner et al. (1971) postulated that individuals attribute the causes of success and failure primarily to four of the causal elements discussed by Heider (1958): ability, effort, task difficulty, and luck. Several studies using structured measures supported the belief that these factors are used by subjects in systematic ways to explain achievement outcomes (e.g., Frieze & Weiner, 1971; Weiner, Heckhausen, Meyer, & Cook, 1972; Weiner & Kukla, 1970). Frieze (1976) employed open-ended questionnaires to ascertain what causes college students naturally used to explain success or failure in two achievement tasks (an exam and an unspecified game). The results indicated that the four causal factors postulated by the Weiner et al. (1971) model were used by subjects for these situations and accounted for the large majority of causal attributions. However, two additional causal factors, mood and other people, were also indicated. Other recent studies (Elig & Frieze, 1975; McHugh, Note 5) that used open-ended questions for a variety of social and achievement situations found still other causal factors that are frequently used to explain success and failure. Among these are personality, interest in the task, and physical appearance. This research supports

the importance of the four causes proposed originally but suggests the importance of other factors not previously considered, especially for dealing with nonacademic situations. These results raise serious questions about studies relying only on the four causal factors of ability, effort, luck, and task difficulty in a structured format.

An additional advantage of open response assessment is that subjects may find open response questions easier and more natural to respond to. However, even though it is hypothesized in this study that subjects will prefer open-ended questions, it is also hypothesized that open-ended questions will be psychometrically inferior to more structured responses. First, the added step of coding the unstructured responses should lead to lower reliability for open-ended questions than for structured measures. Second, structured measures provide a closer approximation to interval or ratio measurement. Third, structured measures allow for degrees of attribution on various dimensions rather than the simple presence–absence or frequency of appearance measures typical for open-ended response coding.

A major distinction between various structured attribution measures is whether they involve ipsative or independent judgments. Ipsative measures are measures in which the score of one attribution must influence the score of other attributions, thus inducing negative correlations. Negative correlations are not forced by measures using independent judgments. With independence of ratings comes ease of analysis, since each attribution can be tested separately. However, independent ratings do not give as direct an assessment of the relative importance of attribution factors as is given by ipsative judgments. This ease of comparison is the major attraction of ipsative measurement.

Among the ipsative measures, the assignment of percentages to various causes is perhaps the best developed. Percentage ratings (e.g., Meyer, 1970) make explicit the basic assumption of interdependent judgments that the causes being rated account for the totality of cause for the outcome and that the total cause of an event can be parceled out to various particular causes. McHugh (Note 3) pointed out that percentage ratings give the clearest indication of the relative importance of various causes (e.g., luck rated relative to other causes seems unimportant).

Another widely used method of causal assessment, the single bipolar scale anchored by ability and luck used by Feather (1969) and Feather and Simon (1971a), does not allow for even the four causal factors identified by Weiner et al. (1971). It measures only two of the possible causes of events. This measure also confounds two dimensions that have been shown to be important for attributions: stability and locus of control (Weiner et al., 1971). However, multiple bipolar scales could be used in such a way as to overcome the particular problems of using a single bipolar scale (Weiner, Nierenberg, & Goldstein, 1976). Weiner et al. (1976) limited comparisons on each scale to a difference within only one dimension (e.g., luck and ability are not on a scale since they differ on both the stability and the location of cause dimensions). Such a limitation solves the theoretical problem of which dimension the causes were judged on as well as easing the practical problem of the very large number of comparisons needed in a paired comparison of many attributions. Within-dimension bipolar scales (Weiner et al., 1976) yield more sensitive measures than paired comparisons used by McMahan (1973), since the range of possible scores is increased by the use of scales.

Among these structured response measures, separate rating scales and percentage measures appeared to be the most interesting measures. These were selected for further study along with an unstructured, open-ended measure.

A few researchers have begun to question the indiscriminate use of the various attribution measures (Deaux & Farris, Note 1; Smith, Note 2). Deaux and Farris analyzed a variety of measures by looking at differentiation across success and failure and at correlations across measures. They found some convergence, but they also found significant difference between what the various scales measure. This study extends their analysis by using a multitrait–multimethod test of the re-

liability of three of the most commonly used measures: open-ended measures, scale ratings, and percentage judgments. In addition, a formal assessment is made about subjects' perceptions of the face validity of the various measures.

Hypotheses

Two specific hypotheses that can be made concerning various attribution measures involve a comparison of the open-ended response measure with the two structured response measures: (a) Convergent and discriminant validities will be lower for the open-ended response measure than for either of the two structured response measures and (b) the face validity to subjects of the open-ended response question will be better than that obtained by either structured response measure. That is, subjects will rank the open-ended measure as better and easier to answer. Another hypothesis is made based on the apparent superiority of the independent scale judgments: (c) The independent ratings will be superior to the percentage ratings in terms of convergent and discriminant validities.

Method

Subjects

The subjects were 252 students in introductory social psychology and personality psychology classes. Students were requested to participate in a study that would serve as a focal point for class discussions of experimental and measurement procedures in psychology. Subjects were free to participate or not as they chose. Two of a total of 254 students elected not to participate because of their dislike of anagrams.

Attribution Measures

Three types of attribution measures were used. The measures were worded separately for success and failure events, and subjects were directed to answer only the questions for their own outcome. The first attribution measure was always an open-ended question asking the subject, "Why do you think you *succeeded* (*failed*) on this task?" The other attribution measures were structured response measures.

Eight causal attributions were assessed in the structured response measures. These included the four basic Weiner et al. (1971) causes: high/low general intelligence, task ease/difficulty, good/bad luck, and high/low unusual effort. These were selected for theoretical reasons and because of their frequent use in the literature. The other four attributions were high/low stable effort (stable = consistent), high/low task interest, good/bad mood, and high/low motivation. These have been suggested as important causal factors for this type of situation by open-ended studies (Elig & Frieze, 1975; Frieze, 1976). The attributions were worded in the format suggested by Elig and Frieze (1975) and by Valle and Frieze (1976).[1] Table 2 presents these causes in a three-dimensional taxonomy (cf. Elig & Frieze, 1975; Rosenbaum, 1972).

Two structured response measures were used: unipolar 9-point rating scales and percentage ratings (see Table 1). Both of these were administered with two alternate forms, (a) determination (e.g., "Please rate [indicate] how important you think each of the following factors was in determining your success or failure on the anagram task") and (b) cause (e.g., "Please rate [indicate] the extent to which each of the following factors caused your success or failure on the anagram task"). Each of these was then followed by instructions to respond to a set of success or failure causes. Exact wording of the causes is shown in Table 2 for the success and failure conditions. Structured response measures were (a) rating scale of determinance, (b) percentage scale of determinance, (c) rating scale of causality, and (d) percentage scale of causality. With the open-ended measure, this resulted in a total of five different attribution measures.

The two forms were needed for alternate-form reliability estimates for the scale and percentage methods of assessing causal attributions. Reliability estimates for the open-ended measure were based on intercoder reliability. Since these causal attributions represent a causal analysis of a particular success–failure event, other forms of reliability estimates such as test–retest or coefficient alpha were not appropriate.

Subjects were given the four structured attribution measures and one unstructured attribution measure in a within-subject design. Sixteen orders of the four structured measures were used in a $2 \times 2 \times 2 \times 2$ variation of (a) whether the two measures using the same method were together or separate, (b) whether the scale or percentage method was asked first, and (c and d) whether the determination form or the causality form was asked first for the percentage method and for the scale method. To an unknown extent, this within-subject design will lead to an inflation of intermethod correlation estimates. Subjects may strive for consistency in their responses even when they are instructed to start each measure afresh and not to look forward or back in the test sheets. Such a problem confronts any assessment of

[1] Another issue, not dealt with in this article, is the way in which attributions are worded. This also needs to be systematically analyzed.

Table 2
Measured Causal Attributions Varying According to Stability, Locus of Control, and Intentionality of the Attribution

Stability	Locus of control: Internal	Locus of control: External
Stable		
Unintentional	1. Your high (low) general intelligence	2. How easy (difficult) this type of task is
Intentional	5. The high (low) level of effort which you consistently demonstrate in whatever you do 8. Your high (low) desire to do well in everything you do	
Unstable		
Unintentional	7. Your good (bad) mood	3. Good (bad) luck
Intentional	4. How high (low) your interest in this task was 6. The unusually high (low) effort which you put forth in doing this task	

Note. Numbers indicate the order in which the attributions were presented. Failure wordings are in parentheses.

a momentary state rather than of a stable trait. However, since our interest is in the pattern of correlations rather than in the overall absolute levels, this problem was not too troublesome. Varying order also allowed the assessment of order effects, but since no consistent order effects were found, the results for various orders are collapsed in the discussion of results.

Procedure

The experiment was administered during class time by a male experimenter in five classes. Subjects were told the following:

> In a few minutes you will be asked to do a task consisting of 15 anagrams. An anagram is a group of scramble letters that can be unscrambled to form a word . . . . The brightest 25% of college students can solve at least eight of these anagrams. If you solve eight or more of these anagrams, you will have succeeded at the task; if you solve seven or less of the anagrams, you will have failed on this task. . . .

Subjects were given 30 sec to answer each of the 15 anagrams. The difficulty of the anagrams was manipulated so that half of the subjects received a set of mostly easy anagrams that led them to succeed (e.g., MNEGAA, WADNET, BOLWE), and the other half of the subjects was given a set of mostly difficult anagrams that led them to fail (e.g., SEALGT, IUMSC, OOLRUC) (Bar-Tal & Frieze, 1977).

After completing the anagram task, subjects were asked to total the number of anagrams solved and to rate their subjective feelings of success and failure on a 9-point scale. Subjects were then given as much time as they needed to answer the attribution questions previously discussed and a final set of questions that assessed the face validity of the attribution measures. Subjects were asked for their general impressions of the methods and then ranked them for difficulty and for global impressions of best and worst.

Subjects answered the questions at their own pace; no subject required more than 45 minutes. In each class, when all students were finished, the purpose and general nature of the study were discussed. All subjects were debriefed on the deceptions involved in the presentation of the anagrams task.

Results and Discussion

Although the majority of subjects responded to the success–failure manipulation of anagram difficulty, there were a few subjects who did not. Thirty-six subjects who were given easy anagrams failed to solve eight of them and therefore failed the success criterion. Seventeen other subjects made up words for enough of the unsolvable anagrams to score a "success." This raised a question about whether these 53 subjects should be included in the analyses. A preliminary analysis indicated that there were no systematic differences in the responses of these "erring" subjects as compared to the responses of the other subjects. Therefore, all subjects were included in the analyses listed below. Similar data were obtained when these 53 subjects were eliminated from the analyses.

Table 3
Open-Ended Responses

Attribution category	Equivalent structured measure	Intercoder reliability[a]	No. subjects mentioning	% codable responses[b]
Ability	Intelligence	.91	51	20.8
Task difficulty	Task difficulty	.99	65	26.5
Luck	Luck	1.00	2	.8
Intrinsic motives	Interest	.97	17	6.9
Stable effort	Stable effort	.57	38	15.5
Unstable effort	Unstable effort	.51	11	4.5
Mood	Mood	.91	34	13.9
Personality	—	.70	8	3.3
Task × Ability interaction	—	.74	19	7.8

[a] $n = 252$. These are correlations of the frequency of the attribution being coded for the subject.
[b] A total of 247 attributions were made that were codable; that is, they did not simply repeat the outcome nor explain the process of labeling the outcome a success or failure. One codable attribution was given by 102 subjects, whereas 64 subjects made more than one codable attribution.

Open-Ended Attributions

The responses to the open-ended attribution measure were coded according to the scheme developed by Elig and Frieze (1975). In this coding scheme, each subject's response is segmented into phrases containing individual causes, each of which can be scored for the specific attribution category and for the dimensions of stability, location of cause, and intentionality. The index of each attribution category (such as ability and task ease–task difficulty) used in this study is the number of times it was mentioned by the subject. Two coders coded all responses; the percentage of agreement for the codings ranged between 82.8% and 93.5%. Percentage agreement, although a useful concept, is not a reliability estimate. Reliability was estimated by using the correlations reported in Table 3.

Table 3 also lists the attributions found for the open-ended question, and it shows that all but a small proportion of these attributions were also measured by the structured response measures. Two types of responses are not shown in Table 3. Sixty-one responses (16.4% of total responses) simply repeated the outcome or denied the outcome. Almost all of these responses were from the subjects who had failed to follow instructions. Another 65 responses (17.5% of total responses) were uncodable as attributions. Most of these responses were from subjects who misunderstood the question as asking for the reasons why they labeled the outcome as a success or failure rather than asking for the causes of the outcome.

Intertest Correlation Validity

Interest correlation validities (convergent and discriminant) were assessed in multivariable–multimethod matrices (Campbell & Fiske, 1959; Magnusson, 1967). Since alternate forms (cause and determination) of the structured response methods were used, separate multivariable–multimethod matrices were prepared for each. Table 4 presents the causal forms of the structured method. The matrix for the determination forms was highly similar.[2]

Factor analysis procedures were used in the interpretation of these validities. The factor-analytic solution chosen was principal components analysis with varimax rotation. This solution accounts for reliable true variance with independent factors. We were most interested in the variance common to two or more attribution measures but had to allow

[2] For purposes of brevity, much of the specific comparison data obtained has been omitted from this article. These data are available on request from the authors.

Table 4
Multivariable–Multimethod Matrix

	Percent								Scale								Open response						
Attributions	I	T	L	I	S	U	M	M	I	T	L	I	S	U	M	M	A	T	L	M	S	U	M
Percent																							
Intelligence	.91																						
Task E–D[a]	−.27	.91																					
Luck	−.14	−.17	.97																				
Interest	−.31	−.17	−.24	.90																			
St. effort[b]	.10	−.43	−.17	−.18	.70																		
Us. effort[c]	−.16	−.33	−.10	.12	.19	.79																	
Mood	−.17	−.29	.00	−.12	−.06	.04	.88																
Motivation	.04	−.36	−.14	−.24	.43	−.02	−.09	.90															
Scale																							
Intelligence	*.58*	−.38	−.08	−.35	.43	−.03	−.11	.45	.85														
Task E–D[a]	−.10	*.49*	−.18	−.19	−.04	−.15	−.10	−.10	.03	.74													
Luck	−.03	−.15	*.63*	−.27	−.07	−.05	.08	−.03	.07	−.05	.77												
Interest	−.07	−.31	−.19	*.43*	.17	.22	.03	.02	.07	.07	−.08	.74											
St. effort[b]	.22	−.42	−.03	−.24	*.61*	.03	−.04	.55	.66	.07	.08	.29	.85										
Us. effort[c]	.08	−.38	−.08	.06	.28	*.43*	.01	.15	.26	.06	.03	.52	.52	.82									
Mood	.10	−.34	.07	−.22	.13	.07	*.53*	.07	.26	.04	.25	.27	.36	.35	.89								
Motivation	.30	−.40	−.08	−.27	.54	−.01	−.08	*.60*	.63	−.00	.02	.24	.87	.48	.36	.88							
Open response																							
Ability	*.06*	−.00	−.14	−.08	.08	.05	.06	.05	*.10*	−.05	−.12	−.03	.11	−.00	−.01	.06	.91						
Task E–D[a]	−.09	*.25*	−.00	−.13	−.02	−.03	−.01	−.14	−.17	*.09*	−.11	−.14	−.14	−.11	−.15	−.16	−.15	.99					
Luck	.00	−.07	*.16*	−.02	−.02	−.01	.02	−.02	.04	−.12	*.10*	−.04	−.03	−.08	.12	−.01	−.04	−.05	1.0				
Motives[d]	.04	−.04	−.09	*.06*	.06	−.01	−.03	.05	.13	.09	−.06	*.17*	.14	.09	.07	.13	.06	−.11	.14	.97			
St. effort[b]	−.05	−.01	−.08	−.01	*.03*	.06	−.08	.23	.00	.01	−.05	.07	*.08*	.11	−.06	.07	−.05	−.10	−.04	.21	.57		
Us. effort[c]	.01	.05	.05	−.01	.02	*.00*	−.09	−.03	.06	.10	.02	.03	−.01	*.03*	−.06	−.03	−.00	−.07	−.02	.02	.02	.51	
Mood	−.12	−.08	−.04	.00	−.07	.10	*.37*	−.06	−.16	−.10	−.03	.12	−.14	.03	*.22*	−.14	−.03	−.09	.08	.06	.00	−.02	.91

Note. Reliability estimates appear in the main diagonal. Italics indicate convergent validities. The cause alternate form is presented here for the percent and scale methods. (The determination alternate form is available from the authors.) The main diagonal is alternate form reliability for the structured methods and intercoder reliability for the open response method. $n = 252$.

[a] Task ease–difficulty.

[b] Stable effort.

[c] Unstable effort.

[d] Motives category includes the equivalent of both interest and motivation attributions.

Table 5
Factor Loadings of Cause and Open-Response Measures

Attribution/method	Factor 1	2	3	4	5	6	7	8	Communality
Stable effort/scale	.89								.87
Stable effort/percentage	.69								.57
Motivation/scale	.88								.84
Motivation/percentage	.79					−.27			.77
Intelligence/scale	.70					.50			.80
Mood/percentage		.87							.79
Mood/scale	.29	.71							.76
Mood/open		.61							.50
Unstable effort/percentage			.67	−.40					.67
Unstable effort/scale	.36		.78						.76
Interest/scale			.74						.63
Luck/scale				.86					.76
Luck/percentage				.87					.80
Task ease/difficulty/scale					.88				.78
Task ease/difficulty/ percentage	−.44		−.35		.62				.83
Intelligence/percentage	.39					.80			.78
Stable effort/open						−.53	.44		.56
Intrinsic motives/open							.73		.60
Luck/open							.57		.49
Task ease/difficulty/open								−.70	.60
Ability/open					−.28			.61	.55
Unstable effort/open								.48	.40

for specific factors to emerge. Table 5 gives the factor loadings of the factor analysis of the multivariable–multimethod matrix presented in Table 4.[3] This analysis accounts for 91% of the reliable variance.

The main diagonal of the matrix (Table 4) gives the reliability estimates for attribution-method unit. For the structured methods, these are alternate form estimates; for the open response method, they are intercoder reliabilities. With few exceptions, these are satisfactorily high. The open response measure shows low reliabilities for the effort attributions and shows generally low communalities indicating a need for further refinement of coding instructions to achieve more uniform, acceptable reliabilities for this measure.

The monomethod triangles (under the reliability diagonals) show the intercorrelations of the attributions measured the same way. It is obvious that the open response measure shows the smallest (in absolute value) intercorrelations, indicating either relatively good independence for the attributions measured or a poor representation of the true interrelationships of the attributions. The true correlations may be attenuated by the small range of frequencies of the open response attributions.

The monomethod triangle for the open response method is more similar to the triangle for the percentage than for the scale method. In part, this seems due to the quasi-ipsative nature of the method. No subject in this study gave more than four responses to the open-ended question; thus, once an attribution was made, there was less chance for any other attribution to be made also, inducing small negative correlations. Negative correlations were directly induced by the ipsative percentage measures. The relatively high positive intercorrelations of the attributions measured

[3] Because the percentage measures are ipsative, one percentage measure had to be dropped from the analysis. Interest was dropped in the analysis reported here. Though this is not a totally satisfactory solution, it does eliminate the singularity problem. Interest is included in other analyses, which are available from the authors.

by the scale method may in part have been due to a response bias of some subjects to use only one part of the 9-point scale. This was a response bias that cannot influence the percentage method or the open response method. Method variance can thus be seen to influence the monomethod intercorrelation pattern. The percentage method and the open response method seem to present a pattern different from the one evident in the scale measures.

Factor analyses can aid us in clarifying these patterns. Returning our attention to Table 5, we find the major evidence for method variance in the factor loadings of the open response measures, which form distinct method factors (Factors 7 and 8). However, with the exception of the intelligence attribution, the structured measures of a specific attribution both load on the same factor (see Footnote 3). Also, the monomethod blocks of the structured methods can be factor analyzed separately to clarify their patterns of correlations and the role of method variance. (The open response measure cannot be factor analyzed within method, since there is no alternate form for it as there are for the structured measures.) Tables 6 and 7 present the factor analyses of percentage and scale measures, respectively. The factors in these analyses account for 97% and 100% of the reliable variance of the percentage and scale measures, respectively. The strongest similarity between the percentage and scale monomethod triangles of the multivariable–multimethod matrix was that in each triangle, the largest correlation was for stable effort with motivation (motives in the open measure). These two attributions form the nucleus of the strongest factor (which appears to be a stable achievement factor) in all analyses. The analyses also agree to having mood and luck constituting independent factors.

There are some disagreements between the scale and percentage measures. Factor 1 includes task ease–difficulty for percentage measures and intelligence for scale measures. The factor analysis of the scale measures is more easily interpreted than that of the percentage measures. Factor 2 of the scale analysis can be called *unstable achievement*. Interest in a task should generate unusual effort leading to achievement. Interest and unstable effort correlate when measured by scale cause and scale determination, .52 and .43, respectively, indicating only 27% and 18% overlap in variance, respectively.

Tables 6 and 7 exhibit remarkably clear

Table 6
Factor Loadings of Percentage Measures

Attribution–alternate form	Factor 1	2	3	4	5	Communality
Motivation–cause	.88					.81
Motivation–determination	.87					.79
Stable effort–determination	.65					.49
Stable effort–cause	.72					.61
Task ease/difficulty —cause	−.55	−.39	−.40	−.41	−.33	.89
Task ease/difficulty —determination	−.55	−.41	−.37	−.36	−.36	.86
Intelligence–cause		.95				.93
Intelligence–determination		.95				.93
Unstable effort–cause			.91			.84
Unstable effort–determination			.91			.85
Mood–determination				.95		.92
Mood–cause				.95		.92
Luck–determination					.97	.96
Luck–cause					.97	.96

Note. Interest was dropped from this analysis because of the ipsative nature of percentage measures.

Table 7
Factor Loadings of Scale Measures

Attribution–alternate form	Factor 1	2	3	4	5	Communality
Motivation–cause	.88	.25				.86
Motivation–determination	.87	.26				.84
Stable effort–cause	.86	.29				.85
Stable effort–determination	.84	.31				.82
Intelligence–determination	.84					.73
Intelligence–cause	.84					.73
Interest–determination		.82				.73
Interest–cause		.85				.71
Unstable effort–cause	.36	.77				.74
Unstable effort–determination	.39	.71				.69
Mood–cause			.91			.94
Mood–determination			.91			.94
Luck–cause				.93		.89
Luck–determination				.92		.88
Task ease/difficulty —cause					.93	.87
Task ease/difficulty —determination					.93	.88

structures. The alternate forms of the structured methods have similar loadings, substantiating their interchangeability.

Returning our attention to Table 4, we note the convergent validities. The convergence of the two structured response methods seems high enough to warrant our further attention. The range of these validities is .43 to .63 for cause measures and .30 to .69 for determination measures. For atttributions measured by cause, there is a mean overlap of 31.1% in variation of the structured methods, whereas the mean is 30.4% for the determination measures.

The convergence of the two structured methods can be clearly seen in Table 5. In this analysis, the scale and percentage measures of an attribution have their highest loading on the same factor except for intelligence. The two structured measures of intelligence converge to the point of loading on the same two factors. (Interest measured by percentages was not in the analysis; see Footnote 3.)

The convergent validities of the open response measure with the structured measures are woefully low (ranging from .00 to .37). These figures indicate that two studies claiming to measure the same attribution, one with a structured measure and one with the open response measure, will only have an average of 4.7% of the variance in common (and a maximum of 14%), with the remainder of the variance being due to noncommon factors (method and error). Mood is the only open response measure to show any amount of convergence. It correlated .37 with the percentage measure and .22 with the scale measure. Mood is the only open response measure to load on the same factor as its structured equivalent (see Table 5).

On either side of the convergent validity diagonals lie heterovariable–heteromethod triangles. Correlations in these triangles are of variables having neither trait nor method variance in common. From this consideration follows the second type of validity, discriminant validity. Discriminant validity is shown when a validity diagonal value is higher than the values lying in its column and row in the heterovariable–heteromethod triangles. For example, attributions to intelligence measured in percentages should correlate more highly with intelligence measured by a scale than they correlate with any other attribution measured by a scale.

The structured responses discriminate fairly well between variables. For the cause form,

no convergent validity estimate is exceeded by a heterovariable–heteromethod value. In the determination form, only task ease–difficulty is problematic. Its convergent validity is exceeded by four other values. This relative lack of discrimination showed up in the factor analysis of the determination form (Table 6) by task ease–difficulty loading somewhat on every factor.

The discrimination shown by the open response method is not nearly as good. Only mood always discriminated. The other attributions failed to show discrimination an average of 4.7 times out of 13 comparisons. It seems reasonable to blame the open response rather than the structured response measures for this lack of discrimination, since good discrimination was shown between the structured measures themselves.

As predicted, results indicate some superiority of structured measures. Structured response reliabilities are higher and are substantiated by communalities, whereas the open response measure intercoder reliabilities are not substantiated by communalities. Convergent and discriminant validities for structured measures are satisfactory, whereas open response convergent validities are low (except for mood). Expect for mood, other open response convergence values are exceeded by an average of one third of the values in their row and column of the heterovariable–heteromethod triangles, indicating poor discrimination.

Face Validation

The second hypothesis concerns the reactions of those subjects whose attributions were being measured. This hypothesis states that to subjects, the face appearance of the open response measure will be more positive than that of either structured response measure. Table 8 presents the frequency with which the subjects termed each method easiest, hardest, best, and worst. As can be seen, subjects differed in ranking the methods on all criteria. Contrary to the hypothesis, however, the differences are not between the open response method compared with structured methods. There were no significant differences between the open response and scale methods; even the largest difference between these methods on the evaluation of best is not significant, $\chi^2(1) = 2.4$. Differences arose from a general disliking of the percentage method. Subjects said they felt that the percentage measure was hard to compute and was not the best reflection of what they felt were the reasons for the outcome.

Table 8
Frequency of Attribution Measurement Methods Selected as Easiest, Hardest, Best, and Worst

	Method			
Evaluation	Open response	Scale	Percent	χ^2
Easiest	76	79	47	9.282**
Hardest	52	49	97	21.909***
Best	77	58	47	7.595*
Worst	45	51	89	18.462***

* $p < .05$.
** $p < .01$.
*** $p < .001$.

Outcome Effects

The final analysis concerned the variations across the methods for success as compared with failure conditions. Results of the analysis of variance for each attribution are presented in Table 9. As can be seen, success was more attributed than failure to ability, stable effort, and motivation for all four structured measures. These outcome effects were not found for the open response measures. There was convergence of the open response measure and both forms of the percentage method (but not the scale method) in that both had greater attributions to low interest and task difficulty for failure than to high interest or task ease for success. Both scale measures also yielded an outcome effect such that success was more attributed to luck, unstable effort, and mood.

Conclusions

As predicted, open-ended response measures of causal attributions have poorer intertest validity and reliability than structured response measures. However, contrary to prediction, structured scale measures were seen

Table 9
Mean Attributions for Success and Failure

Outcome and attribution	Structured response measure				Open response measure
	Percentage		9-point scale		
	Cause	Determination	Cause	Determination	
Success					
High intelligence	22.10***	21.41***	5.42***	5.54***	.19
Task ease	17.02***	16.55***	5.14	5.27	.15**
Good luck	8.47	8.04	2.97	3.06**	.02
High interest	11.71***	11.24***	4.86	5.08	.11*
High stable effort	12.40***	12.60***	5.43	5.58	.15
High unstable effort	7.50	7.76	4.57***	4.91***	.05
Good mood	6.60	6.80	4.16***	4.32***	.09
High motivation	14.35***	14.90***	6.21***	6.02***	—
Personality	—	—	—	—	.03
Task–ability match	—	—	—	—	.05
Failure					
Low intelligence	8.89	9.72	2.31	3.38	.21
Task difficulty	32.27	32.63	5.02	4.90	.34
Bad luck	7.64	7.33	2.33	2.30	.00
Low interest	22.50	21.77	4.51	4.96	.03
Low stable effort	4.85	4.18	2.08	2.36	.15
Low unstable effort	9.03	8.82	3.14	3.34	.04
Bad mood	9.13	9.37	2.65	2.74	.16
Low motivation	5.38	5.49	2.11	2.37	—
Personality	—	—	—	—	.03
Task–ability mismatch	—	—	—	—	.10

Note. Significant difference between success and failure attributions are indicated on the success means.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

by subjects as easy to respond to, like the open-ended measures. Percentage ratings were seen by these subjects as having the least face validity. Thus, one of the postulated advantages of the percentage measure was not confirmed. And a supposed weakness of scale measures was not found. Thus, at least for college students solving anagrams or doing similar tasks in which the basic causal categories are well understood, the scale method is clearly a superior technique. This strong support for the scale method of independent assessment of the contribution of various causal factors also confirms findings of Deaux and Farris (Note 1), who used other criteria for their conclusions. They also found little support for the percentage method.

Although open-ended procedures are weaker on the basis of psychometric criteria than scale ratings, they have utility for the researcher who is asking for causal attributions in a new situation. Clearly, people use different categories of causal explanations in different settings (e.g., Elig & Frieze, 1975), and without open-ended pretesting, the experimenter cannot know which causal factors to include in structured measures. Using an open-ended question in later stages of research can serve as a continuing validation of the attribution scales that subjects are asked to rate.

Data from the factor analyses and the success–failure comparisons suggest that open-ended and scale and percentage ratings not only vary in their psychometric properties but also yield different types of data. Hypotheses that concern particular causal attributions can be supported by one type of measurement

and disconfirmed by another; this was demonstrated in the outcome effects reported here.

The attributional measures discussed in this article have been the measures typically used and are representative of three general methods of attribution measurement. Other methods and other within-method variations are possible (e.g., "How much ability did he have?" vs. "How much was ability a cause of his success?"). Many of these possible variations have been used in attribution research. As indicated in this article, these measurement variations may account for some of the discrepancies in the attribution literature.

Fiske (1971) has suggested that in cases in which multiple measures yield conflicting results, the construct be operationally defined by a single measure. These are multiple criteria for selection of a measure, but many of them would lead to the use of the scale measure for attribution research. Results reported here indicate that scale measures have moderately good intermethod correlations with percentage measures, do not force intercorrelations among attributions, and have good face validity. Another important criterion for selecting the specific measurement technique that defines the construct is the construct validity of various procedures. Other data not reported here (see Elig & Frieze, Note 6) indicate that the scale method used in this study provides generally better support for some of the basic theoretical relationships between causal attributions and future expectancies and affect than do either the percentage or open response methods. Even if attribution researchers do not agree on a single measurement technique, these reliability and validity issues should be carefully considered before selecting the measures to be used in future investigations of causal attributions.

Reference Notes

1. Deaux, K., & Farris, E. *Causal attributions for performance: Approaching substance via method.* Unpublished manuscript, Purdue University, 1977.
2. Smith, E. R. *On the concept and measurement of cause in studies of causal attribution.* Paper presented at the annual meeting of the American Psychological Association, San Francisco, August 1977.
3. McHugh, M. C. *Sex differences in causal attributions: A critical review.* Paper presented at the annual meeting of the Eastern Psychological Association, New York, April 1975.
4. Cooper, H. M., & Burger, J. M. *Categorizing open ended academic attributions: Replication of earlier findings.* Paper presented at the annual meeting of the American Psychological Association, Toronto, August, 1978.
5. McHugh, M. C. *Attributions of causality made by alcoholics and nonalcoholics for success and failure in selected situations.* Unpublished manuscript, University of Pittsburgh, 1978.
6. Elig, T. W., & Frieze, I. H. *Construct validity of causal attributions for success and failure.* Manuscript in preparation, University of Pittsburgh, 1978.

References

Bailey, R. C., Helm, B., & Gladstone, R. The effects of success and failure in a real-life setting: Performance, attribution, affect, and expectancy. *Journal of Personality,* 1975, *89,* 137–147.

Bar-Tal, D., & Darom, E. Pupils attributions for success and failure. *Child Development,* in press.

Bar-Tal, D., & Frieze, I. H. Achievement motivation for males and females as a determinant of attributions for success and failure. *Sex Roles,* 1977, *3,* 301–313.

Campbell, D. T., & Fiske, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin,* 1959, *56,* 81–105.

Elig, T. W., & Frieze, I. H. A multi-dimensional coding scheme of causal attributions in social and academic situations. *Personality and Social Psychology Bulletin,* 1974, *1,* 94–96.

Elig, T. W., & Frieze, I. H. A multi-dimensional scheme for coding and interpreting perceived causality for success and failure events. JSAS *Catalog of Selected Documents in Psychology,* 1975, *5,* 313. (Ms. No. 1069)

Feather, N. T. Attributions of responsibility and valence of success and failure in relation to initial confidence and task performance. *Journal of Personality and Social Psychology,* 1969, *13,* 129–144.

Feather, N. T., & Simon, J. G. Attribution of responsibility and valence of outcome in relation to initial confidence and success and failure of self and other. *Journal of Personality and Social Psychology,* 1971, *18,* 173–188. (a)

Feather, N. T., & Simon, J. G. Causal attributions for success and failure in relation to expectations of success based upon selective or manipulative control. *Journal of Personality,* 1971, *39,* 527–541. (b)

Feather, N. T., & Simon, J. G. Luck and the unexpected outcome: A field replication of laboratory findings. *Australian Journal of Psychology,* 1972, *24,* 113–117.

Fiske, D. W. *Measuring the concepts of personality.* Chicago: Aldine, 1971.

Frieze, I. H. Causal attributions and information seeking to explain success and failure. *Journal of Research in Personality,* 1976, *10,* 293–305.

Frieze, I. H., & Weiner, B. Cue utilization and attributional judgments for success and failure. *Journal of Personality,* 1971, *39,* 591–606.

Heider, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

Magnusson, D. [*Test theory*] (H. Mabon, trans.). Reading, Mass.: Addison-Wesley, 1967.

McMahan, I. D. Relationships between causal attributions and the expectancy of success. *Journal of Personality and Social Psychology,* 1973, *28,* 108–114.

Meyer, W. U. *Selbstverantwortlichkeit und Leistungsmotivation.* Unpublished doctoral dissertation, Ruher Universitat, Bochum, Germany, 1970.

Rosenbaum, R. M. *A dimensional analysis of the perceived causes of success and failure* (Doctoral dissertation, University of California, Los Angeles, 1972). *Dissertation Abstracts International,* 1973, *33,* 5040B. (University Microfilms No. 73-10,475)

Valle, V. A., & Frieze, I. H. Stability of causal attributions as a mediator in changing expectations for success. *Journal of Personality and Social Psychology,* 1976, *33,* 579–587.

Weiner, B. Achievement motivation as conceptualized by an attribution theorist. In B. Weiner (Ed.), *Achievement motivation and attribution theory.* New York: General Learning Press, 1974.

Weiner, B., et al. *Perceiving the causes of success and failure.* New York: General Learning Press, 1971.

Weiner, B., Heckhausen, H., Meyer, W., & Cook, R. E. Causal ascriptions and achievement behavior: A conceptual analysis of effort and reanalysis of locus of control. *Journal of Personality and Social Psychology,* 1972, *21,* 239–248.

Weiner, B., & Kukla, A. An attributional analysis of achievement motivation. *Journal of Personality and Social Psychology,* 1970, *15,* 1–20.

Weiner, B., Nierenberg, R., & Goldstein, M. Social learning (locus of control) versus attributional (causal stability) interpretations of expectancy of success. *Journal of Personality,* 1976, *44,* 52–68.

Received June 23, 1978 ■

Editors Named for New *Journal of Personality and Social Psychology*

The Publications and Communications Board has named Melvin Manis, Ivan D. Steiner, and Robert Hogan to edit the *Journal of Personality and Social Psychology* (JPSP), which will appear in a new format in January 1980. The affiliations of the new editors are, respectively, University of Michigan and Veteran's Hospital, Ann Arbor; University of Massachusetts; and Johns Hopkins University. Manis will edit the section on ATTITUDES AND SOCIAL COGNITION, Steiner the section on INTERPERSONAL RELATIONS AND GROUP PROCESSES, and Hogan the section on PERSONALITY PROCESSES AND INDIVIDUAL DIFFERENCES.

An announcement on the restructuring of JPSP appeared on p. 668 of the June 1978 issue of JPSP. As indicated at that time, the three editors will work independently with separate groups of advisory editors. Their acceptances will appear as distinct sections in monthly issues of JPSP. Detailed descriptions of each section's coverage appear on pages 468–469 of the March issue of JPSP.

Authors should submit manuscripts to the appropriate section editor as follows:

ATTITUDES AND SOCIAL COGNITION

April 1 through June 15:

Melvin Manis
c/o Anita DeVivo
Suite 700
1400 North Uhle Street
Arlington, Virginia 22201

After June 15:

Melvin Manis
Research Center for Group Dynamics
Institute for Social Research
University of Michigan
Ann Arbor, Michigan 48104

INTERPERSONAL RELATIONS AND GROUP PROCESSES

Ivan D. Steiner
Department of Psychology
University of Massachusetts
Amherst, Massachusetts 01003

PERSONALITY PROCESSES AND INDIVIDUAL DIFFERENCES

Robert Hogan
Johns Hopkins University
Department of Psychology
33rd and Charles Streets
Baltimore, Maryland 21218

PSYCINFO™

Offering you a family of database services in psychology and related behavioral and social sciences. A trusted name established through reliable services available in the forms users need.

ONLINE SERVICE

Nearly 300,000 database entries available for searching and review through your interactive terminal whenever you need access.

PRINT PRODUCTS

Psychological Abstracts monthly journal and subject and author indexes in six month and multiple year cumulations.

TAPE LEASES & LICENSES

PsycINFO monthly updates and historical volumes on magnetic tape ready for on site processing by information centers.

DEMAND SEARCHES

The Psychological Abstracts Search and Retrieval service (PASAR) bringing our staff expertise to bear on your retrospective search topics from strategy formulation through final review.

PsycINFO offers effective reference tools and training services for online and manual utilization of the database:

- **The Thesaurus of Psychological Index Terms** (2nd Edition)
- **The PsycINFO Users Reference Manual**
- PsycINFO user training seminars

We welcome your inquiries or comments.

CALL OR WRITE.

NEW!

PsycINFO Action Line (800) 336-4980
(Toll free in the continental U.S.)

PsycINFO, 1200 Seventeenth St., N.W., Washington, D.C. 20036, (202) 833-7600
A service of the American Psychological Association

Bargaining Behavior

Edited by Heinz Sauermann. 1978. X, 375 pp. (Contributions to Experimental Economics, Vol. 7). Cloth DM 78.–

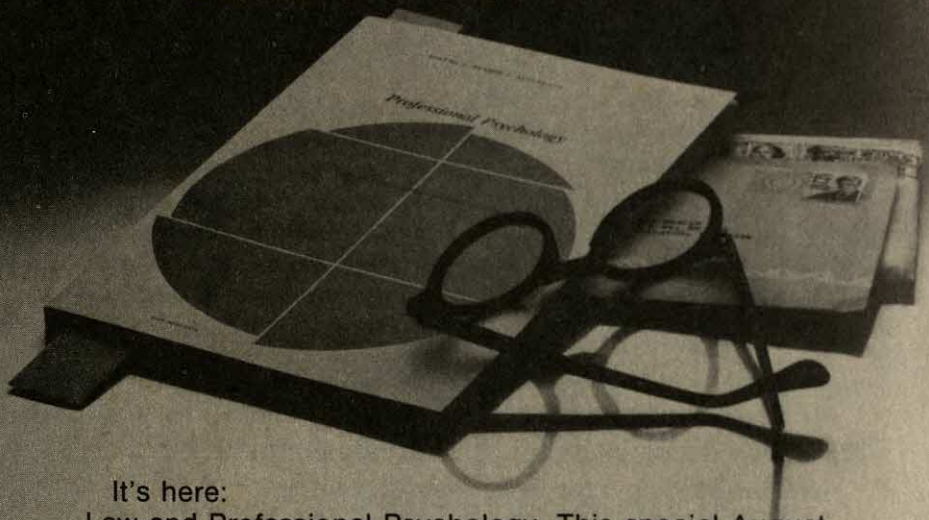
Coalition Forming Behavior

Edited by Heinz Sauermann. 1978. X, 375 pp. (Contributions to Experimental Economics, Vol. 8). Cloth DM 78.–

Both volumes contain recent experimental results in the field of multivariable decision behavior. Numerous factors which influence the behavior of subjects in conflict situations are isolated and analyzed. The necessity of interdisciplinary cooperation to develop adequate theories of conflict resolution becomes obvious. Both books present various behavioral approaches and theoretical concepts to explain the observed decision behavior of subjects and to improve the traditional decision theories. Volume 7 focuses on bilateral bargaining problems. In Volume 8 the decision behavior in multilateral decision situations is analyzed. Most contributions discuss the coalition forming behavior especially in characteristic function games.

J.C.B. Mohr (Paul Siebeck)
Tübingen

NOW: INFORMATION ABOUT THE LAW AND PSYCHOLOGY.



It's here:
Law and Professional Psychology. This special August 1978 issue of **Professional Psychology** brings you thirteen original articles full of fresh thought and constructive recommendations on the issues that vitally affect your day-to-day practice. You'll find precautions you may take to protect yourself as a professional and a plan for improving investigations of malpractice claims. Expert witness testimony, jury selection, civil commitment, confidentiality and privilege, minor's consent to treatment, and the use of psychological devices are critically examined.

Send your order today. And soon you'll receive this special issue.

Single copies of the issue are available for $5 each. All orders $25 or less must be prepaid. Make and send checks payable to:

American Psychological Association
Order Department
1200 17th Street, N.W.
Washington, D.C. 20036

Enclosed is $_____ for ___ copies of **Professional Psychology's** special issue—Law and Professional Psychology—at $5.00 each.

NAME _____

ADDRESS _____

CITY _____

STATE _____ ZIP CODE _____

PP-8

(Continued from front cover)

Erotica and Aggression: The Influence of Sexual Arousal, Positive Affect, and Negative Affect on Aggressive Behavior 591
Leonard A. White

Predictions of Others' Responses in a Mixed-Motive Game: Self-Justification or False Consensus? 602
Lawrence A. Messé and John M. Sivacek

Information, Choice, and Reactions to Stress: A Field Experiment in a Blood Bank With Laboratory Analogue 608
Richard T. Mills and David S. Krantz

Measuring Causal Attributions for Success and Failure 621
Timothy W. Elig and Irene Hanson Frieze

Editors Named for New *Journal of Personality and Social Psychology* 635

Understanding Social Psychology

Revised Edition *Just Published*

Worchel & Cooper

Stephen Worchel
University of Virginia

Joel Cooper
Princeton University

The **Revised Edition** of **Understanding Social Psychology** keeps abreast of the rapid growth in the field of social psychology. Each chapter has been thoroughly revised and updated to reflect current trends. This currency has been achieved without sacrificing the historical perspective presented in the first edition.

Special Features

- ★ "Real world" incidents are used at the beginning of each chapter.
- ★ Excellent coverage of attribution theories.
- ★ Very thorough coverage of all traditional areas of social psychology.
- ★ Separate chapter on environmental psychology.
- ★ Includes update on classic theories and research.

Changes

- ★ New chapter on socialization.
- ★ New coverage of: locus of control, learned helplessness, biases in attribution, social power, trust, groups, conflict mediation, theories of crowding.
- ★ Glossary.
- ★ Suggested readings at the end of the chapters.
- ★ **Student Study Guide** written by Carl I. Greenberg, University of Nebraska, Omaha, with learning objectives, multiple choice questions, essays, and student projects.

PERSONALITY: Strategies and Issues, Third Edition
Robert M. Liebert, State University of New York at Stony Brook, and
Michael D. Spiegler, Providence College
Published in 1978

Examination copies for adoption consideration available on request; please indicate course title and text presently used.

THE DORSEY PRESS • Homewood, Illinois 60430

Library

JOURNAL OF

Personality AND Social Psychology

VOLUME 37 NUMBER 5 · MAY 1979

The Popularity of Conspiracy Theories of Presidential Assassination: A Bayesian Analysis 637
Clark McCauley and Susan Jacques

It is not irrational to judge that successful assassinations are more likely to be conspiracies than are failed attempts.

Implicit Theories of Relationship: An Intergenerational Study 645
Marylyn Rands and George Levinger

Senior citizens' and college students' judgments of pair interactions of own-generation young adults were analyzed by multidimensional scaling.

Evaluation of Outcome Improvement Resulting From Threats and Appeals 662
André deCarufel and John Schopler

Thematic Effects of Person Judgments on Impression Organization 674
John H. Lingle, Nehemia Geva, Thomas M. Ostrom, Michael R. Leippe, and Michael H. Baumgardner

Effects of Preparatory Information About Sensations, Threat of Pain, and Attention on Cold Pressor Distress 688
Howard Leventhal, Donald Brown, Saya Shacham, and Gretchen Engquist

Effects of Pitch and Speech Rate on Personal Attributions 715
William Apple, Lynn A. Streeter, and Robert M. Krauss

Smiles Can Be Back Channels 728
Lawrence J. Brunner

Motoric and Symbolic Mediation in Observational Learning 735
Seymour M. Berger, Linda L. Carli, Kathy S. Hammersla, Judith F. Karshmer, and M. Estela Sanchez

A Step in the Socialization of Altruism as Hedonism: Effects of Negative Mood on Children's Generosity Under Public and Private Conditions 747
Douglas T. Kenrick, Donald J. Baumann, and Robert B. Cialdini

The Organizational Functions of the Self: An Alternative to the Duval and Wicklund Model of Self-Awareness 756
Jay G. Hull and Alan S. Levy

Disruptive Effects of Disconfirmed Expectancies About Crowding 769
Kitty Klein and Bruce Harris

Persistence of Opinion Change Induced Under Conditions of Forewarning and Distraction 778
William A. Watts and Lewis E. Holt

(Continued on inside back cover)

Published monthly by the

AMERICAN PSYCHOLOGICAL ASSOCIATION

ISSN 0022-3514

Editor

Anthony G. Greenwald, *Ohio State University*

Topic Editors

Irwin Altman, *University of Utah*
Robert A. Baron, *Purdue University*
Daryl Bem, *Stanford University*
Eugene Burnstein, *University of Michigan*
Robert B. Cialdini, *Arizona State University*
Reid Hastie, *Harvard University*
Douglas N. Jackson, *University of Western Ontario*
Leslie Z. McArthur, *Brandeis University*

Consulting Editors

Icek Ajzen, *University of Massachusetts—Amherst*
Robert S. Baron, *University of Iowa*
Gordon Bear, *Ramapo College of New Jersey*
Ellen Berscheid, *University of Minnesota*
Michael Birnbaum, *University of Illinois at Urbana-Champaign*
Philip Brickman, *Institute for Social Research*
J. Merrill Carlsmith, *Stanford University*
John S. Carroll, *Loyola University of Chicago*
Jerry L. Cohen, *University of Illinois at Urbana-Champaign*
Sheldon Cohen, *University of Oregon*
Rick Crandall, *Texas Christian University*
William D. Crano, *Michigan State University*
Robyn M. Dawes, *University of Oregon*
Kay K. Deaux, *Purdue University*
Marshall Dermer, *University of Wisconsin—Milwaukee*
Edward Diener, *University of Illinois at Urbana-Champaign*
Anthony Doob, *University of Toronto, Toronto, Canada*
Susan V. Eisen, *McLean Hospital, Belmont, Massachusetts*
Baruch Fischhoff, *Decision Research, Eugene, Oregon*
Irene H. Frieze, *University of Pittsburgh*
David Funder, *Stanford University*
David L. Hamilton, *University of California, Santa Barbara*
Ranald D. Hansen, *Oakland University*
Richard J. Harris, *University of New Mexico*
Edward Helmes, *University of Western Ontario, London, Canada*
Donald S. Hiroto, *VA Hospital, Brentwood, Los Angeles, CA*
Edward E. Jones, *Princeton University*
Martin F. Kaplan, *Northern Illinois University*
Douglas T. Kenrick, *Montana State University*
Norbert L. Kerr, *University of California, San Diego*
Samuel Komorita, *University of Illinois at Urbana-Champaign*
Vladimir J. Konečni, *University of California, San Diego*
Robert M. Krauss, *Columbia University*
Ellen J. Langer, *Harvard University*
Mark R. Lepper, *Stanford University*
Melvin Manis, *University of Michigan (Ann Arbor)*
Hazel Markus, *University of Michigan (Ann Arbor)*
Margaret Matlin, *SUNY College at Geneseo*
Dale T. Miller, *University of British Columbia, Vancouver*
Thomas M. Ostrom, *Ohio State University*
Paul B. Paulus, *University of Texas at Arlington*
Richard Petty, *University of Missouri—Columbia*
Harry T. Reis, *University of Rochester*
Judith Rodin, *Yale University*
Seymour Rosenberg, *Livingston College, Rutgers*
Ralph L. Rosnow, *Temple University*
Lee Ross, *Stanford University*
Michael Ross, *University of Waterloo, Waterloo, Canada*
Diane Ruble, *Princeton University*
Michael J. Saks, *Nat'l Center for State Courts, Williamsburg, VA*
Michael F. Scheier, *Carnegie-Mellon University*
Barry R. Schlenker, *University of Florida*
Philip Shaver, *New York University*
Steven J. Sherman, *Indiana University*
Jerome E. Singer, *University of the Health Services, Bethesda, Maryland*
Eliot R. Smith, *University of California, Riverside*
Mark Snyder, *University of Minnesota*
Melvin L. Snyder, *University of Texas at Austin*
Lawrence Stricker, *Educational Testing Service*
Shelley E. Taylor, *Harvard University*
Abraham Tesser, *University of Georgia*
Bernard Weiner, *University of California, Los Angeles*
Gary L. Wells, *University of Alberta, Edmonton, Canada*
Stephen Worchel, *University of Virginia*
Robert S. Wyer, *University of Illinois at Urbana-Champaign*
Mark Zanna, *University of Waterloo, Waterloo, Canada*
Miron Zuckerman, *University of Rochester*

Assistant to the Editor
Holly Wong

APA Journal Staff

Anita DeVivo, *Executive Editor*
Ann I. Mahoney, *Manager, Journal Production*
Barbara R. Richman, *Production Supervisor*
Marjorie M. Henry, *Production Editor*
Robert J. Hayward, *Advertising Representative*
Juanita Brodie, *Subscription Manager*

The Journal of Personality and Social Psychology publishes original reports of research, methodology, theory, criticism, and literature review in the fields of personality and social psychology of normal processes and populations. Submitted manuscripts should delineate an important question and provide a clear answer to it. In the case of research reports, this will often require demonstration that the reported results are generalizable beyond a single research setting. Consult the editorial in the January 1976 issue (Vol. 33, pp. 1–7) for a detailed statement of policies and instructions to authors.

Manuscripts: Submit manuscripts in quadruplicate, prepared according to the *Publication Manual of the American Psychological Association* (2nd ed.). Copies should be highly legible and will not be returned to authors unless they are original typewritten copies. Mail manuscripts according to instructions on p. 809.

Required supplementary information: Manuscripts cannot be reviewed for publication unless the author provides supplementary information in the following categories: (a) a cover letter that indicates the relation (or the lack of relation) of the submitted manuscript to other papers by the same author(s) that are currently submitted, in press, or recently published; (b) one copy of each such related paper; (c) one copy of a statement of the procedures of the research reported in the submitted manuscript in sufficient detail to serve as a guide for conduct of an exact replication; and (d) one copy of the summary tables (e.g., correlation tables, analysis of variance summaries) for the *major* analyses reported. If it appears that the supplementary materials in categories (c) or (d) might be bulky, the author should exercise judgment and restaint in selecting what to send. It is not necessary to send supplementary material that repeats verbatim manuscript content.

After a manuscript is accepted, authors will be expected to provide assurance that their data are available in unanalyzed form for reanalyses by other investigators and that their research procedures were consistent with the American Psychological Association's published principles of research ethics, except insofar as may be detailed otherwise in the manuscript.

Copyright and Permission: All rights reserved. Written permission must be obtained from the American Psychological Association for copying or reprinting text of more than 500 words, tables, or figures. Permission is normally granted contingent upon like permission of the author, inclusion of the APA copyright notice on the first page of reproduced material, and payment of a fee of $10 per page, table, or figure. Abstracting is permitted with credit to the source. Permissions and fees are waived for authors who wish to reproduce their own material for personal use; fees only are waived for authors who wish to use their own material commercially. Permission and fees are waived for the photocopying of isolated articles for nonprofit classroom or library reserve use by instructors and educational institutions. Libraries are permitted to photocopy beyond the limits of U.S. copyright law: (1) those post-1977 articles with a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through the Copyright Clearance Center, P.O. Box 765, Schenectady, NY 12301; (2) pre-1978 articles, provided that the per-copy fee stated in the Publishers' Fee List is paid through the Copyright Clearance Center, P.O. Box 765, Schenectady, NY 12301. Address requests for reprint permission to the Permissions Office, APA, 1200 Seventeenth Street, N.W., Washington, D.C. 20036.

Subscriptions: Subscriptions are available on a calendar year basis only (January through December). Nonmember rates for 1979: $60 domestic, $64 foreign, $6 single issue. APA member rate: $20. Write to Subscription Section, APA.

Change of Address: Send change of address notice and a recent mailing label to the attention of the Subscription Section, APA, 30 days prior to the actual change of address. APA will not replace undelivered copies resulting from address changes; journals will be forwarded only if subscribers notify the local post office in writing that they will guarantee second-class forwarding postage.

Reprints: Authors may order reprints of their articles from the printer when they receive proofs.

Back Issues and Back Volumes: For information regarding back issues or back volumes write to Order Department, APA.

Microform Editions: For information regarding microform editions write to any of the following: Johnson Associates, Inc., P.O. Box 1017, Greenwich, Connecticut 06830; University Microfilms, Ann Arbor, Michigan 48106; or Princeton Microfilms, Princeton, New Jersey 08540.

Published monthly in one volume per year by the American Psychological Association, Inc., 1200 Seventeenth Street, N.W., Washington, D.C. 20036. Printed in the U.S.A. Second-class postage paid at Arlington, Va., and at additional mailing offices.

JOURNAL OF
Personality
AND
Social Psychology

Volume 37 May 1979 Number 5

The Popularity of Conspiracy Theories of Presidential Assassination: A Bayesian Analysis

Clark McCauley and Susan Jacques
Bryn Mawr College

Journalist Tom Bethell has advanced the hypothesis that conspiracy explanations of Presidential assassination are popular because people have an irrational need to explain big and important events with proportionately big and important causes. This is a species of consistency hypothesis and clearly predicts that a shot that kills the President is more likely than a miss to be attributed to a conspiracy. Four studies are reported that support this prediction. Three of the four studies provided a check on whether conspiracy was overly favored, in the case of successful assassination, by comparison with the normative Bayesian formulation. No evidence of this kind of departure from rationality was found. It appears that people associate conspiracy with successful assassination, not because of any kind of special need for proportionality of cause and effect, but because of a belief that conspiracies are more effective and successful than lone assassins.

John F. Kennedy was shot dead in Dallas in 1963, and the Warren Commission reported in 1964 that the assassination was the work of a lone assassin, Lee Harvey Oswald. In 1979, the issue is evidently not yet settled. The House Select Committee on Assassinations has concluded its work, still uncertain about a fourth shot. Polls indicate that the majority of Americans, around 80% in fact, believe that others besides Oswald were involved in the assassination (Gallup, 1976). Books and articles theorizing about the assassination are still appearing regularly. Recent examples of the genre include *They've Killed the President* by Robert Sam Anson (1975), *Coincidence or Conspiracy?* by Bernard Fensterwald (1977) and the Committee to Investigate Assassinations, and *Legend: The Secret World of Lee Harvey Oswald* by Edward Jay Epstein (1978). Clearly, Americans are not satisfied with the conclusion of the Warren Commission.

Whatever the merits or defects of the Warren Commission report, the continuing popularity of conspiracy theories is itself a remarkable fact. Most events, no matter how traumatic, do not last in the public awareness as the Kennedy assassination has. Most events, no matter how great, quickly drop from headlines to history. Accidents, scandals, great men, and even wars are left behind, forgotten or at least much faded. But

Requests for reprints should be sent to Clark McCauley, Department of Psychology, Bryn Mawr College, Bryn Mawr, Pennsylvania 19010.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3705-0637$00.75

Diary No. 69
Date 2/11/1979

new books about John F. Kennedy's assassination are selling 15 years after the event, selling in supermarkets and in drug stores against competition from pop psychology and sex books. It is the premise of this article that the continued popularity of conspiracy theories of the Kennedy assassination is a surprising social fact that is worthy of investigation.

The studies reported here are aimed at testing one hypothesis about the popularity of conspiracy theories, namely, that people irrationally seek big causes to explain big effects. According to this hypothesis, advanced by journalist Tom Bethell in *The Washington Monthly* (1975), a lone assassin is too small and insignificant a cause to provide a satisfactory explanation for such large effects on policy and people as follow a President's death.

> We are expected to believe, according to the official explanation, that the Johnson Administration and all that it entailed, possibly including the debacle of Vietnam, was set in motion by one man who had quarreled with his wife; who had, as it were, gotten out of bed on the wrong side that morning, and found a gun lying there.
>
> The cause doesn't fit the effect. But the fact is, when great power is vested in one man, as in the President of the United States, it is always possible that a small cause (a microbe in his blood, for example, leading to a fatal disease, leading to a new President, leading to a "Vietnam") can trigger a large effect.
>
> In such cases many people will seek a new cause that is commensurate with the effect—seek, in other words, large and global explanations that thereby imbue the event with appropriate meaning. In the case of the Kennedy assassination, of course, this means looking for a conspiracy—preferably a large one. (Bethell, 1975, p. 39)

Bethell's hypothesis is recognizable as a species of consistency hypothesis, of which dissonance theory is the most prominent previous example (Brown, 1965, chapter 11). A need for consistency of cause and effect clearly implies that the need for a big cause should be greater, the greater the effect to be explained. Thus the Bethell hypothesis predicts that the perceived probability of a conspiracy should be much higher when the President is shot at and killed than when he is shot at but missed. Study 1 was designed to test this prediction.

Study 1

Method

Subjects. Subjects were 20 undergraduate students of Bryn Mawr and Haverford Colleges, 10 female and 10 male.

Questionnaire. The questionnaire consisted of two pages. On the top of each page was the following introduction: "News reports of violent events and their causes are sometimes surprising and sometimes not. This study aims to measure your personal feeling about the likelihood of several stories of several events." Below the introduction was typed a headline in capital letters: "A MAN SHOOTS AT THE PRESIDENT AND MISSES" or "A MAN SHOOTS AT THE PRESIDENT AND KILLS HIM." On each page, after the headline, the same two questions appeared: "What is the probability that this man is acting alone and unaided?" and "What is the probability that this man is acting as a member of a group organized to kill the president?"

Procedure. The order of the two pages was reversed for half the subjects. Two female experimenters each obtained 10 completed questionnaires.

Results

In this and succeeding studies, it is the relative probabilities associated with group and individual explanations of assassination that are of interest. These relative probabilities are naturally expressed as ratios, and we report our data in the form of median ratios. Mean ratios do not properly represent the central tendency of distributions of these ratios, since a few ratios in every distribution are likely to be very extreme values. When most subjects are giving ratios such as 2:1, 5:1, or 1:4, one subject giving a ratio of 100:1 can make the mean ratio totally unrepresentative. Thus, means and parametric statistics are inappropriate with our data, and we use medians and nonparametric statistics instead.

The data of Study 1 did not appear to depend on the order of the pages of the questionnaire or on the experimenter, so the data of all 20 subjects were pooled for analysis.

The first column of Table 1 shows that the median odds for conspiracy were 1:1 when the President had been killed, but were 1:2 when the President had been missed. That is, the likelihood of conspiracy relative to the likelihood of a lone assassin was typically seen as greater when the assassination was

successful. We can use a sign test to test the significance of this tendency. Of the 20 subjects in Study 1, 13 subjects gave odds of conspiracy higher when the President was killed than when he was missed. Six subjects indicated no difference in these odds, and one subject reported the reverse difference in odds. These data indicate ($p < .05$ by one-tailed sign test for correlated samples) that a successful assassination is more likely than a failure to be attributed to a conspiracy.

Discussion

Study 1 supported the prediction that information about success or failure of an assassination attempt makes a big difference in the popularity of a conspiracy explanation. This prediction came from the general hypothesis that people need a big cause to explain a big effect. A special need for consistency in the size of cause and effect is not, however, directly demonstrated by the data of Study 1. For instance, it might well be that the effect of information about success comes rationally from a judgment that groups are more effective and likely to succeed than individuals. In order to demonstrate the hypothetical consistency need, it must be shown that people systematically exaggerate the probability of a conspiracy beyond what the news of a successful assassination rationally calls for. Clearly, this demonstration requires a formulation of the rational impact of information, and Bayes' rule provides just this normative formulation.

In probability form, Bayes' rule requires that

$$p(\text{conspiracy/President killed}) = p(\text{conspiracy}) \cdot \frac{p(\text{President killed/conspiracy})}{p(\text{President killed})}.$$

Note that this formulation calls for revision in the prior probability of a conspiracy to the extent that the ratio reflecting the efficacy of a conspiracy—p (President dead/conspiracy)/p(President dead)—is greater than 1.0. More useful for present purposes is Bayes' rule in odds form:

$$\frac{p(\text{conspiracy/President killed})}{p(\text{individual/President killed})} = \frac{p(\text{conspiracy})}{p(\text{individual})} \cdot \frac{p(\text{President killed/conspiracy})}{p(\text{President killed/individual})}. \quad (1)$$

This form indicates that the odds favoring a conspiracy over an individual assassin, given that the President is killed, need to be revised and increased over the prior odds favoring a conspiracy to the extent that a conspiracy is seen as more likely than an individual to succeed in killing the President. If the posterior odds are found to be systematically higher than called for by the prior odds and the efficacy ratio, then the hypothesis of a consistency need to explain the departure from rational prescription would be strongly supported.

Studies 2 and 3

Method

Subjects. The subjects of Study 2 were six males and six females recruited individually by a female experimenter in the environs of the Bryn Mawr train station. The 12 subjects ranged in (estimated) age from early 20s to late 50s. Subjects of Study 3 were 15 males and 9 females recruited by a different female experimenter in a restaurant and bar near Bryn Mawr College. These subjects appeared to be in their 20s and 30s and were generally approached as same-sex groups (though subjects filled out the questionnaire without discussion with their friends).

Questionnaire. The questionnaire for Studies 2 and 3 was composed of four pages, two of which asked questions much like the two questions used in Study 1. At the top of these two pages was the instruction: "Imagine the following news headline": There followed, on one page, "MAN SHOOTS, KILLS PRESIDENT," and on the other page, "MAN SHOOTS AT PRESIDENT, MISSES." On both pages the succeeding question was the same: "Which is more likely? (A) The man is acting alone and unaided or (B) the man is acting as a member of an organized group." The question continued with a quantification: "If you checked A, how much more likely is A? (Twice as likely as B? Three times as likely? Five times? Ten times?)." The parallel alternative was also given: "If you checked B . . . ," etc.

The prior odds favoring conspiracy were assessed on a third page: "The next person to try to kill the President will likely be . . . (check one)." There fol-

lowed the same A versus B choice and the same quantification of that choice as just described for the first two pages.

The likelihood ratio reflecting the relative efficacy of a conspiracy was assessed on a fourth page, as follows: "Suppose that a man acting alone and unaided is trying to kill the President. Suppose also that a man acting as a member of an organized group or conspiracy is trying to kill the President. Which is more likely to succeed in killing the President?" There followed the same A versus B choice and quantification of that choice as already described.

Procedure. In Study 2, the prior-odds question was always on the last page, and the 6 possible orders of the other three pages appeared twice each to form the 12 questionnaires. In Study 3, all 24 possible orders of the four pages appeared once to make up the 24 questionnaires.

Results

Columns 2 and 3 in Table 1 show that, as in Study 1, the conspiracy explanation was typically more popular when the President was killed than when he was missed (median odds of 2:1 vs. 1:3 for Study 2, and median odds of 2.5:1 vs. 1:3 for Study 3). This result is confirmed ($p < .05$ by one-tailed sign test for correlated samples) by noting that 7 subjects in Study 2 judged conspiracy odds higher when the President was killed than when he was missed, whereas only 1 judged the reverse. For Study 3, the corresponding numbers were 13 and 2, respectively (also $p < .05$ by sign test).

The two additional questions and the odds format of all the questions permit assessment of the degree to which subjects exaggerate the chance of conspiracy. For each subject, the posterior odds of conspiracy given that the President was killed were compared with the product of the prior odds for conspiracy and the efficacy ratio. In Study 2, 5 subjects gave posterior odds of conspiracy higher than required by Bayes' rule, and 6 subjects gave posterior odds too low. In Study 3, 8 subjects gave posterior odds too high, and 13 gave posterior odds too low. Clearly, there is no evidence here of systematic departure from rationality; conspiracy is not overly favored when the President is killed.

Earlier we supposed that higher odds for conspiracy, given successful assassination, might be rational if people believe that groups are more effective than lone assassins. That supposition received some support in Studies 2 and 3, since the median efficacy ratios (third row in Table 1) of 2.5:1 and 2:1 indicate that groups are typically seen as more likely than individuals to succeed in an attempt to kill the President. In Study 2, 9 subjects thought groups were more likely to succeed, 2 subjects thought individuals were more likely to succeed, and 1 subject thought there was no difference ($p < .05$, one-tailed binomial test). In Study 3, 14 subjects thought groups were more effective, and 10 subjects thought the reverse ($p < .27$, one-tailed binomial test, *ns*). Thus, Study 2, but not Study 3, shows significantly more than half the subjects judging groups as more likely than individuals to succeed in an attempt to kill the President.

Study 4

It appears that Studies 2 and 3 have shown that the odds for conspiracy are not exaggerated, compared to the perception of the prior odds of conspiracy and the diagnosticity of the news that the President has been killed. Before accepting this conclusion, however, there is a flaw to be considered in the questionnaire used in Studies 2 and 3. That is, the aim was to ask about prior and posterior odds of conspiracy that differ only in giving the information that the President was killed in the posterior assessment. Unfortunately, the difference between our prior ("The next person to try to kill the President . . .") and our posterior ("Man shoots, kills President") question is two pieces of information: that the attempt got as far as getting a shot off and that the shot was successful. Our efficacy ratio, on the other hand, assessed only the information value of success and not the information value of getting a shot off. In order to be sure that this confounding is not important to our results, we revised the questionnaire and used it in a new study.

Method

Subjects. Subjects were 15 males and 9 females recruited at a shopping center near Bryn Mawr College by the same female experimenter who ran Study 3. The subjects ranged in (estimated) age from early 20s to late 50s.

Table 1
Median Probability Ratio Associated With Conspiracy (Group) Versus Lone Assassin (Individual) Explanations of Presidential Assassination

Probability	Study 1 (n = 20)	Study 2 (n = 12)	Study 3 (n = 24)	Study 4 (n = 24)
$\frac{p(\text{group/President killed})^a}{p(\text{individual/President killed})}$	1:1	2:1	2.5:1	2:1
$\frac{p(\text{group/President missed})^a}{p(\text{individual/President missed})}$	1:2	1:3	1:3	1:2
$\frac{p(\text{President killed/group try})^b}{p(\text{President killed/individual try})}$		2.5:1	2:1	3:1
$\frac{p(\text{group try})}{p(\text{individual try})}$		1:2	2:1	1:1

[a] In each of the four studies, a conspiracy was judged more likely when the President was killed than when the President was missed ($p < .05$ by one-tailed sign test for correlated samples).
[b] In Studies 2 and 4, but not in Study 3, more than half the subjects judged a group more likely to succeed than an individual ($p < .05$ by one-tailed binomial test).

Questionnaire. The questionnaire was identical to that used in Studies 2 and 3 except for the wording of the prior-odds question and the relative efficacy question. Whereas the previous questionnaire asked about "The next person to try to kill the President . . ," the revision asked, "The next person to shoot the President will likely be. . . ." And whereas the previous form asked about the success expected of an individual or group "trying to kill the President," the revision went as follows: "Suppose that a man acting alone and unaided gets a shot at the President. Suppose also that a man acting as a member of an organized group gets a shot at the President. Which is more likely to succeed in shooting and killing the President?"

Procedure. The procedure was the same as in Studies 2 and 3.

Results

Column 4 of Table 1 shows that, as in Studies 1–3, the conspiracy explanation was more favored when the President was killed than when he was missed (median odds of 2:1 vs. 1:2). This result is confirmed by noting that 13 subjects judged conspiracy more likely when the President was killed than when he was missed, whereas only 4 subjects judged the reverse ($p < .05$ by one-tailed sign test for correlated samples).

As in Studies 2 and 3, an analysis at the level of the individual compared the posterior odds of conspiracy with the product of the prior odds for conspiracy and the likelihood ratio giving the relative efficacy of conspiracy. Nine subjects gave posterior odds of conspiracy higher than required by Bayes' rule, and 15 subjects gave posterior odds too low. In short, the results of Study 4 are like those obtained with the flawed questionnaire in Studies 2 and 3. Studies 2–4 are consistent in finding no systematic exaggeration, compared to the Bayesian prescription, of the probability of a conspiracy when the President is killed.

Likewise, the tendency to see groups as more effective than individuals is confirmed in Study 4. The efficacy ratios judged by subjects in Study 4 were for the case of an assassin who had gotten as far as a shot at the President: the relative probability of killing the President for a member of an organized group versus an individual acting alone and unaided. The median efficacy ratio (third row in Table 1) of 3:1 indicates that group members are typically seen as better shots. In Study 4, 19 subjects thought group members were more effective, and 5 subjects thought the reverse ($p < .05$, one-tailed binomial test). Thus, Study 4 shows significantly more than half the subjects judging groups as more likely than individuals to succeed in killing the President, once having gotten off a shot.

"Pure" Data From Studies 1–4

The within-subjects design of the present studies, where each subject answered two

(Study 1) or four (Studies 2–4) questions, raises the possibility of sensitization effects. Perhaps subjects answer a question differently because of biases or hypotheses engendered by having already answered previous questions. In order to be sure that our results do not suffer from sensitization problems, we brought together "pure" data corresponding to the first three rows of Table 1, that is, for the ratios about which we have made claims by statistical test. The "pure" data are from only the first pages of questionnaires, that is, are responses to questions by subjects who had not previously answered other questions. These data cannot therefore be prejudiced by any kind of sensitization effect.

Putting together "pure" data from Studies 1–4, we had 26 judgments of the odds for conspiracy given success (first row in Table 1) and 26 judgments of the odds for conspiracy given failure (second row in Table 1). The median odds for conspiracy when the President was killed were 2:1; they were 1:2 when the President was missed. These two "pure" medians are numerically quite close to the medians for all data in rows 1 and 2 (respectively) of Table 1, indicating that sensitization effects were not a problem for these data. Further, these two "pure" medians are significantly different ($p < .05$ by one-tailed median test) and clearly indicate that in a cross-groups design, we still find that the odds for conspiracy are higher when assassination is successful than when it fails.

Similarly, we put together data from Studies 2–4 to get a total of 15 "pure" judgments of the efficacy ratio (row 3 in Table 1). The median of these 15 efficacy ratios was 2:1, which is numerically close to the medians for all data in row 3 of Table 1. Unfortunately, it is not true that significantly more than half our 15 "pure" subjects saw groups as more effective than individuals (in fact, only 8 of 15 did). Still, we believe that with only 15 judgments, the failure to obtain statistical support cannot be too surprising and that the numerical value of the "pure" median offers some indication that the efficacy ratio data do not suffer from serious sensitization problems.

General Discussion

We began with Tom Bethell's (1975) hypothesis that conspiracy theories of Presidential assassination are popular because of an irrational need for big causes to explain big events. From this hypothesis we predicted that the success or failure of an assassination attempt should make a big difference in the perceived likelihood of conspiracy. Studies 1–4 confirmed this prediction: People are much more likely to entertain a conspiracy explanation when the President is shot and killed than when he is shot at and missed.

This effect is all the more striking given that our manipulation of the information of success and failure was only a few words in an imaginary newspaper headline. We provided none of the richness of detail and description that one might suppose necessary for compounding a conspiracy theory. The cold and abstract nature of our manipulation was intentional, since Bethell's consistency hypothesis depends only on an abstract proportionality of cause and effect, and since we wanted to avoid the possibility of introducing some general cognitive bias in our description. It is clear from our results that the greater popularity of conspiracy explanations when the President is dead is an effect that does not depend on the specifics of a particular assassination story.

Studies 2–4 were aimed at discovering whether the effect of success versus failure is irrational, that is, whether the posterior odds for conspiracy are exaggerated compared to the prior odds for conspiracy and the diagnosticity of the information that the President has been killed. No evidence of such exaggeration was found.

Here it should be noted that the difference between the odds for success when the President is killed and the odds for success when the President is missed—the difference suggested by Bethell's hypothesis and found by us in each of the four studies reported here—might still be irrational in a fashion not tested for in our studies. That is, it might be that the posterior odds for conspiracy are systematically underestimated compared to the prior odds for conspiracy and the diag-

nosticity of the information that the President has been missed.

What we know from all four studies is that

$$\frac{p(\text{group/President killed})}{p(\text{individual/President killed})} > \frac{p(\text{group/President missed})}{p(\text{individual/President missed})}. \quad (2)$$

Bethell's consistency hypothesis suggests that people have an irrational need to see big causes for big effects, which implies an inequality in place of the normative Bayesian equality:

$$\frac{p(\text{group/President killed})}{p(\text{individual/President killed})} > \frac{p(\text{group try})}{p(\text{individual try})} \cdot \frac{p(\text{President killed/group try})}{p(\text{President killed/individual try})}. \quad (3)$$

We tested for this systematic exaggeration and found no evidence of it, that is, no evidence that the left side of Equation 2 is typically too large. But it might yet be true that Equation 2 is irrational to the extent that the right side of it is too small:

$$\frac{p(\text{group/President missed})}{p(\text{individual/President missed})} < \frac{p(\text{group try})}{p(\text{individual try})} \cdot \frac{p(\text{President missed/group try})}{p(\text{President missed/individual try})}. \quad (4)$$

This amounts to the hypothesis that people have an irrational need to see small causes for small effects. Although Equation 4 is also a species of consistency hypothesis, it is different from and less intuitive than Bethell's hypothesis. We did not test for Equation 4, which requires an odds judgment—

$$\frac{p(\text{President missed/group try})}{p(\text{President missed/individual try})}$$

—that we did not ask of our subjects.

Thus, people do associate successful assassination with conspiracy, but not because of an irrational need to find big causes for big effects. Rather, our data indicate that people believe conspiracies are more efficient and effective than lone assassins. Whether conspiracies are in fact more dangerous than lone assassins is unknown, save perhaps to the Federal Bureau of Investigation, but the belief in group efficiency does make sense of the substantial probabilities people attach to conspiracy explanations of Presidential assassination.

The association of successful assassination with conspiracy and the belief in group efficacy make sense in another way. They are an interesting example of support for Kelley's (1967, 1972) analysis of variance model of subjective causality. According to this model, "the more extreme the effect to be attributed, the more likely the attributor is to assume that it entails multiple necessary causes" (Kelley, 1972, p. 6). With the relatively easy assumption that the members of a conspiracy are thought of as multiple necessary causes, it follows from Kelley's model that a conspiracy is seen as more likely when the President is dead—an extreme effect—than when he has been missed.

At this point we must recognize that we have not entirely answered the question with which we began. We set out to understand the popularity of conspiracy explanations of the John F. Kennedy assassination. In one sense we have done that, since the belief that conspiracies are more effective than lone assassinations makes understandable the substantial likelihood of conspiracy that people perceive when the President is killed. But the level of public interest in conspiracy explanations of the Kennedy assassination, as distinct from the level of confidence in these explanations, is yet to be understood. The strength and duration of public interest in conspiracy explanations suggests a strong motive at work, but the Bayesian analysis—a static, cognitive analysis—cannot by itself reveal the motive we seek. If the Bayesian analysis had given us evidence of systematic exaggeration in the probability of conspiracy, we would have inferred the kind of consistency motivation we hypothesized at the beginning of this research. In the absence of this kind of exaggeration, however, we must look elsewhere for the motivation that makes the con-

spiracy explanation of the Kennedy assassination so interesting.

One interesting possibility is that Americans are experiencing a gigantic Zeigarnik effect. Zeigarnik (1927) found that interrupted tasks are remembered better than tasks that are completed. Although this effect dissipated within a day or two for the laboratory tasks Zeigarnik used, stronger effects of the same kind may occur with more involving tasks. It seems likely, for instance, that the same lack of closure Zeigarnik was studying can make a fisherman dwell on the one fish that got away long after he has forgotten the many he boated. Perhaps as long as the official explanation conflicts with a common belief linking successful assassination with conspiracy, Americans will see the John F. Kennedy assassination as the kind of unresolved problem that continues in memory and attention.

References

Anson, R. S. *They've killed the president!* New York: Bantam Books, 1975.

Bethell, Tom. The quote circuit. *Washington Monthly*, December 1975, pp. 34–39.

Brown, R. *Social psychology*. New York: Free Press, 1965.

Epstein, E. J. *Legend: The secret world of Lee Harvey Oswald*. New York: McGraw-Hill, 1978.

Fensterwald, B. *Coincidence or conspiracy?* New York: Zebra Books, 1977.

Gallup, G. 80% see a conspiracy in JFK assassination. In *The Baltimore Sun*, Monday, December 27, 1976, p. 11.

Kelley, H. H. Attribution theory in social psychology. In David Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 15). Lincoln: University of Nebraska Press, 1967.

Kelley, H. H. *Causal schemata and the attribution process*. New York: General Learning Press, 1972.

Zeigarnik, B. Das behalten erledigter and unerleighter handlungen. [The memory of completed and uncompleted actions.)] *Psychologische Forschung*, 1927, *9*, 1–85.

Received June 29, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 5, 645-661

Implicit Theories of Relationship: An Intergenerational Study

Marylyn Rands and George Levinger
University of Massachusetts—Amherst

College students and senior citizens two generations apart in age estimated the probabilities of each of 30 behaviors for each of 14 pair relationships varying in their closeness (casual acquaintances, good friends, close relationship, married partners) and in their sex composition. The younger respondents rated the pair relationships of 22-year-olds today, while the older ones rated those of 22-year-old pairs of 50 years ago. The behavior items referred to a variety of social behaviors pertaining to joint activity, self-disclosure, other-enhancement, other-disparagement, physical contact, and norm regulation. Closeness and sex composition of the relationships as well as content of the behaviors strongly affected the probability estimates. The raters' generation also exerted strong effects; today's pairs, especially good or very close heterosexual friends, were believed to be much more likely to express positive and negative feelings and to have physical contact.

How do social relationships of today differ from those of yesterday? And how do they correspond? In what ways have norms changed from generation to generation? Answers to such questions are difficult to come by.

Historical analyses sometimes provide us with theories or hypotheses about contemporary conduct by comparing it with that of earlier times. For instance, in his historical novel *Angle of Repose,* Wallace Stegner (1971) draws a picture of the intimate relationships of three pairs of characters, each pair separated by two generations. The main character is a 58-year-old historian, Lyman Ward, who writes a book about his grandparents' marriage, contrasting it intermittently with his own and that of his 20-year-old secretary, Shelly. Lyman's two grandparents lived together for over 60 years, until parted by death; each suffered the other's transgressions silently, but their relationship deteriorated over the course of their long marriage. Lyman's own long marriage was recently ended, leaving him feeling helpless and betrayed. His secretary, meanwhile, unsure of her commitment, alternately lives with and hides from her lover.

These three pairs display vast differences in their overt behaviors, their expectations of right and wrong, their comforts and discomforts. Lyman's grandparents display stoicism, politeness, discipline, and inhibition toward each other, their kin, and close friends. Lyman and his wife, too, maintain their relationship publicly in the face of private difficulties, but their failure to resolve their problems ultimately results in the breakup of their marriage. In contrast, Shelly's and her lover's on-and-off affair is marked by self-expression, directness, and unpredictability.

This set of characters conveys the idea that norms of interpersonal conduct have changed dramatically over the past century and that each generation both suffers and benefits from its own constellation of circumstances. The novel's message derives from Stegner's per-

This article is based on a master's thesis submitted to the University of Massachusetts by Marylyn Rands (Lacey) with the guidance of the second author and the support of National Science Foundation Grants GS-33641 and BNS 76-02575. For their assistance, we wish to thank the other members of the master's committee, Icek Ajzen and Arnold Well. We are also indebted to Lawrence O'Brien for his valuable help in implementing the multidimensional analyses and to Andy Anderson, Ted Huston, and Myron Wish for their extensive comments.

Requests for reprints should be sent to Marylyn Rands, Department of Psychology, University of Massachusetts, Amherst, Massachusetts 01003.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3705-0645$00.75

sonal impressions—from his *theory* of changes in interpersonal conduct from yesterday to today.

Are there ways of confirming such theories about past and present interpersonal relationships? The present study was an effort to do so. It was a systematic attempt to study people's "implicit theories" about pair relationships—people's expectations about what behavior will occur in different pairings. Each of us is assumed to hold expectations for appropriate conduct that depend on our relationship with another, whether as a child, sibling, friend, spouse, parent, co-worker, or colleague.

The term "implicit theory of relationship" was first suggested to us by Wolfe (Note 1), who proposed that people share "notions about what characteristics of relationships should co-occur" (p. 20). Research on such notions appeared to be a way of pursuing our own interest in distinguishing among levels of pair relatedness (Levinger, 1974; Levinger & Snoek, 1972). Before ascertaining what sorts of characteristics are expected to co-occur, however, it seemed important first to discover what behaviors people expect to *occur*, as was explored earlier in our research program by Richard Mack (see Levinger, 1974).

Previous Research

Several previous studies have investigated perceptions of interpersonal relationships. DeSoto and Kuethe (1959) examined male college students' subjective probabilities for expecting positive or negative feelings in hypothetical pairs of male acquaintances. Respondents estimated the probability that, under various conditions, A would like B (or trust, feel superior to, dislike, lie to, listen to, or dominate B, and so on). Respondents were found to assign higher probabilities to positive than to negative affective relations. Some A–B relations were perceived as symmetric, others as asymmetric; some as transitive, others as intransitive.

Marwell and Hage (1970) obtained ratings of 100 different dyadic role relationships—such as father–son, best friend–best friend, landlord–tenant, actor–agent—on a set of 16 scales measuring characteristics of interaction (such as its frequency, the effort required, or its voluntariness). A factor analysis of these ratings uncovered three dimensions of pair relationships: intimacy, visibility, and regulation of behavior.

A third study located cross-cultural generalities in close relationships (Triandis, Vassiliou, & Nassiakou, 1968). American and Greek respondents rated the appropriateness of 120 behaviors (e.g., show concern for, disagree with, kiss, help, argue with other) for 100 dyadic roles. Four factors common to both cultures emerged: affect, intimacy, dominance, and hostility.

In a fourth study, Wish, Deutsch, and Kaplan (1976) searched for fundamental dimensions in people's perceptions of interpersonal relationships by asking subjects to rate 25 "typical" role relationships and 20 of their own interpersonal relationships on scales pertaining to a large variety of traits (such as cooperative–competitive, emotional–intellectual, flexible–rigid). Their multidimensional scaling analysis revealed four major dimensions of relationships, which can be interpreted as mutual cooperativeness, equality, intensity, and informality.

All these studies found aspects of closeness, positivity, and social influence to be important for characterizing variations among social relationships. Only one of these studies (Triandis et al., 1968), however, examined interpersonal *behaviors*, and it did not select those behaviors from any defined population of actions. Nor did any of these studies have an explicit schema for selecting its sample of relationships; the findings did not permit one to draw conclusions about the effects of systematic variations in relationship.

The Present Research

The present work is concerned with the effects of interpersonal closeness on people's perceptions of appropriate behavior in pairs. Following from Levinger and Snoek's (1972) suggestion of a continuum of pair relatedness, Richard Mack had earlier obtained pertinent descriptive data (see Levinger, 1974, pp. 109–111). Mack's study focused on the behaviors of heterosexual couples of four different degrees of intimacy, from "casually acquainted"

to "much in love and fully committed." His informants considered 52 different possible behaviors—from "smile at the other" to "refuse to date other persons"—and rated how likely each behavior was in each of the four sorts of relationships. For most of the behaviors, likelihood of pair interaction was perceived to vary directly with intimacy (see Levinger, 1974, p. 110).

In the present study, it was decided to vary not only the two partners' closeness but also the pair's sex composition, so as to study perceptions of a systematically selected set of pair relationships. Such relationships were to be rated by raters of both sexes who were drawn from two different age groups. A set of guiding hypotheses and questions is listed below.

Relationships: Degree of closeness. It was hypothesized that expectations of interaction would increase as the perceived closeness of a pair increased. According to current conceptions of relationships (Altman & Taylor, 1973; Levinger & Snoek, 1972), the breadth and the intensity of mutual behavior grows along with increasing interpersonal involvement.

Furthermore, respondent characteristics were assumed to affect such expectations. In line with earlier findings in our research program (Levinger, Rands, & Talaber, Note 2), we hypothesized no differences between the judgments by raters of different sex, but we did look for significant age differences. Raters from today's college generation were hypothesized to expect more pair interaction than raters from their grandparents' generation rating young pairs of 50 years ago.

Relationships: Sex composition. We hypothesized that different behaviors would be expected in same-sex pairs than in cross-sex ones, as well as in male–male as compared to female–female pairs. For example, cross-sex pairs would be perceived as more likely to have physical contact than would same-sex pairs, and female–female pairs as more likely to disclose intimacies than would male–male dyads.

Furthermore, it was hypothesized that younger raters would differentiate less between same-sex and cross-sex relationships than would older raters. This hypothesis was based on the assumption that today's sex role stereotypes are less distinct than those of 50 years ago.

Considering the sex of the actor, we hypothesized that raters would judge female partners to be more likely than males to initiate expressive actions, and males to be likely to initiate more actions pertaining to the management of overt activities. Such differences, it was thought, would be greater for older than for younger respondents.

Questions were also posed concerning the perceived likelihood of various sorts of behaviors that might occur in close pairs. We hypothesized that activity-oriented behaviors would be perceived as less affected by the partners' closeness or sex composition than would those expressing positive or negative feelings and that expectations of physical contact would be most affected by variations in closeness and sex composition.

Method

Participants in the study were of two age groups and of two sexes. Each person rated the likelihood of 30 different behaviors for each of 14 relationships. A sample of young, college-age raters judged "typical relationships" between 22-year-old partners today. The age of 22 was chosen to be near enough to that of the college-age respondents while old enough for people to be stably married. A sample of older raters, generally in their 70s, described relationships of 22-year-olds as they recalled them from the time when they themselves were in their early 20s.

Raters

The group of young raters consisted of 40 University of Massachusetts students, half male and half female. They came to group sessions in response to either an announcement in an introductory sociology class or an advertisement posted in the psychology building. Thirty-four raters were paid $1.50 each for filling out the questionnaire, and six chose to receive experimental credit from a psychology class. They ranged in age from 18 to 27 years, with a mean age of 19.9 years.

The group of older raters were 40 senior citizens, also half male and half female, who lived in a retirement community near Philadelphia. The community's "Reserve Fund" received $1.50 for each rater's participation. These respondents ranged in age from 65 to 87 years, with a mean age of 75.7 years. Respondents in both age groups were all Cau-

casian and predominantly from middle- to upper-middle-class background.

Relationships

Each person rated 14 relationships representing four levels of closeness (casual acquaintances, good friends, close relationship, and married pairs). At each degree of closeness except marriage, the two partners could be either both male, both female, or one male and one female. In the heterosexual pairs, the sex of the actor was also varied. Thus each rater considered eight heterosexual and six same-sex relationships.

The questionnaire first introduced the rater to four degrees of relatedness. Each pair relationship was pictured as two circles, varying in their degree of "intersection" (Levinger & Snoek, 1972) from very small to substantial. They were described as follows: (a) *casual acquaintances*—two people who feel friendly toward each other; the partners have many other relationships like this one; (b) *good friends*—two people who care about each other; each person has several other relationships like this one; (c) *close relationship*—two people who care about each other very much; the partners have no other relationship like this one; and (d) *married partners*—two people who care about each other very much; the partners have no other relationship like this one; they have made their relationship permanent.

Each rater received a booklet with the 14 relationships placed in a different random order. For each pairing, the relationship was described and the sex of each partner indicated, with the first person in each pair designated as the actor.

Behaviors

The 30 items used in this study (see Table 1) were selected from an initial list of 50 items to represent overt actions in the categories of joint activity, self-disclosure, other-enhancement, other-disparagement, norm regulation, and physical contact (positive and negative). The items chosen represented both physical activity and emotional expression, both verbal and nonverbal acts, and both positive and negative behaviors. A preliminary study showed that these items discriminated among the four degrees of closeness and covered a range of activities appropriate for most social relationships.

Choices of behaviors were guided by several criteria: (a) The behaviors were selected to represent specific acts, not abstractions such as many of the verbs in Osgood's (1970) study, nor behavioral intentions. (b) Items were to represent initiatory actions, since reactive behaviors would depend on their context. (c) Behaviors were to be applicable to all possible relationships; except for "making love," none of the items appeared likely to seem improper for same-sex pairs but proper for cross-sex ones.

Each behavior was rated on a 10-point scale ranging from 0 (not at all likely—i.e., would occur in 0%–10% of all relationships of the kind rated) to 9 (extremely likely—i.e., would occur in 90%–100% of all relationships of the kind rated).

The 30 items were presented 10 to a page, with 3 pages per relationship. The order of the three pages was randomly varied across the 14 pairings, so that no two respondents made their ratings in exactly the same order.[1]

A note on the rater's time perspective. While the young raters judged relationships among today's 22-year-olds, their contemporaries, older raters were asked to recall behavior norms of about 50 years ago. Although judgments of their own contemporaries seemed more meaningful than if senior citizens were to express norms about young pairs today, this procedure introduced a difficulty for interpreting the data: The older raters' judgments depended on their recollections. This difference will be discussed further at the end of the article.

Results

Our findings will be presented as follows: First we will consider the implicit dimensions of closeness that seem to underlie the respondents' ratings; the relationships judged by each group of raters will be plotted into a two-dimensional space. Next we will look at mean expected probabilities for the average social behavior across relationships and across generations. Following that, we will examine variations in different kinds of behaviors, using the results of a five-factor analysis of variance. Finally, we will focus intensively on intergenerational differences in expectations for appropriate interaction in close pairs.

Dimensions for Distinguishing Among the Relationships

In order to compare the 14 relationships simultaneously across several dimensions, a multidimensional scaling analysis was performed on the ratings of each age group. Individual Differences Scaling (INDSCAL), developed by Carroll and Chang (1970), was

[1] A preliminary study, using 50 behavior items, was carried out with two other groups of 40 young and 40 old respondents (Lacey, 1977). Because of a design error, its results could not be submitted to multidimensional scaling analyses. The methods of that study otherwise parallel the present ones, and its results will be referred to wherever they are relevant to the present findings.

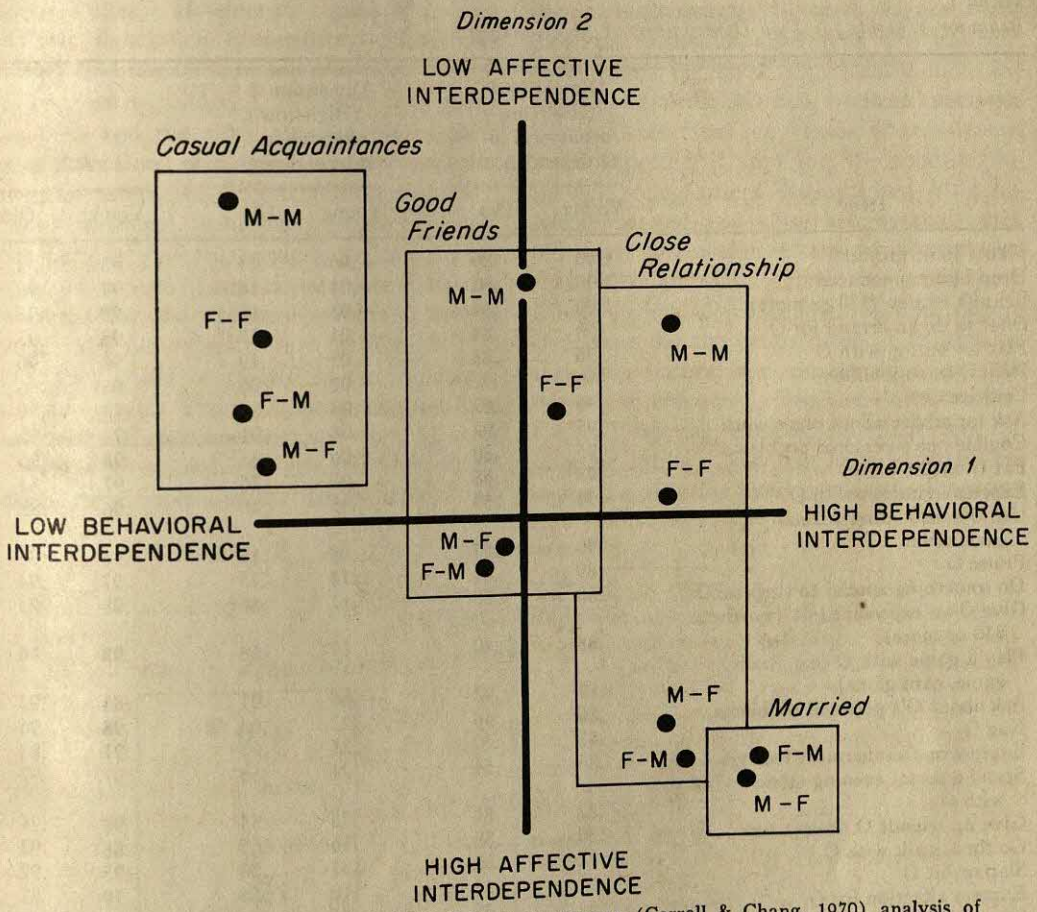


Figure 1. Relationship dimensions derived from an INDSCAL (Carroll & Chang, 1970) analysis of the younger respondents' behavior ratings. (For each pair, the first letter denotes the actor, the second the target of the action. M = male; F = female.)

used; the statistical program allows simultaneous analyses of all 14 relationships and provides unique solutions without any rotation of axes.[2] The INDSCAL analyses produced a relationship space for each group of raters with several dimensions and weights on each dimension for each behavior, to reflect the importance of each item for contributing to that dimension.[3] Distances between points in the space indicated the perceived dissimilarity among relationships. The analyses showed that a one-dimensional solution accounted for 81% and 74% of the variance for the young and the old raters, respectively; a two-dimensional solution accounted for 89% and 83%, respectively; and a three-dimensional solution accounted for 91% and 86% of the variance for the young and old raters, respectively. Since the further increment from two to three dimensions had little meaning, the two-dimensional solution was used.

Figure 1 shows the 14 relationships judged

[2] Input for the INDSCAL analyses was the dissimilarity matrix for each relationship, derived from the following formula:

$$\delta_{jk(s)} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_{ijs} - x_{iks})^2},$$

where δ_{jk} is the dissimilarity between relationships; x_{ijs} and x_{iks} are Subject i's ratings of Relationships j and k on Behavior s; and N is the number of subjects who rated the relationships on Scale s (see also Wish, Deutsch, & Kaplan, 1976, p. 411).

[3] These dimension weights are analogous to partial correlations, whereas the R values (see last column of Table 1) are analogous to multiple correlations.

Table 1
Behavior Weights for Two Dimensions of Relationships

Item[a]	Dimension 1 (High–low behavioral interdependence)		Dimension 2 (High–low affective interdependence)		*R*	
	Young	Old	Young	Old	Young	Old
Plan a joint project	1.00	.89	−.06	.15	.98	.98
Drop by unannounced	1.00	.86	−.04	.14	.99	.95
Loan O money ($10 or more)	1.00	.84	−.03	.17	.98	.95
Offer to do an errand for O	.98	.84	−.01	.19	.98	.95
Plan an outing with O	.96	.84	.05	.19	.98	.96
Help O learn a hobby	.96	.92	−.01	.06	.95	.96
Criticize O	.96	.67	.03	.34	.97	.90
Ask for advice about one's work	.95	.92	−.04	.01	.93	.93
Confide one's personal problems	.95	.80	.06	.25	.98	.96
Pat O on the back	.95	.68	.04	.25	.97	.84
Express irritation with O	.94	.59	.05	.40	.96	.88
Use O's belongings without permission	.90	.81	.08	.19	.95	.83
Praise O	.89	.61	.14	.45	.97	.94
Do something special to surprise O	.88	.50	.17	.56	.98	.93
Give O an expensive gift (worth $10 or more)	.88	.40	.17	.68	.98	.96
Play a game with O (e.g., ball game, card game)	.87	.92	−.08	.01	.84	.92
Ask about O's personal problems	.86	.46	.22	.63	.98	.96
Nag O	.85	.30	.21	.63	.97	.83
Express one's innermost feelings	.83	.54	.24	.56	.97	.97
Spend a social evening alone with O	.82	.51	.28	.55	.98	.94
Give up friends O doesn't like	.80	.38	.16	.65	.88	.93
Go for a walk with O	.72	.80	.37	.20	.95	.92
Slap or hit O	.64	.35	.10	.56	.70	.81
Express affection for O	.64	.35	.50	.74	.98	.97
Stand close to O (within 1 foot)	.61	.16	.52	.88	.97	.98
Hug O	.53	.07	.57	.89	.95	.93
Avoid touching O	.24	.34	.79	.62	.94	.86
Compete with O	.19	.24	.58	.17	.69	.37
Make love (have intimate sexual contact)	.10	−.08	.87	.88	.92	.83
Hold hands	.01	−.26	.96	1.00	.96	.93
Percentage of variance for two-dimensional solution					89	83

Note. O = the other.
[a] Items are listed according to decreasing weights for the young raters on Dimension 1.

by the young raters. They are arrayed on two dimensions. Dimension 1, accounting for the largest amount of variance, can be interpreted as a pair's degree of behavioral interdependence. Most of the 30 behaviors had high weights on this dimension (e.g., planning a joint project, helping the other learn a hobby, or asking for advice about work; see Table 1). Casual pairs were rated lowest on behavioral interdependence and married pairs highest, with good friends and close relationships intermediate.

Dimension 2 refers to the pair's degree of affective interdependence. This dimension is characterized by physical contact—by behaviors such as making love, hugging, holding hands, or showing affection (and, for the older raters, by self-disclosure to and criticism

of the other). Heterosexual pairs were considered highest on Dimension 2, male pairs lowest, and female pairs intermediate.

The correspondence between the young sample's and the older sample's use of these two dimensions was estimated by correlating their respective coordinate positions in their two-dimensional solutions. Product-moment correlations were extremely high for both Dimension 1 (.969) and Dimension 2 (.936). Accordingly, the separate solutions for the young and old raters were rotated slightly so that they could be plotted in the same relationship space. Figure 2 shows the resulting intergenerational comparison.

While behavioral and affective interdependence were important for describing the ratings of both samples of raters, Figure 2 also indicates important differences. For example, older raters perceived significantly more behavioral interdependence between same-sex than between cross-sex friends, whereas young raters expected, if anything, the opposite. Although young raters distinguished little between married and close heterosexual pairs, old raters distinguished considerably between these intimate pairs; they also rated good friends as less interdependent on both dimensions than did the young raters. Furthermore, the older raters saw relationships at each degree of contact as less affectively differentiated by sex composition than did the younger raters. This may have occurred because the affective interdependence dimension

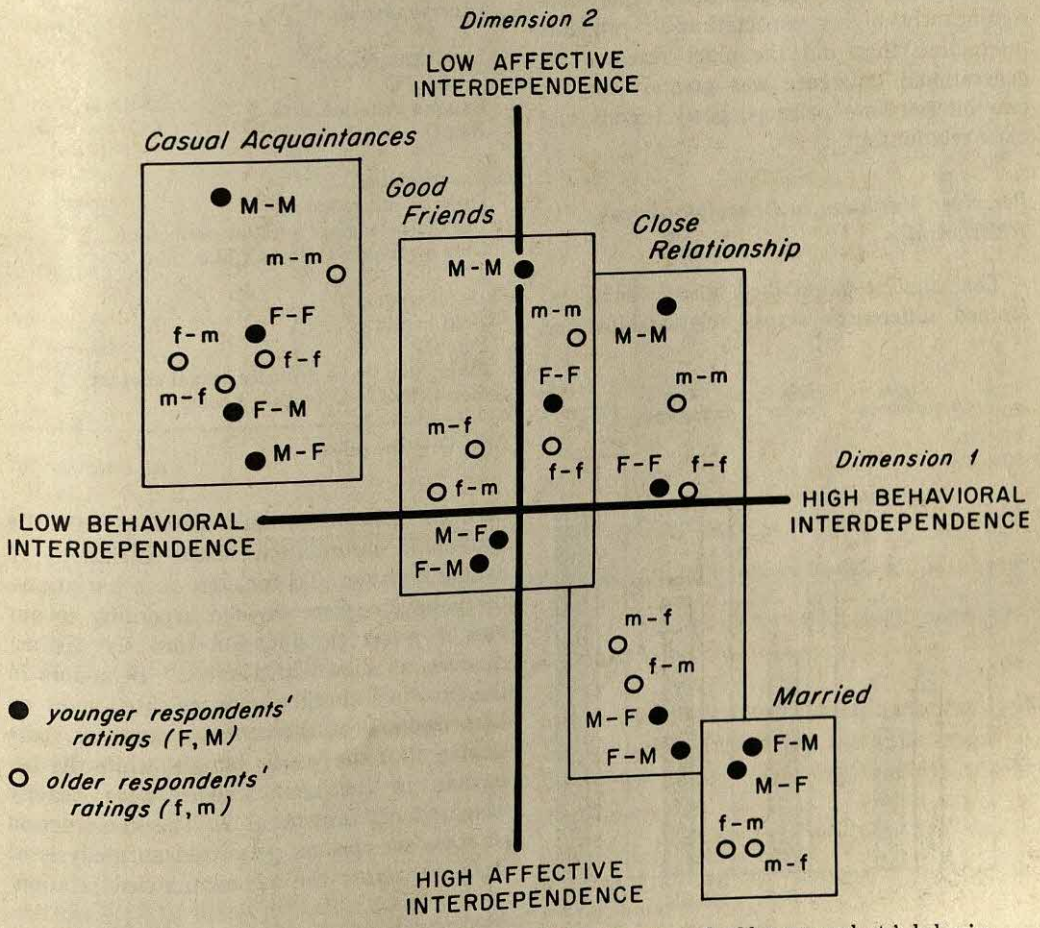


Figure 2. Relationship dimensions derived from both younger and older respondents' behavior ratings. (For each pair, the first letter denotes the actor, the second the target of the action. M, m = male; F, f = female.)

was largely based on physical contact behaviors; the older participants rated such contact quite improbable across all unmarried relationships, whereas the younger ones' ratings varied from low to very high.

Behavior Probabilities Across Relationships and Generations

The multidimensional analysis does not in itself reveal absolute percentage differences between the younger and the older samples, or among the relationships. Figure 3 shows such mean differences. It shows the extent to which probability of interaction was perceived to increase with relationship closeness. At each degree of closeness, and for almost all behaviors, the younger raters reported a significantly higher expectation of probable interaction than did the older raters. This generational difference was greatest for the two intermediate pairings, good friends and close relationship.

Perceived Variation in Behaviors Across Relationships

The analyses described above have examined differences across relationships and between generations but have given little systematic information about variations in dyadic behavior. To consider such variations, the behaviors were divided according to our own *a priori* classification into six distinct clusters, as shown in Table 2.[4] Behaviors in the first five clusters were weighted most on Dimension 1 of the INDSCAL analysis, particularly for the young sample, while the behaviors in the sixth cluster were heavily weighted on Dimension 2. The construction of these six clusters permitted an analysis of variance across the 12 nonmarried relation-

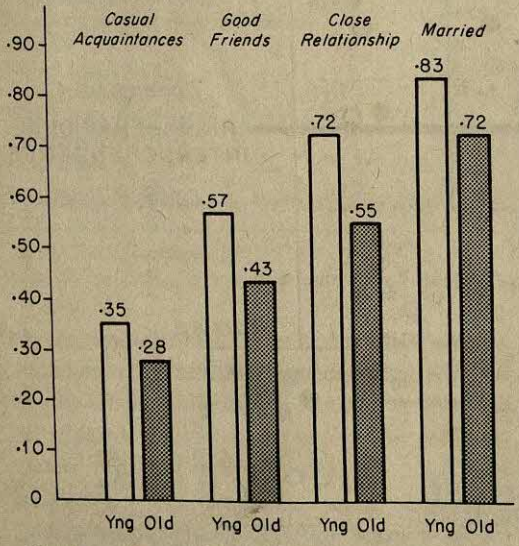


Figure 3. Mean probabilities of interaction across four degrees of closeness as estimated by both younger and older raters.

Table 2
Six Behavior Clusters Used in the Analysis of Variance

Social contact
Play game with O (e.g., ball game, card game)
Ask for advice about one's work
Help O learn a hobby
Offer to do an errand for O
Plan an outing with O
Go for a walk with O
Self-disclosure
Confide one's personal problems
Ask about O's personal problems
Express one's innermost feelings
Other-enhancement
Praise O
Pat O on back
Do something special to surprise O
Express affection for O
Other-disparagement
Criticize O
Express irritation with O
Nag O
Norm regulation
Drop by unannounced
Use O's belongings without permission
Give up friends O doesn't like
Physical contact
Hold hands
Hug O
Make love (have intimate sexual contact)
Stand close to O (within 1 foot)

Note. O = the other.

[4] Seven items were omitted from this analysis of variance across categories because they did not clearly fit into any of the six behavior categories.

Table 3
Mean Probabilities Across 14 Relationships and 6 Clusters of Behavior

Relationship	Joint activity	Self-disclosure	Other-enhancement	Other-disparagement	Norm regulation	Physical contact	Total
Casual acquaintances							
Same sex							
M–M	.48	.17	.20	.24	.17	.13	.25
F–F	.51	.25	.37	.24	.19	.21	.29
Cross-sex							
M–F	.50	.23	.40	.23	.16	.31	.31
F–M	.46	.22	.33	.22	.15	.26	.27
Total	.49	.22	.35	.23	.17	.23	.28[a]
Good friends							
Same sex							
M–M	.73	.42	.53	.37	.40	.23	.45
F–F	.73	.55	.61	.39	.41	.36	.51
Cross-sex							
M–F	.70	.44	.59	.35	.33	.48	.48
F–M	.69	.45	.56	.35	.31	.48	.48
Total	.71	.47	.58	.37	.36	.39	.48
Close relationship							
Same sex							
M–M	.81	.61	.66	.48	.55	.33	.57
F–F	.81	.70	.73	.47	.55	.48	.62
Cross-sex							
M–F	.80	.68	.77	.47	.53	.75	.67
F–M	.80	.72	.76	.48	.50	.72	.66
Total	.81	.67	.73	.48	.53	.57	.63
Marriage[b]							
Cross-sex							
M–F	.86	.83	.86	.61	.68	.90	.79
F–M	.88	.84	.87	.61	.68	.90	.80
Total	.87	.83	.86	.61	.68	.90	.79
All relationships							
Total	.72	.55	.63	.42	.44	.52	.55

Note. For each pair, the first letter denotes the actor; the second, the target of the action. M = male; F = female.

[a] Any discrepancy between the probabilities reported in this column and in Figure 1 is due to the different number of items on which the means are based.

[b] Probabilities for the married partners are shown here, even though the married relationship was not included in the analysis of variance.

ships. The one between-group factor was the raters' age (young or old);[5] the four within-subjects factors were the closeness of the relationship (casual acquaintances, good friends, close relationship), sex composition (same sex, cross-sex), sex of actor (male, female), and the type of behavior (joint activity, self-disclosure, other-enhancement, other-disparagement, norm regulation, and positive physical contact). Each of these variables exerted a statistically significant effect on the ratings.[6] Table 3 depicts the mean perceived probabilities across all four within-subjects variables;

[5] Sex of rater was omitted from this analysis, since previous analyses had shown no significant main effect ($p > .10$) nor any significant interaction effects of the raters' sex.

[6] Parallel findings were obtained for the data collected in the preliminary study, which also had included many additional behaviors.

Table 4
Mean Probabilities for Six Behavior Clusters Across Three Levels of Closeness

Relationship	Joint activity	Self-disclosure	Other-enhancement	Other-disparagement	Norm regulation	Physical contact	Total
Casual acquaintances							
Same sex	.49	.21	.33	.24	.18	.17	.27
Cross-sex	.49	.23	.37	.23	.16	.29	.29
Good friends							
Same sex	.73	.49	.57	.38	.40	.30	.48
Cross-sex	.69	.45	.58	.35	.32	.48	.48
Close relationship							
Same sex	.81	.65	.70	.48	.55	.41	.60
Cross-sex	.80	.69	.76	.47	.51	.73	.66

for the sake of simplicity, it disregards generational differences, which will be considered later.

Effects of closeness. As anticipated, interaction was expected to increase directly with the increasing closeness of the partners; mean probabilities were 28% for casual acquaintances, 48% for good friends, and 63% for the close relationship. These differences were highly significant statistically, $F(2, 156) = 463.62$, $p < .0001$.

The six clusters of behavior items differed significantly, $F(5, 390) = 229.64$, $p < .001$. For some clusters, increases in likelihood across degrees of relationship closeness were steep and others were gradual; some began from a low baseline and others from a moderate baseline (see Table 3). Joint activity and other-enhancement were perceived to be moderately probable (49% and 35%, respectively) among casual acquaintances and to rise high (87% and 86%, respectively) for the married pair; most other behaviors were considered to be improbable between casual acquaintances (less than 25%). Self-disclosure and physical contact were perceived to rise steeply (to 83% and 90%, respectively) as closeness increased. The behaviors subsumed under norm regulation and other-disparagement leveled off at a lower ceiling (68% and 61%, respectively).

Sex composition. The sex composition of a relationship was also significantly associated with the expected probability of behavior, $F(1, 78) = 23.76$, $p < .0001$. Overall, cross-sex pairs were expected to interact somewhat more than same-sex pairs (see Table 3). That effect was conditioned, however, both by the degree of closeness (there was no such difference for good friends) and by the sex of the same-sex pairs (female–female pairs were believed more likely than male–male pairs to initiate interaction across almost all sorts of behavior). Further, while cross-sex means tended to be about the same regardless of who was the actor, in same-sex dyads females were expected to interact more than were males (see row totals in Table 3).

How did same-sex and cross-sex pairs differ in regard to particular behavior clusters? Considering first the rather neutral behaviors classed under "joint activity," interaction was expected to be somewhat greater in same-sex pairs—for good friends, significantly so (see Table 4). Regarding self-disclosure, same-sex and cross-sex means were overall about equal, but same-sex good friends and cross-sex close ones were believed to disclose the most. Other-enhancement was believed to be significantly higher between cross-sex partners, but this difference was entirely due to the low means for male–male dyads. Other-disparagement, in contrast, was expected to be higher between same-sex than cross-sex partners, primarily for good friends. The cluster of norm regulation behaviors (i.e., dropping by unannounced, borrowing belongings and giving up friends the other does not like) was believed more characteristic of same-sex than of cross-sex interaction, but mainly for the good

friends. Finally, as hypothesized, the largest same- versus cross-sex difference, at every level of closeness, occurred for behaviors associated with physical contact.

Other notable sex differences may be seen in comparing the means for male and female same-sex dyads in Table 5. Overall, interaction was estimated as higher in the female pairs at each level of closeness; but these differences were due almost entirely to three of the six behavior clusters: to greater self-disclosure, other-enhancement, and physical affection between females. There were insignificant differences between male and female pairs with regard to expected joint activity, other-disparagement, or norm regulation.

Sex of actor. The analysis of variance located a significant effect of the actor's sex, $F(1, 78) = 14.19$, $p < .0001$, but only in the heterosexual pairs is it meaningful to consider male versus female actor differences. Turning to the data on cross-sex pairs in Table 5, one can examine differences in behavioral interaction attributable to the sex of the actor. In general, the male was expected to initiate more physical activity, ranging from working to playing games and making love, whereas the female partner was believed more likely to initiate verbal activity, such as giving praise or disclosing feelings.

Such male–female differences appeared to decrease with increasing pair closeness. The Total column of Table 3 shows that in the casual acquaintanceship, the male partner was believed more likely than the female partner to initiate all behaviors, particularly other-enhancement and physical contact. When the two persons are good friends or closer, however, the difference becomes negligible and is actually reversed in the marriage relationship. In other words, our respondents perceived it as more appropriate for male than for female members of casual relationships to initiate interaction, but in more intimate relationships, initiating an interchange was seen as equally appropriate for either sex.

Generational Differences

As shown above, the raters' generation strongly influenced their expectations of appropriate behavior in various relationships, $F(1, 78) = 45.40$, $p < .0001$. Altogether, older raters expected less interaction between partners than did young raters (44% vs. 58%). We now consider additional differences between groups both for relationships and for behaviors. We also explore within-group variations in behavioral expectation.

Relationship differences. Although older raters expected less interaction than did young raters, this difference was largest for the pairs intermediate in closeness (i.e., for the good friends and close relationship; see Figure 3). To assess these differences between the young and the old sample's mean ratings, a set of 420 significance tests was computed—one for each of the 14 relationships, for each of the 30 behaviors. About 35% of these comparisons were significant (by t test) at the .003 level or higher—the appropriate significance level for the conservative Bonferroni

Table 5
Mean Probabilities for 6 Behavior Clusters for 12 Relationships Differing in Sex Composition

Relationship	Joint activity	Self-disclosure	Other-enhancement	Other-disparagement	Norm regulation	Physical contact	Total
Same sex							
M–M	.67	.40	.50	.37	.37	.23	.42
F–F	.68	.50	.57	.37	.38	.35	.48
Cross-sex							
M–F	.67	.45	.58	.35	.34	.51	.48
F–M	.65	.46	.55	.35	.32	.48	.47

Note. For each pair, the first letter denotes the actor; the second, the target of the action. M = male; F = female.

Table 6
Mean Probabilities for Six Behavior Clusters Estimated by Young and Old Raters

Group	Joint activity	Self-disclosure	Other-enhancement	Other-disparagement	Norm regulation	Physical contact	Total
Young	.73	.58	.65	.50	.46	.54	.58
Old	.66	.44	.54	.29	.34	.39	.44
Total *M*	.70	.51	.60	.39	.40	.47	.51
Young–Old difference	.07	.12	.11	.21	.12	.15	.14

t test comparison. These differences between the young and the old raters' means were significant for only 15% of the casual acquaintances and 27% of the married pair ratings. In contrast, the young–old mean differences were significant for 41% of the good friends and for 55% of the close relationship ratings.

Generational differences were more noticeable for ratings of cross-sex than of same-sex relationships (40% vs. 29% of the comparisons), and these differences were significant most frequently for female-initiated interaction in the female–male pair: 49% of these contrasts were significant compared to 32% for the male–female, 30% for the female–female, and 28% for the male–male pairs.

Behavior differences. Although the young raters consistently perceived more interaction across the 14 relationships than did the old raters, these differences varied according to the type of behavior. Table 6 shows the mean probabilities for the six behavior clusters. (The following discussion utilizes other data for individual items and relationships that are not presented here.)

Intergenerational differences were least evident in expectations about activity-centered interaction; the data indicate little change over the years, except for today's females being thought more likely than their counterparts of 50 years ago to initiate interaction with their close male partner.

The largest difference occurred in the older generation's perceptions of greater inhibition in regard to criticizing another; today's raters saw as more permissible the disparagement of the partner, including such actions as nagging, criticizing, and expressing irritation.

Generational differences were also quite large for disclosing feelings to another. For each relationship other than the casual acquaintanceship, today's 22-year-olds were perceived as significantly more likely than those of 50 years ago to express their feelings, to ask about a partner's personal problems, and to confide their own personal problems.

Other-enhancement also was perceived as more likely between today's than yesterday's 22-year-olds. Particularly in the close relationship, friends today were perceived as more likely to praise the other and to express affection; especially, it was indicated that females are freer to acknowledge affection toward males than they were two generations ago.

Physical contact was another noticeable area of generational difference. Heterosexual partners today were believed far more likely to pat one another on the back, to hug, or to make love; and such interactions today were considered nearly as likely to be initiated by females as by males. Such differences were extended to same-sex pairings; young raters believed it significantly more likely that both female and male pairs would express some physical affection for each other. (Another finding, concerning negative physical contact behaviors not included in the six major clusters—for instance, slapping or hitting a partner—was that they, too, were perceived as more likely in today's relationships.) The expression of all sorts of feelings appears more permissible today than yesterday.

Within-group age variations. After finding such pervasive differences between the young and the old group of raters, it became important to ask whether similar trends might be detected *within* a single age group. If younger raters *within* an age group were

consistently found to expect more interaction than older raters in that *same* group, such a finding would strengthen the validity of the between-groups comparisons.

The ages of the college raters were very homogeneous, with a range of only 8 years, but among the senior citizens the age range was 22 years (from 65 to 87). Therefore we expected to find no meaningful correlations within the young group but did hope to find negative Age × Probability correlations within the old sample. Within each age group of raters, and for each of the 14 relationships, we computed correlations between the raters' ages and their probability ratings on a subsample of six behaviors that had earlier shown significant between-groups differences.[7]

For the young group, as was expected, no overall trend emerged. For the older age group, 95% of the correlations were negative; 21% of them were significant at the 5% level. In other words, 65-year-old raters describing relationships of 45 years ago tended to recall a higher probability of interaction than did 75- or 85-year-old raters describing pairs of 55 or 65 years ago. This correlational finding gives strong confirmation to the age effects found in the between-groups comparisons.[8]

Discussion

It was found that variations in the closeness and sex composition of pair relationships exerted strong effects on normative expectations of interaction across a wide variety of dyadic behaviors. Those findings applied both to college students judging interaction between today's 22-year-olds and to senior citizens judging interaction between 22-year-olds when they themselves were young, although there also were important intergenerational differences. There were no significant effects of the judges' own sex in either of the two samples. After reviewing such effects and discussing their implications, we will assess the usefulness of this method for comparative research on close relationships.

Effects of Closeness

As a relationship moves from casual to close, its interaction probabilities were perceived to increase correspondingly. Behaviors pertaining to touching or disclosing oneself to the partner were believed to be most affected by increases in closeness, whereas less intimate joint activities, such as planning a joint project or doing an errand for the other, were believed to be least affected. These findings confirm earlier theoretical proposals about the meaning of a growing person–other "intersection" (Levinger, 1974; Levinger & Snoek, 1972). With the exception of fairly neutral task-oriented activity or of strongly affective sexual activity, the closeness of a relationship exerted an increasing, monotonic effect on perceived probabilities of interaction. Nonetheless, the relationships of good friends were perceived to be much more similar to those of close friends than to those of casual friends, as was indicated by their location on Dimension 1 of Figures 1 and 2 and by the means shown in Figure 3. This finding parallels those obtained in Mack's study of perceived behaviors in heterosexual couples (cited in Levinger, 1974) and in Stambul's (1975) retrospective investigation of stages in the courtships of premarital pairs; Stambul found that the stage of serious dating was much more similar to the stage of engagement than to that of casual dating.

The two age groups were largely parallel in these effects but also showed some differences. At each level of closeness, older respondents rated the interaction probabilities significantly lower than did the younger respondents. Greater young–old contrasts occurred for the relationships intermediate in closeness (good friends and close relationship) than for those at the extremes of casualness (casual acquaintances) and closeness (married); such contrasts were noted particularly in the older respondents' lower rat-

---

[7] For this analysis we selected a sample of six behaviors from different clusters that had shown larger intergenerational differences: stand close to the other, ask about the other's personal problems, hug the other, criticize the other, confide one's personal problems, and express irritation with the other.

[8] In the preliminary study, too, the older sample of raters had shown significant negative correlations between raters' ages and the behavior probabilities they estimated for the pairs of friends.

ings of physical contact and other-disparagement.

Effects of Sex Composition

Norms for cross-sex pairs were seen to differ from those for same-sex pairs, but those for female and male same-sex pairs were perceived as even more dissimilar. The sex of the actor, too, had an effect on ratings of behavioral expectations.

Cross-sex versus same-sex differences. The largest differences between cross-sex and same-sex pairs occurred for the close relationship. At this level of closeness, cross-sex pairs were perceived as more likely than same-sex pairs to engage in self-disclosure, other-enhancement, and physical contact. In contrast, for good friends, same-sex pairs were seen as more likely to interact than were cross-sex ones (except for having physical contact).

Male–male versus female–female differences. The effect of sex composition is revealed further in the ratings of interaction in same-sex pairs. Female pairs were expected to engage in greater self-disclosure, other-enhancement, and physical contact than were male pairs. These expectations parallel findings from other studies. Previous research has found that female friends do indeed disclose more intimate details about their lives than do male friends (Jourard, 1971), report greater emotional involvement in their friendships (Rubin, 1973; Purdy, Note 3), have longer interactions (Wheeler & Nezlek, 1977), and do things together more spontaneously (Booth, 1972). Weiss and Lowenthal (1975) found that males emphasize sharing activities and interests more than do females, whereas females emphasize the importance of warm supportiveness.

Pleck (1975) has suggested that women's friendships are traditionally intimate, whereas men's friendships are more sociable than intimate—a distinction he derived from an analogy to Lewin's (1948) cross-cultural analysis of German versus American relationships. Our data support Pleck's suggestion. The multidimensional analysis shows that at each degree of closeness, most sorts of interactions were expected to be lowest for same-sex male pairs, intermediate for female pairs, and highest for cross-sex pairs. Males, then, are believed to obtain emotional gratification from cross-sex relationships, whereas females are intimate with friends of either sex. This inequality, also alluded to in studies of adolescent friendships by Douvan and Adelson (1966) and Komarovsky (1976), implies that males are more emotionally dependent than females on their heterosexual relationships, in that they have learned to minimize the emotional value of their same-sex friendship.

Sex of actor. Differences between male and female actors were in accord with tradition. Males were believed more likely to initiate overt physical activities, such as playing games, and females to initiate verbal interaction, but such sex differences were expected to diminish with the deepening of a relationship.

This finding accords well with independent empirical research. For example, in a study of discussions in heterosexual pairs, Heiss (1962) found that among "casual daters," males were more task oriented and females more social–emotional, but "committed" partners differed little along such traditional lines. Furthermore, two studies of married couples (Levinger, 1964; Raush, Barry, Hertel, & Swain, 1974) found little difference between male and female spouses in their dyadic social–emotional behavior. In other words, increasing closeness is associated with a reduction of sex typing.

Variations in Behavior

Multivariate analyses indicated the presence of a primary dimension indicating general activity or behavioral interdependence and of a secondary dimension indicating intimacy or affective interdependence. The first dimension means that to the extent that partners share one type of interaction, they are judged likely to engage also in other kinds of exchanges. A wide range of behaviors was believed to become more probable as relationships become closer.

The second dimension refers to behaviors that are associated with affection; they include physical touch, hugging, making love,

and expressing affection. For the older raters, negative remarks to the partner as well as strongly positive disclosures were also a component of this "affectivity" dimension.

The present study was designed to compare relationships only with regard to variations in their expected behavior; the two orthogonal dimensions obtained in the present study, therefore, are undoubtedly influenced by the selection of the particular behaviors. Nonetheless, our two dimensions of behavioral and affective interdependence appear generally meaningful. Behavioral interdependence refers to the sheer frequency or likelihood of joint action across a wide variety of behaviors. Affective interdependence, in contrast, refers to the frequency of emotional expression toward the other.

Selection of behaviors in the present study. Previous studies compared relationships with regard to traits (e.g., Wish et al., 1976) or interpersonal feelings (DeSoto & Kuethe, 1959), but in this study we focused on overt action. In doing so, we found great difficulty in selecting a satisfactory sample of interpersonal behaviors, both in getting lists of pair behaviors from samples of naive judges and in constructing such lists ourselves. The list of 30 behaviors was derived largely from the suggestions of college students and other younger persons. It may have been less representative of behaviors that older respondents would have considered important, even though they were able to respond to these items.

Our preliminary study of 50 behaviors had included equal numbers of associative and dissociative acts, but raters had found negative acts to be less clearly interpretable. Because of that ambiguity, and also because the many negative (unlikely) acts may have lowered rater motivation to consider relationships realistically, the main study included only seven negative behavior items. These points are mentioned in the hope that future investigators will arrive at a more elegant sample of behavior items.

Effects of the Raters' Generation

Older raters, recalling the behavior of 22-year-old pairs when they themselves were that age, showed systematically different impressions of pair interaction from those of college-age raters. For almost all relationships and behaviors, old raters recalled significantly lower probabilities of interaction than did young raters; the differences were particularly pronounced for the good friends and for the close relationship. Today's pairs were expected to be more expressive, to do more things together, to reveal a greater proportion of their positive and negative feelings, and to have more physical contact. Such young–old differences were least for superficial activities (e.g., playing games or spending a social evening together) and were less marked for the casual and the married pairs. Parallel young–old differences were found in an internal analysis of the two samples of older raters; for a set of representative behaviors, the raters' ages and their estimates of interaction probabilities were inversely correlated.

One other age difference pertained to the heterosexual relationships. Old raters made a significantly greater distinction between the behaviors of female and male actors. Yesterday's female was perceived as less equal to her male partner than is today's. The generational difference in this study referred particularly to the female's initiation of physical contact and of joint activity.

Limitations of the young–old comparison. It was noted earlier that these data refer to raters' *perceptions* of typical relationships. We do not know how nearly such ratings correspond to "reality." It is possible, for example, that older raters, who had to think back 50 years to recall their own and other people's friendly relationships when they were young, would have distorted their ratings of interaction in some systematic way. In other words, it is possible that yesterday's stereotypes were less accurate than today's. It is also likely that 50 or so years ago, there were in fact fewer instances of "close" heterosexual relationships short of marriage between 22-year-olds; our questionnaire did not ask respondents to indicate the expected frequency of such relationships, but that fact may have influenced these results. Furthermore, while most of the older raters had a middle-class, fairly well educated socioeconomic back-

ground, it is impossible to make a clear-cut socioeconomic comparison between two age groups at such very different stages in their lives.

Nonetheless, our correlational results do support the idea of growing trends toward more openness and expression. Despite the necessary absence of customary experimental controls, our results do confirm intuitive analyses (Stegner, 1971) and documentary evidence (Gadlin, 1977; Murstein, 1974) of historical changes in intimate relationships.

Comparative Assessment of Relationships

The present study has examined a technique for collecting and analyzing "snapshots" of different stages of interpersonal relationships, beginning with casual acquaintanceship and moving through friendship to very close relationship and marriage. Without tracking the interaction of actual pairs over time, it enables one to gather systematic data about people's implicit theories about the conduct and the development of relationships.

Considered alone, the approach yields summaries of people's subjective impressions that permit intergroup comparisons. Considered in conjunction with other more objective observations, a useful overall picture may emerge. Despite its susceptibility to possible distortion, the present procedure permitted a systematic intergenerational comparison that can be integrated with documentary data. Furthermore, a variety of other empirical studies (see Chaikin & Derlega, 1974; Huston & Levinger, 1978; Levinger, 1974; Morton, 1978) have shown that relations at superficial levels are oriented largely toward doing a particular activity with any other person; at deeper levels they become increasingly oriented toward doing any activity with a particular other. The present research tends to confirm this.

Our survey method provides a mapping of people's beliefs about behavior—a means of comparing differences across relationships that vary in depth and sex composition. A weakness in earlier relationship research has been its lack of system, its tendency to study some variables in one kind of relationship (e.g., marriage) and quite other variables for another relationship (e.g., same-sex friendship or first impressions of an acquaintance; Huston & Levinger, 1978). The present method may encourage better comparative research. It may be useful for historical and cross-cultural comparisons and thus may facilitate the recognition that the meaning of interpersonal behavior is embedded in its sociocultural context.

Reference Notes

1. Wolfe, H. *Mate selection, interaction selection, and commitment: A conceptual framework.* Paper presented at the Research and Theory Section of the Meeting of the National Council on Family Relations, St. Louis, October 1974.
2. Levinger, G., Rands, M., & Talaber, R. *The assessment of involvement and rewardingness in close and casual pair relationships.* Tech. Rep. Amherst: University of Massachusetts, June 1977.
3. Purdy, T. F. *Perceptions of involvement in close relationships.* Senior honors thesis, University of Massachusetts—Amherst, 1978.

References

Altman, I., & Taylor, D. A. *Social penetration: The development of interpersonal relationships.* New York: Holt, Rinehart & Winston, 1973.

Booth, A. Sex and social participation. *American Sociological Review,* 1972, *37,* 183–192.

Carroll, J. D., & Chang, J. J. Analysis of individual differences in multidimensional scaling via an *N*-way generation of "Eckart Young" decomposition. *Psychometrika,* 1970, *35,* 283–319.

Chaikin, A. L., & Derlega, V. J. *Self-disclosure.* Morristown, N.J.: General Learning Press, 1974.

DeSoto, C. B., & Kuethe, J. L. Subjective probabilities of interpersonal relationships. *Journal of Abnormal and Social Psychology,* 1959, *59,* 290–294.

Douvan, E., & Adelson, J. *The adolescent experience.* New York: Wiley, 1966.

Gadlin, H. Private lives and public order: A critical view of the history of intimate relations in the United States. In G. Levinger & H. L. Raush (Eds.), *Close relationships: Perspectives on the meaning of intimacy.* Amherst: University of Massachusetts Press, 1977.

Heiss, J. S. Degree of intimacy and male–female interaction. *Sociometry,* 1962, *25,* 197–208.

Huston, T. L., & Levinger, G. Interpersonal attraction and relationships. In M. R. Rosenzeig & L. W. Porter (Eds.), *Annual review of psychology* (Vol. 29). Palo Alto, Calif.: Annual Reviews, 1978.

Jourard, S. M. *Self-disclosure: An experimental analysis of the transparent self.* New York: Wiley, 1971.

Komarovsky, M. *Dilemmas of masculinity: A study of college youth.* New York: Norton, 1976.

Lacey, M. *Implicit theories of relationship.* Unpublished master's thesis, University of Massachusetts —Amherst, 1977.

Levinger, G. Task and social behavior in marriage. *Sociometry,* 1964, *27,* 433–448.

Levinger, G. A three-level approach to attraction: Toward an understanding of pair relatedness. In T. L. Huston (Ed.), *Foundations of interpersonal attraction.* New York: Academic Press, 1974.

Levinger, G., & Snoek, J. D. *Attraction in relationship: A new look at interpersonal attraction.* Morristown, N.J.: General Learning Press, 1972.

Lewin, K. Social-psychological differences between the United States and Germany. In G. W. Lewin (Ed.), *Resolving social conflicts.* New York: Harper & Row, 1948.

Marwell, G., & Hage, J. The organization of role relationships: A systematic description. *American Sociological Review,* 1970, *35,* 884–900.

Morton, T. L. Intimacy and reciprocity of exchange: A comparison of spouses and strangers. *Journal of Personality and Social Psychology,* 1978, *36,* 72–81.

Murstein, B. I. *Love, sex, and marriage through the ages.* New York: Springer, 1974.

Osgood, C. E. Interpersonal verbs and interpersonal behavior. In J. L. Cowan (Ed.), *Studies in thought and language.* Tucson: University of Arizona Press, 1970.

Pleck, J. H. Man to man: Is brotherhood possible? In N. Glazer-Malbin (Ed.), *Old family/new family: Interpersonal relationships.* New York: Van Nostrand, 1975.

Raush, H. L., Barry, W. A., Hertel, R. K., & Swain, M. A. *Communication, conflict, and marriage.* San Francisco: Jossey-Bass, 1974.

Rubin, Z. *Liking and loving.* New York: Holt, Rinehart & Winston, 1973.

Stambul, H. B. *Stages of courtship: The development of premarital relationships.* (Doctoral dissertation, University of California, Los Angeles, 1975). *Dissertation Abstracts International,* 1976, *36*B, 5872. (University Microfilms No. 76-11535)

Stegner, W. *Angle of repose.* Greenwich, Conn.: Fawcett, 1971.

Triandis, H. C., Vassiliou, V., & Nassiakou, M. Three cross-cultural studies of subjective culture. *Journal of Personality and Social Psychology,* Monograph Supplement, 1968, *8,* No. 4, Part 2, 1–42.

Weiss, L., & Lowenthal, M. F. Life-course perspectives on friendship. In M. F. Lowenthal & D. Chiriboga (Eds.), *Four stages of life.* San Francisco: Jossey-Bass, 1975.

Wheeler, L., & Nezlek, J. Sex differences in social participation. *Journal of Personality and Social Psychology,* 1977, *35,* 742–754.

Wish, M., Deutsch, M., & Kaplan, S. J. Perceived dimensions of interpersonal relations. *Journal of Personality and Social Psychology,* 1976, *33,* 409–421.

Received March 27, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 5, 662-673

Evaluation of Outcome Improvement Resulting From Threats and Appeals

André deCarufel
Faculty of Administration
University of Ottawa, Ottawa, Canada

John Schopler
University of North Carolina at Chapel Hill

In a simulated industrial setting, subjects performed a clerical task, believing that their pay was being determined by a peer allocator. After being treated inequitably, subjects were able either to threaten the (fictitious) allocator or appeal to fairness principles, or they had no say ("mute" condition). During a second series of pay periods, subjects' pay either remained constant, improved such that they and the allocator received equal shares, or improved such that subjects received *more* than the allocator ("compensation" condition). The *total* pay was identical in all conditions and created a context of overall inequity. Increased satisfaction and perceived fairness were observed with improved outcomes in both the mute and the threat conditions. In the appeal conditions, satisfaction and perceived fairness were highest in the equality cell. These results are interpreted in terms of relative deprivation. Implications for responses of recipients in ameliorative social programs are presented.

The concept of *relative deprivation* was introduced by Stouffer et al. (1949) to account for "paradoxical" observations in their studies of the American soldier during World War II. Theoretical development in this area has continued to the present day (e.g., Crosby, 1976; Davis, 1959; Gurr, 1970; Runciman, 1966). In all of these formulations, felt injustice has been characterized by a discrepancy between the individual's outcomes and experiences and those that he or she feels are deserved. A similar rationale is evident in recent formulations of equity theory (Adams, 1965; Berkowitz & Walster, 1976), distributive justice (Homans, 1974), and justice theory (Lerner, 1977). The consequences of injustice involve psychological states roughly equivalent to anger and dissatisfaction. These states are hypothesized to motivate behavioral attempts to remedy the injustice, and there exists a strong historical connection between these states and conflict, violence, and revolution (c.f. Davies, 1962; Sears & McConahay, 1973). Though the results of attempts to rectify injustice have often been in the direction of objective outcome improvement, the satisfaction associated with these increases in outcome level is sometimes temporary. This is because the subjective standard by which outcome levels are evaluated may also rise when outcomes improve, a process referred to as the "hedonic treadmill" by Brickman and Campbell (1971).

The improvement itself may be a source of rising expectations, by making salient valued end states to which the individual feels entitled. To the extent that these rising, legitimate expectations are violated by improvements that fail to rise at the same rate or occur in a capricious manner, a second condition for dissatisfaction is present. Finally, if the violation occurs by illegitimate means, as in the case in which an allocator of outcomes is perceived to be intentionally holding back the rate or extent of improvement, the experience of injustice should be acute. Under these conditions, outcome improvement may be predicted to lead to dissatisfaction and hostility rather than to satisfaction (cf. Brown, in Brown & Herrnstein, 1976; Crosby, 1976).

Requests for reprints should be sent to André deCarufel, Faculty of Administration, University of Ottawa, Ottawa, Ontario, Canada K1N 9B5.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3705-0662$00.75

Research Evidence

Anecdotal evidence for this proposition abounds. For example, de Tocqueville observed that "evils which are patiently endured when they seem inevitable become intolerable when once the idea of escape from them is suggested" (quoted in Davies, 1962, p. 6). However, experimental evidence concerning the link between improvement and discontent is scarce. Thibaut (1950) found that disadvantaged parties who were successful in influencing the experimenter to allow them to perform the interesting tasks in an experimental game exhibited more hostility toward the formerly advantaged group than did those who were unsuccessful.

A recent $2 \times 2 \times 2$ experiment by Folger (1977) examined these issues more closely. He constructed a simulated industrial setting in which sixth-grade boys performed a clerical task and were paid over a series of trials by a (fictitious) manager. After an initial period of inequitable treatment, half the boys were allowed to express allocation desires to the manager ("voice" condition), whereas half were not ("mute" condition). During the remaining pay periods, the subjects' pay either improved or remained constant, and the final totals were either inequitable ($\frac{2}{3}$–$\frac{1}{3}$ in the manager's favor) or equitable ($\frac{1}{2}$–$\frac{1}{2}$). The results were that in the inequity half, improvement led to greater perceptions of fairness and lower requests for more pay in a bonus session than did no improvement when it was granted arbitrarily (mute condition), but the *reverse* was true in the voice conditions. In the equity half, perceived fairness was high in all of the cells. What could account for these data, particularly the pattern of means in the inequity half?

Folger suggested that the manager's positive response to the subject's expression of allocation desires (i.e., the improvement) represented an acknowledgment that the appeal was legitimate and created the expectation that fair allocations would follow. The inequity that remained in the final totals dashed these hopes in a particularly cruel fashion. From the subjects' perspective, the manager confirmed their right to higher pay and then blithely continued allocating for his own benefit. The negative response (no improvement) did not endow the appeal with this legitimacy and may even have represented a *denial* of subjects' right to more pay. This response undercut any basis for strong discontent. In the mute conditions, on the other hand, the allocations were experienced relative to those of the first set of pay periods. The improved outcomes were satisfying; the constant ones were not.

Rationale for the Experiment

The Folger (1977) account relies very heavily on the *nature* of the interchange between worker and manager. The "voice" was clearly an *appeal* to principles of fairness. Subjects filled out a "fair pay card" to communicate how they thought pay ought to be allocated. There are other possible interpretations, however, that are simpler in the sense that they suggest that voice per se may have produced the differing evaluations of the improvement in the mute and voice conditions. Discontent in the voice–improvement cell, relative to that in the mute–improvement cell, may have reflected a greater willingness on the part of these subjects to *express* their feelings ("expression hypothesis"). These subjects, after all, had already been encouraged once before by the experimenter to voice their discontent, whereas those in the mute conditions had not. Voice–improvement subjects may have thought the manager less fair because they attributed the improvement to the "external pressure" of their complaint, rather than to the manager's spontaneous desire to increase their outcomes ("attribution hypothesis"; cf. Kelley, 1971). Finally, voice–improvement subjects may have viewed the manager as a bully who "acted tough" in the first few pay periods but who was easily cowed when they confronted him, leading to a lowered evaluation of him and his actions ("derogation hypothesis").

A direct way to disentangle these possibilities is to manipulate the nature of the opportunity for expression available to subjects. If it is voice per se that produces the effect, then allocations following from differ-

ent kinds of voice should be evaluated similarly. If the effect depends on the arousal of fairness principles, then communications to the allocator not involving issues of fairness should not result in evaluations of the subsequent allocations similar to those that do. This raises the general issue of the impact of the *procedures* by which improvements are wrested from advantaged parties on how these "gains" are evaluated.

A particularly interesting type of voice, not involving fairness, is one in which the worker is able to threaten the allocator with dire consequences if the pay is not increased. Note that this maneuver involves expression, allows the improvement to be attributed to "external" pressure, and involves social influence. Thus, from the perspectives of the expression, attribution, and derogation hypotheses it is equivalent to an appeal to fairness. If, on the other hand, fairness principles are necessary for discontent, are there any grounds for predicting how an improvement following a threat to the allocator will be evaluated? The most straightforward prediction is that subjects will feel satisfied when their outcomes increase and dissatisfied when they do not. Satisfaction should be increased when the improved outcomes exceed those experienced previously; it should also be increased by the sense of mastery or control inherent in being able to force an apparently unwilling allocator to move away from the initial allocation pattern. It is also possible that subjects in these threat conditions would be *more* satisfied even than those in the *mute* conditions, where the increase comes about beyond the subjects' control (cf. Wortman & Brehm, 1975). Thus, this "threat" condition is one of special interest, because it could result in either satisfaction or dissatisfaction with the improvement.

To this point, outcome improvement has been discussed in general terms. It may be interesting, however, to examine how improvements of different objective magnitudes are experienced. Several alternatives are available to an allocator confronted by a request from a disadvantaged party for higher outcomes. One is to maintain the status quo (i.e., to continue the deprivation). With respect to improvement, the allocator could terminate the existing discrimination and begin allocating fairly from that point on ("equality"), or he or she could compensate the worker for the past inequity by giving the (former) victim a *better* than even share of the remaining allocations ("compensation"). It is clear that the magnitude of the improvement in the compensation case is objectively greater than that in the equality case, which in turn exceeds that in the constant case. The issue of what is fair compensation for victims of discrimination is a complex and explosive one, as the furor over allegations of "reverse discrimination" indicates. The focus here, however, is on how the victims themselves experience these improvements.

Hypotheses

The hypotheses will be presented in terms of how the improvements should be experienced within each level of the voice factor. It should be noted that *all* of the experimental conditions occurred within the context of overall inequity, since it was in these conditions that Folger (1977) found his most interesting results.

In the *mute conditions*, where improvements are evaluated against initial allocations, satisfaction and perceived fairness should increase as the new outcome level exceeds the old (i.e., be greatest in the compensation case and least in the constant case). In the *appeal conditions*, this pattern should be *reversed*. If the constant allocations represent a denial of the worker's right to higher pay, as Folger suggested, there should be little, if any, discontent when no improvement is forthcoming. In the improvement cases, the positive response should increase the perceived legitimacy of the subject's claim. Further, it is likely that the magnitude of the improvement itself should affect subjects' perception of their own deservingness. The further subjects are taken in the direction of a complete restoration of equity (where Folger found high satisfaction), the more they should expect it and feel entitled to it. These subjects should also feel most chagrined when this standard is not reached.

In a sense, it is like a "goal gradient," where they are taken ever closer to the goal before it is snatched away. Thus, satisfaction and perceived fairness should be *lowest* in the compensation case and greatest in the constant case.

The *threat conditions* could go either way. With no fairness principles involved, it is likely that the improvements would be experienced as increasingly satisfying as they exceed the initial outcome level. Further, the larger the increase that had been forced from the unwilling allocator, the more control or mastery of the situation subjects should feel, which should also increase their satisfaction. On the other hand, the expression, attribution, and derogation hypotheses would predict a pattern similar to that described above for the appeal conditions, since the increase in both cases resulted from "voice."

These predictions suggest a Voice × Improvement interaction. A 3 × 3 experiment crossing outcome improvement (constant, equality, compensation) and opportunity for expression (mute, threat, appeal) was conducted to examine these issues.

Method

Participants

Ninety-seven persons (63 females, 34 males) were recruited via sign-up sheets posted around the University of North Carolina at Chapel Hill. The experimenter scheduled each person by phone. Subjects were run two at a time, but since interaction between them was minimal, the individual was used as the unit of analysis. Five females and two males turned out to be "professional subjects" who attended any experiment involving payment. These subjects expressed some suspicion about the identity of the allocator. Their data were excluded from the reported analyses. The remaining 58 females and 32 males were randomly assigned to conditions until there were 10 respondents per cell.

Procedure

Two subjects at a time reported for an experiment called "Organizational Behavior" and were immediately ushered into separate rooms, which were linked to the experimenter's control room via an intercom system and a set of signal lights. The experimenter began by explaining that the focus of the experiment was on "factors affecting job satisfaction and performance." Subjects were told that the organization in which they would be working would consist of two roles—worker and allocator.

At this point subjects were assigned to roles by the use of a series of signal lights. All subjects discovered that they would be the worker and assumed that the other subject would be the allocator. In fact, the role of allocator was played by the experimenter. Each subject was aware of the sex of the allocator, but there was no systematic relation between experimental condition and sex of subject or the perceived sex of the allocator. The experimenter visited each subject and gave him or her a "worker's folder" containing experimental materials.

The experimenter returned to the control room and specified the duties of the two roles. The worker was told that he or she would perform a clerical task—transcribing numbers set up in the form of business transactions from a computer printout onto a work sheet. The number of items transcribed was to be used as a measure of productivity. Subjects were told that they would be paid by the allocator over a series of 10 pay periods, each lasting 1 minute. The allocator was to receive 24¢ in each pay period from the experimenter and would then divide up this sum between him- or herself and the worker. The experimenter indicated that he would announce the allocator's pay decisions after each pay period. Subjects were asked to write down these allocations on the payment record card in their folders. Subjects were also told that the allocator would check the worker's performance, but would only be able to do so at the end of the final pay period. In order to make the experimental situation more plausible, the experimenter indicated that this phase of the study was concerned with the degree of contact among organization members. They were told that this was the minimal contact condition, which accounted for their being in separate rooms and the allocator dividing up the pay at each pay period despite only being able to check the worker's performance at the end.

At this point, the experimenter visited each subject and answered questions about the procedure. Subjects were asked to fill out a fair pay card that asked what they thought would be a fair way for the allocator to divide up the pay.

The experimenter then announced that the first series of pay periods was about to begin and that subjects should begin their tasks. At 1-minute intervals, the experimenter informed them of the allocations. During this first series, the allocations were consistently unfair to the worker. After Pay Period 5, subjects were asked to total up the allocations on the payment record card and to fill out the first questionnaire.

The opportunity for expression manipulation was introduced at this point. In the mute conditions, the experimenter visited the subjects and asked whether they had any questions about the procedure. In the voice conditions, subjects were able to select a message for the experimenter to deliver to the allocator.

During the remaining five pay periods, the improvement manipulation was introduced. In the con-

stant conditions, the worker received the same outcomes as in the first half, whereas in both improvement conditions, his or her outcomes increased. In the equality conditions, the worker and allocator received equal shares, whereas in the overcompensation conditions, the allocator gave the worker *more* than he or she kept.

After Pay Period 10, the experimenter announced that the work session was over. He asked subjects to calculate their second-half and final allocation totals and to fill out the final questionnaire. When subjects were finished, they were led into a common room, fully debriefed, and paid a standard sum of $2 each for their participation.

Independent Variables

Two factors were manipulated: (a) opportunity for expression and (b) outcome improvement. The opportunity-for-expression factor was introduced between the two series of pay periods. The purpose of this manipulation was to vary the "tone" of the interchange with the allocator and cast the subsequent allocations in different lights. In the mute conditions, there was no opportunity for the subject to contact the allocator. The experimenter simply visited the subject's room and answered any questions the subject might have had. In both voice conditions, however, the subject was able to select one of three messages for the experimenter to pass to the allocator. One of these messages expressed satisfaction with the initial allocations, one suggested that the allocator should take *more* in the second half, and the third expressed dissatisfaction. Of course, it was expected that subjects would choose the third message.

This message went on to include further information specific to the threat or appeal conditions. In the threat conditions, the tone was clearly power-oriented and belligerent. Subjects believed that there would be a short bonus session following the 10th pay period, during which they would be able to bargain with the allocator. During this session, they were told they would be able to deliver loud bursts of noise into the allocator's headphones to back up their demands. The message that the subjects were able to choose emphasized this capability and stated that if their outcomes were not improved in the second half, they would use this weapon during the bonus session. The message concluded by demanding that the total shares of pay end up equal. Thus, this message did not mince words and had a distinctly "or else" quality about it.

In the appeal conditions, however, the message appealed to the allocator's sense of fairness and tried to convince him or her of the legitimacy of the request for a larger share. This message also suggested that the total shares ought to end up equal. To further enhance the legitimacy of the worker's request for more pay, the experimenter supposedly gave the allocator the subject's first questionnaire, which spelled out his or her dissatisfaction and felt injustice, and the subject's work sheet, which detailed the worker's performance during the first half. Thus, this message was based on an appeal to fairness and emphasized the worker's *entitlement* to higher pay.

The opportunity-for-expression factor was crossed factorially by a manipulation of outcome improvement. This manipulation was defined in terms of the amount allocated in the second half relative to that allocated in the first half. In the constant conditions, the allocator took two thirds of the available pay in both halves. The totals were $1.60 for the allocator and 80¢ for the worker. In the two improvement conditions, the final totals were identical, but the outcomes were distributed differently over the 10 pay periods. In the equality conditions, the allocator gave the worker an exactly even split (60¢–60¢) during the second half, whereas in the compensation conditions, the allocator gave the worker more (70¢–50¢) than he or she kept. Table 1 provides an array of the allocation schedules on a trial-by-trial basis for both worker and allocator.

Dependent Variables

The dependent variables were assessed at various points in the procedure. Before the pay periods, subjects filled out a fair pay card, which was a check on the perceived equity norm. Following Pay Period 5, subjects completed the first questionnaire, which assessed their satisfaction with, and perceptions of the fairness of, the initial allocations. Finally, at the conclusion of the work session, subjects completed the final questionnaire, which contained the main dependent variables. Subjects were asked how satisfied they were with their share of the pay in the second half ("satisfaction"), and how fairly they felt the allocator was dividing up the pay in the second half ("fairness"). They were also queried on their feelings about the *total* allocations, and the number of items they transcribed was tallied.

Results

Effectiveness of the Manipulations

The improvement manipulation was checked by asking subjects if their share of the pay changed from the first to the second half. Eighty-nine of the 90 subjects responded correctly ("no" in the constant conditions, "yes" in the improvement ones). They were also asked how much of a change there was in their share of the pay from the first to the second half (1 = none, 11 = very much). The means for the constant, equality, and compensation conditions were 1.2, 8.9, and 9.4, respectively, $F(2, 72) = 315.52$, $p < .001$.

The opportunity-for-expression manipula-

Table 1
Worker and Allocator Pay Schedules for the 10 Pay Periods

	Condition					
	Constant		Equality		Compensation	
Pay period	Worker	Allocator	Worker	Allocator	Worker	Allocator
			First half			
1	8	16	4	20	2	22
2	8	16	4	20	2	22
3	9	15	5	19	3	21
4	7	17	3	21	1	23
5	8	16	4	20	2	22
Sum	40	80	20	100	10	110
			Second half			
6	8	16	12	12	14	10
7	7	17	12	12	15	9
8	8	16	12	12	14	10
9	9	15	12	12	13	11
10	8	16	12	12	14	10
Sum	40	80	60	60	70	50
Total	80	160	80	160	80	160

Note. The amounts are in cents.

tion was checked by giving subjects a choice of four alternatives with which to sum up the content of the message sent to the allocator after Pay Period 5. The choices were (a) sending no message, (b) threatening possible retaliation, (c) appealing to fairness principles, and (d) other (specify). All 30 mute-condition subjects chose (a), 23 of the 30 threat-condition subjects chose (b), and 28 of the 30 appeal-condition subjects chose (c), $\chi^2(6) = 148.10$, $p < .001$.

Finally, subjects were asked how much influence they had been able to exert over the allocator to change the division of pay from the first to the second half (1 = none, 11 = very much). The Voice × Improvement interaction was significant, $F(4, 72) = 9.52$, $p < .01$. The mean for the four voice–improvement cells was 6.25, whereas the mean for the three mute cells and the two voice–constant cells was 1.16.

Evaluation of the Improvement

The analyses reported below were initially run using sex of subject as an additional factor to check on the generality of the effects. However, there were no significant effects involving this variable, so the data for male and female subjects were collapsed.

The first effect of interest is the significant improvement main effect: satisfaction, $F(2, 72) = 45.53$, $p < .001$, and fairness, $F(2, 72) = 63.56$, $p < .001$. The pattern of means indicated that in general, subjects were more satisfied and perceived the allocator to have been more fair when their outcomes improved in the second half than when they did not (see Table 2). The voice main effect was also significant for the satisfaction measure, $F(2, 72) = 5.20$, $p < .01$, but not for the fairness measure ($p > .35$).

These main effects were qualified by the hypothesized Voice × Improvement interaction: satisfaction, $F(4, 72) = 4.62$, $p < .01$, and fairness, $F(4, 72) = 2.43$, $p < .055$. The interaction was explored with Scheffé tests to specify more closely the nature of the effect. Only those tests achieving significance ($p < .05$) on *both* measures will be presented. Within both the mute and threat conditions,

Table 2
Means for the Evaluation-of-Improvement Items

	Opportunity for expression								
	Mute			Threat			Appeal		
Improvement	Constant	Equality	Compensation	Constant	Equality	Compensation	Constant	Equality	Compensation
Satisfaction measure	1.6	6.1	7.3	2.0	8.6	8.6	1.3	8.3	3.6
Fairness measure	1.5	8.6	8.0	1.9	8.0	6.9	1.5	9.3	4.4

Note. Higher numbers indicate higher satisfaction and fairness (1 = not at all, 11 = very much).

subjects felt more satisfied and more fairly treated when their outcomes improved than when they did not. The equality and compensation cells did not differ. Within the appeal condition, however, the improvement in the equality cell resulted in increased satisfaction and perceived fairness compared with the constant cell. The objectively larger improvement in the compensation cell did *not* produce mean differences from the allocations in the constant cell and resulted in significantly *lower* satisfaction and fairness than did that in the equality cell! In fact, the appeal–compensation cell did not differ from the average of all the constant cells and was significantly lower than the average of the rest of the improvement cells. The means for the satisfaction and perceived fairness measures are presented in Table 2.

Total Satisfaction and Fairness Measures

Subjects were asked to indicate their satisfaction and perceptions of fairness of the *total* allocations. The improvement main effect was significant: total satisfaction $F(2, 72) = 7.93$, $p < .001$, and total fairness, $F(2, 72) = 4.14$, $p < .05$. The pattern of means indicated that subjects were more satisfied with, and felt there was more fairness in, the total allocations when there was an improvement in the second half than when there was not. It should be noted again that subjects' total allocations were *identical* in every condition. Finally, the behavioral measure of number of items transcribed yielded no significant effects.

Discussion

The focus of the present research was to examine how disadvantaged parties in a relationship experience improvements of varying magnitudes as a function of the procedures by which these increases in their objective outcome level are introduced.

Evaluation of the Improvement

The main effect for the improvement factor indicates that subjects were more satisfied and felt that the allocator had been more fair when their pay improved in the second half than when it did not. This is consistent with a hedonistic view of human beings as motivated to maximize their outcomes. The demand for compensation is a popular response to inequity (Leventhal & Bergman, 1969), because the disadvantaged party benefits materially, as well as psychologically (with the reduction of "inequity distress"). The present results indicate that in general, compensation will be preferred to no compensation, even if the amount is not sufficient to restore overall equity.

This main effect was qualified by a Voice × Improvement interaction. To facilitate the presentation, the evaluation of the improvement will be discussed within each level of voice separately.

Mute conditions. In the mute conditions, subjects were more satisfied when their outcomes increased then when they did not. The two levels of improvement did not produce differential satisfaction, although the com-

pensation mean was higher than the equality mean. It is likely that the positive experience of the improvement resulted from the increased outcomes exceeding subjects' comparison levels (CL; Thibaut & Kelley, 1959). Since the increase occurred beyond subjects' control, their CL was probably not affected by the higher outcomes and remained rooted in the allocations of the first half. The result was a positive outcome–CL discrepancy in the improvement cells, which was absent in the constant cell where outcomes remained at CL.

Threat conditions. In the threat conditions, the pattern was similar to that described for the mute conditions. Subjects were more satisfied when they were successful in forcing the allocator to give them a larger share of the pay than when they were unsuccessful. This pattern of means is clearly counter to the predictions from the expression, attribution, and derogation hypotheses. It seems that voice per se is not sufficient to produce discontent with improving outcomes. Subjects in the threat conditions were able to voice their feelings during the procedure, the improvements represented a retreat by the allocator from his or her initial "tough stance," and the threat certainly could have been interpreted as "external pressure," yet subjects were *more* satisfied in the improvement cells than in the constant cell. Thus, it may require special types of voice to produce the discontent effect. This possibility will be discussed further below.

The question for now is, Why were subjects satisfied in the improvement conditions? Certainly their outcomes were higher than they had been in the first half. It may also have been that the successful exercise of power and the sense of efficacy it provided was itself satisfying (cf. Wortman & Brehm, 1975). The voice main effect, which showed highest overall satisfaction in the threat condition, is consistent with this combination of hedonic satisfaction and perceived control interpretation. An alternative explanation of the voice main effect relies on subjects' beliefs about the "bonus session," during which they thought they would actually have an opportunity to carry out their threat to deliver a burst of noise into the allocator's headphones. Perhaps the anticipated future contact caused them to be more attracted to the allocator (cf. Darley & Berscheid, 1967) and thus to evaluate his or her actions positively as well. Measures of liking for the allocator were included on the final questionnaire to tap these feelings in the second half, and overall as well. In both cases, the voice main effect was not significant (Fs < 1), making this explanation quite unlikely. A final point is that Thibaut and Kelley (1959) would predict that the improvement, having come about under subjects' control, would contribute to the CL, causing it to rise and eventually blunt the satisfaction associated with the increase. We will say more about this later.

Appeal conditions. The pattern in the appeal conditions deviated somewhat from the predictions, but what did happen seems to be of great theoretical interest. Recall that the predicted pattern required discontent to be highest in the compensation cell and least in the constant cell.

Subjects in the constant cell did not show signs of accepting their continued low outcomes, as one might expect if they had perceived the constant allocations as representing a denial of their right to more pay. Satisfaction in this cell was not higher than in the improvement conditions, as had been hypothesized, and in fact did not differ from the constant cells in the mute and threat conditions. Perhaps this difference between the present study and Folger's (1977) may be explained by differences in the clarity of the inequity induction. The opportunity to voice objections came after *five* periods of inequitable allocation in the present study as compared with only two in Folger's study. Our subjects' beliefs that they had been unfairly treated and were deserving of compensation may have been less subject to change following the allocator's response because of this increased experience. In this regard, the present study probably represents more adequately the state of disadvantaged groups in society, whose beliefs about their own entitlement are not easily shaken.

In the improvement cells, satisfaction was lower in the compensation than in the equal-

ity cell despite the fact that the improvement was objectively larger in the former. It had been expected that bringing subjects closer to the goal of overall equality without reaching it would constitute a greater violation of expectations and produce more dissatisfaction. Overall equity had been assumed as the criterion of fairness, since satisfaction was high in all of Folger's "equity" cells and since the messages sent by subjects in our voice conditions stated that the *total* shares should end up equal. However, the mean satisfaction in the equality cell (8.3 on the 11-point scale) is really too high to be called discontent. Some theoretical modifications appear necessary.

Subjects were clearly dissatisfied with the initial allocations (according to their responses to the first questionnaire, to be discussed below) and felt that they ought to be compensated (28 of 30 subjects in the appeal conditions actually chose the "appeal" message). However, they may have been somewhat uncertain about the level of compensation to expect until they experienced the improvements during the initial pay periods of the second half. We are now suggesting that the magnitude of the improvement given by the allocator served to legitimize one of several acceptable standards. In the equality cell, subjects experienced equal shares (12¢–12¢) in each of the initial periods of the second half. The allocator was thus communicating that a "proximal" standard of "fairness from now on" was the appropriate fairness criterion. In fact, the remaining allocations exactly matched this standard. In terms of relative deprivation, subjects received what they and the allocator had decided was fair, and they felt quite satisfied. In the compensation cells, the initial improvement gave the subjects more than the allocator (14¢–10¢ and 15¢–9¢ in Pay Periods 6 and 7, respectively). This division communicated to the subjects that they deserved not only equal treatment but also compensation for the initial unfairness—a "distal" standard. Later allocations failed to match these expectations, resulting in a strong sense of discontent. In the words of one subject in this condition who wrote in the margin of the final questionnaire, the allocator "attempted to co-opt me by giving me a little, but not enough to make fair allocation."

Further evidence on this point comes from the fact that the fairness measure was highly correlated with the satisfaction measure we have been discussing (within-cells correlation = +.59). Subjects not only felt less satisfied in the compensation than in the equality cell but also felt the allocations to have been less fair. This is exactly what one would expect if there existed a greater discrepancy between deserved and experienced outcomes in the compensation cell. In addition, a recently completed study by deCarufel (in press) directly manipulated the level of the standard and the magnitude of the improvement. The results indicated strong discontent when the distal standard was legitimized but "only" the proximal standard was matched by the improvement.

The Role of Initial Inequity

One final matter requiring comment is the role of the initial inequity. The precondition of initial deprivation was assessed with the first questionnaire, after Pay Period 5. Subjects were clearly dissatisfied ($M = 1.75$) and felt the allocations to have been unfair ($M = 1.67$). It is obvious however that in order to manipulate the second-half allocations systematically, this difference had to result in different total allocations or correspondingly greater initial inequity in the improvement cells. Because of our interest in the evaluation of identical totals differentially distributed over time, we decided to allow initial inequity vary with the improvement.

This is also, by the way, how Folger (1977) chose to deal with this dilemma. Fortunately, in the present study subjects' responses to the first questionnaire afford a method of assessing whether this inevitable confounding in any way damages the interpretations offered above. It might be suggested, for example, that if subjects were initially more dissatisfied in the compensation cell, then the results in the appeal conditions could simply reflect a continuation of these trends. Alternatively, it might have been the greater initial inequity

that caused subjects in the equality and compensation cells to use different standards to evaluate the improvement right from the start. The first point to note is that there was *no* difference in satisfaction after the first half across levels of the improvement factor ($p > .15$) despite the objective differential inequity. There was, however, a main effect for the improvement factor on the fairness measure, $F(2, 72) = 6.75$, $p < .01$. The pattern of means (constant = 2.5; equality = 1.0; compensation = 1.5) indicated that subjects were *less* dissatisfied in the constant cells than in either of the improvement cells ($p < .01$ in each case), which did not differ, $t(58) = .17$, *ns*. Thus, with the lack of a consistent satisfaction–fairness difference across the improvement factor and no difference between the improvement cells on *either* measure, any conjectures based on the effects of this differential inequity on evaluations of the subsequent improvement would appear less plausible. What these measures do show, however, is massive deprivation in *all* of the cells prior to the second half (recall that the means were under 2 on the 11-point scale).

Finally, it might be noted that the de-Carufel (in press) study mentioned above held initial inequity constant and allowed the final totals to reflect the improvement. Similar patterns of discontent with improvement were evident in that study as well, reinforcing the fact that experiencing discontent following outcome improvement does not require initial differences in degree of inequity.

Total Satisfaction and Fairness Measures

The main effect for the improvement factor was significant for both the total satisfaction and total fairness measures, despite the fact that the total outcomes in every condition were *identical*. This result clearly reflects the fact that the second-half allocations carried more weight in determining total satisfaction and fairness than did those of the first half. This recency effect may be due to the later allocations *replacing* the earlier ones (cf. Jones & Goethals, 1971) as indicators of the allocator's intentions. Evidence for this point comes from an auxiliary measure from the final questionnaire that asked subjects to indicate what they believed the allocator *now* felt they deserved. The main effect for improvement was significant, $F(2, 72) = 9.67$, $p < .001$, indicating that subjects in the improvement cells perceived that the allocator regarded them as more deserving than did subjects in the constant cell. These order effects point out the importance of taking into account how outcomes are distributed in time, as well as how the final totals come out, in considering the effects of outcome distributions on satisfaction and perceived fairness.

General Discussion and Conclusions

The major theoretical tool guiding the present experiment has been that of relative deprivation. The alternative mechanisms based on the operation of voice per se (expression, attribution, derogation) received no support from the pattern in the threat conditions. However, the analysis of the appeal conditions based on the "illegitimate violation of legitimate expectations" received support, both from the divergence of the threat and appeal patterns and from the convergence of the satisfaction and fairness measures within the appeal cells.

Most of the research conducted within the relative deprivation framework is concerned with "static" comparisons of one person's outcome total with another's total outcomes. The present article suggests two additional matters that contribute to an individual's satisfaction with his or her outcome level: "dynamic" comparisons of outcome distributions *over time* and procedural circumstances surrounding the distributive actions of the allocator (cf. Thibaut & Walker, 1975).

The effects of allocation schedule and procedure may contribute to satisfaction beyond the effects of the final totals, because these factors may precipitate changes in the standards by which outcomes are evaluated. It was mentioned earlier that perhaps satisfaction in the threat conditions would eventually be lower than in the mute conditions. The CL would rise, since the outcomes increased under the subject's control, and blunt the effect of the improvement. This is the classic analysis

of CL derived from Thibaut and Kelley (1959). Let us refer to this as a "market-type" CL, because it rises and falls with experienced outcomes.

There may be other standards, however, that are developed out of social interaction. In the appeal conditions, it was suggested that the nature of the voicing of opinions and the allocation schedule interacted to invoke various "normative" standards of fairness such as equality or full equity, against which the subsequent allocations were evaluated. Let us refer to this type of CL as a "contract-type" standard, because the fairness principle binds the participants to a particular standard. This would seem to be an instance of a more general process by which standards are negotiated by two (or more) parties and in which the outcomes of the interaction are evaluated in terms of each's obligations under that contract. This type of standard has not been explicitly recognized in the relative deprivation literature, which has focused almost exclusively on temporal (past outcomes) and social (others' outcomes) standards. Yet the interaction of type of standard and pattern of allocation is potentially important. For example, with a market-type CL, gradually improving outcomes that *exceed* CL would lead to the greatest satisfaction, whereas with a contract-type CL, sudden shifts, corresponding to renegotiations, followed by stable periods in which the allocations *match* the terms of the contract would produce greatest satisfaction. According to this logic, the threat conditions and the appeal conditions could still be distinguished when examined over a longer time period.

The present results provide another demonstration of the importance of the subjective evaluation of outcomes, rather than the objective level of outcomes, in determining satisfaction and feelings of fairness with dynamic, or changing, outcome distributions. This finding is particularly striking because outcome improvement would seem, on an intuitive basis, most likely to result in high satisfaction. There are some interesting implications for the evaluation of improvements by disadvantaged groups in society. One especially disturbing one is that those who use violent means to extract the increase would feel satisfied, and those who use nonviolent means would not. A future research direction might be to examine how subjects who have been in some of the conditions of this experiment would respond to a *choice* of voicing strategies, violent and nonviolent, before an additional series of pay periods. One might predict that the threat subjects who felt good about their recent outcomes would choose the threat message again. Those in the appeal-compensation cell, who objectively benefited from their nonviolent approach yet remained dissatisfied, might be prone to choose something other than the appeal again—perhaps the exercise of whatever form of power is available to them. These possibilities underscore the importance for administrators of ameliorative social programs to be very careful about triggering rising expectations in nonviolent groups and reinforcing the actions of violent ones.

References

Adams, J. S. Inequity in social exchange. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2). New York: Academic Press, 1965.

Berkowitz, L., & Walster, E. (Eds.) *Equity theory: Toward a general theory of social interaction. Advances in experimental social psychology* (Vol. 9). New York: Academic Press, 1976.

Brickman, P., & Campbell, D. T. Hedonic relativism and planning the good society. In M. M. Appley (Ed.), *Adaptation level theory: A symposium.* New York: Academic Press, 1971.

Brown, R., & Herrnstein, R. J. *Psychology.* Boston: Little, Brown, 1976.

Crosby, F. A model of egoistical relative deprivation. *Psychological Review,* 1976, *83,* 85–113.

Darley, J. M., & Berscheid, E. Increased liking caused by the anticipation of personal contact. *Human Relations,* 1967, *20,* 29–40.

Davies, J. C. Toward a theory of revolution. *American Sociological Review,* 1962, *27,* 5–19.

Davis, J. A. A formal interpretation of the theory of relative deprivation. *Sociometry,* 1959, *22,* 280–296.

deCarufel, A. Factors affecting the evaluation of improvement: The role of normative standards and allocator resources. *Journal of Personality and Social Psychology,* in press.

Folger, R. G. Distributive and procedural justice: Combined impact of "voice" and improvement in experienced inequity. *Journal of Personality and Social Psychology,* 1977, *35,* 108–119.

Gurr, T. R. *Why men rebel.* Princeton, N.J.: Princeton University Press, 1970.

Homans, G. C. *Social behavior: Its elementary forms* (Rev. ed.). London: Routledge & Kegan Paul, 1974.

Jones, E. E., & Goethals, G. R. *Order effects in impression formation: Attribution context and nature of the entity.* Morristown, N.J.: General Learning Press, 1971.

Kelley, H. H. *Attribution in social interaction.* Morristown, N.J.: General Learning Press, 1971.

Lerner, M. J. The justice motive: Some hypotheses as to its origins and forms. *Journal of Personality,* 1977, *45,* 1–52.

Leventhal, G. S., & Bergman, J. T. Self-depriving behavior as a response to unprofitable inequity. *Journal of Experimental Social Psychology,* 1969, *5,* 153–171.

Runciman, W. G. *Relative deprivation and social justice: A study of attitudes to social inequality in twentieth-century England.* Berkeley: University of California Press, 1966.

Sears, D. O., & McConahay, J. S. *The politics of violence.* Boston: Houghton Mifflin, 1973.

Stouffer, S. A., Suchman, E. A., DeVinney, L. C., Star, S. A., & Williams, R. M. *The American soldier: Adjustment during army life* (Vol. 1). Princeton, N.J.: Princeton University Press, 1949.

Thibaut, J. An experimental study of the cohesiveness of underprivileged groups. *Human Relations,* 1950, *3,* 251–278.

Thibaut, J., & Kelley, H. H. *The social psychology of groups.* New York: Wiley, 1959.

Thibaut, J., & Walker, L. *Procedural justice.* Hillsdale, N.J.: Erlbaum, 1975.

Wortman, C. B., & Brehm, J. W. Responses to uncontrollable outcomes: An integration of reactance theory and the learned helplessness model. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 8). New York: Academic Press, 1975.

Received May 11, 1977 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 5, 674-687

Thematic Effects of Person Judgments on Impression Organization

John H. Lingle
Livington College
Rutgers—The State University

Nehemia Geva
Ohio State University

Thomas M. Ostrom
Ohio State University

Michael R. Leippe
St. Norbert College

Michael H. Baumgardner
Ohio State University

To understand how people communicate information about others and make memory-based interpersonal judgments, it is necessary to understand how person impressions become organized in memory. Consistent with recent research in cognitive psychology on thematic organization of information in memory, Experiments 1 and 2 found that subjects were able to recall (as well as recognize up to a week later) descriptive traits that were relevant, as opposed to irrelevant, to an initial impression judgment. Furthermore, they tended to use judgment-relevant characteristics when asked to freely describe the stimulus person. Subjects' memory for the initial impression judgment itself was excellent and not measurably affected by the relevance of the stimulus information upon which it had been based. A third experiment investigated two possible ways in which thematic organization of an impression might influence a subsequent memory-based judgment about a person. Consistent with other recent research, it was found that when making memory-based decisions, subjects relied on memory for an earlier thematic judgment rather than on memory for descriptive stimulus cues.

A great deal of research in impression formation has been directed toward studying how people integrate items of person information for the purpose of making a single, stimulus-based judgment. "Stimulus based" is used to refer to the fact that such judgments are based on descriptive information provided by the experimenter simultaneously with, or immediately prior to, the time the judgment is made.

Two limitations to this research can be identified. Most day-to-day judgments that people make about each other are "memory based" rather than stimulus based. That is, they are based on a sampling of thoughts from a cognitive representation of the person that a perceiver has built up over time (years in most cases) and stored in memory. To understand the way in which people make memory-based judgments, it is necessary to understand how they organize person information in memory and draw on that information in later social behavior.

A second limitation of most previous work in impression formation is that attention has been focused almost exclusively on the process of making judgments. Yet, a great deal of interpersonal communication involves transmitting facts rather than making judgments. People share remembrances of others' actions, appearance, and other characteristics. Sometimes it is necessary to recall information

Portions of these experiments were reported at the annual meetings of the American Psychological Association, Chicago, September 1975, and the Midwestern Psychological Association, Chicago, May 1976. The research was supported by National Science Foundation Grant GS-38604.

Requests for reprints should be sent to John H. Lingle, Department of Psychology, Livingston College, Rutgers—The State University, New Brunswick, New Jersey 08903.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3705-0674$00.75

about others, whereas at other times it is necessary to decide whether someone else's remembrances are accurate. (These two situations correspond to the recall and recognition tasks used in memory research.)

There are, therefore, two important social consequences of how person impressions are organized in memory. These are (a) how people remember previously acquired person information when it must be transmitted or judged for accuracy, and (b) how memory is drawn upon when a judgment is made. Both these questions are examined in the present article.

Thematic Organization of Impressions

Verbal learning research has demonstrated that memory for stimulus events can be organized around a superordinate theme. For example, it has been shown that sentences are more easily remembered when they are preceded by a thematic title than when no title is provided (Bransford & Johnson, 1973; Dooling & Mullet, 1973). Other research investigating how people remember written texts indicates that a salient theme helps determine what specific items, from a set of items, will later be recalled (Frederiksen, 1975; Rumelhart, 1975; Schank, 1975a, 1975b; Thorndyke, 1977). Finally, it has been found that a salient theme not only influences what factual information is remembered but also what inferences are likely to be made about a stimulus event (Picek, Sherman, & Shiffrin, 1975; Potts, 1972; Sulin & Dooling, 1974). Generally, the inferences people generate and remember tend to be relevant to, and congruent with, the theme that is salient when the stimulus information is initially considered.

Little is known about the themes that serve to organize people's impressions. Some work has shown that distinctive characteristics of a person may serve as an organizing theme. Asch (1946) originally suggested that certain central traits, such as warm and cold, serve as organizing frameworks for people's impressions. It has been shown that memory for information about a person can be organized around traits (Cantor & Mischel, 1977) or around a person's occupation (Anderson, 1977; Anderson & Hastie, 1974).

Rather than seek additional classes of person characteristics that are used as organizing themes, the present research examines whether the act of making an impression judgment will, in and of itself, influence the nature of memory organization. It is often the case that a person's first encounter with another is in a context that calls for a decision or judgment. For example, it may involve something as simple as deciding whether to continue a conversation or something as consequential as offering a job to the other. In both cases the person is receiving and organizing information about the other while actively engaged in making a judgment.

All three studies in this article required subjects to make an occupational suitability judgment (e.g., would this person be successful or unsuccessful as a dentist?) at the time they initially received information about a stimulus person. If such a judgment serves as an organizing theme for a person's impression, this judgment should affect memory (and subsequent memory-based social behavior) in the ways discussed above. Memory for the stimulus features of the person should be affected by the occupational theme. Occupation (theme)-relevant stimulus information should be better remembered than theme-irrelevant information. Furthermore, subsequent judgments and inferences about the stimulus person should be congruent with the thematic organization introduced by the judgment.

The first two studies reported here investigated the effects of a judgment on memory for information used to describe a person. Both recall (Experiment 1) and recognition (Experiment 2) memory were examined. Experiment 3 investigated two possible alternative ways in which thematic organization of an impression might influence a subsequent memory-based person judgment.

Experiment 1

If an initial judgment organizes a person's impression of another, this should be reflected both in the information that is later remem-

Table 1
Mean Number of Judgment-Relevant and Judgment-Irrelevant Traits Recalled: Experiment 1

Stimulus person replication	Trait relevance	
	Relevant	Irrelevant
1	2.69	2.25
2	2.44	2.06
3	2.56	1.81
4	2.25	2.25
M	2.49	2.09

Note. Cell $n = 16$. Scores in each cell could range from 0 to 4 traits.

bered about the person and in the attributes that are used to describe the person. In Experiment 1, subjects were asked to make occupational suitability judgments about stimulus persons described by traits that were both relevant and irrelevant for making the judgment. Afterwards, the subjects were asked to recall the descriptive traits and list additional traits they thought the person might possess. These inferred traits were evaluated for relevance to the occupational judgment. It was expected that subjects would (a) recall more judgment-relevant, as opposed to judgment-irrelevant, traits and (b) describe the person with more judgment-relevant (as opposed to irrelevant) characteristics.

Method

Subjects. Subjects were 116 male and female undergraduates from Ohio State University who participated in partial fulfillment of an introductory psychology course requirement. Of these, 12 assisted in selecting the stimulus materials, 64 participated in the experiment proper, and 40 were used to judge the relevance of the subject-generated descriptions. Where appropriate, subjects were randomly assigned to the experimental conditions.

Stimulus materials. Two pairs of occupations and four eight-trait person descriptions (two for each occupation pair) were used as stimulus materials. The occupations were academician and sportsman (Occupation Pair 1) and comedian and pilot (Occupation Pair 2). The person descriptions accompanying the first occupational pair contained four traits relevant to performance as an academician (e.g., logical, studious) and four traits relevant to performance as a sportsman (e.g., fearless, fast moving). The person descriptions accompanying the other occupational pair contained four traits related to performance as a comedian (e.g., humorous, theatrical) and four traits related to performance as a pilot (e.g., accurate, decisive). The traits for each person description were selected by having five subjects generate adjectives they considered to be relevant to one occupation within a pair but not to the other (e.g., relevant to academician but not to sportsman). These traits were subsequently rated by seven independent judges, and those traits were selected that received the highest ratings of relevance to one of the occupations within a pair but not to the other. Four traits relevant to academician in one of the Pair 1 person descriptions and four traits relevant to comedian in one of the Pair 2 descriptions were negative in evaluative tone. All the other traits were positive in tone.

Experimental procedures and design. Subjects were tested in small groups, and the stimulus materials were presented in booklet format. An introductory page explained that the research was concerned with how people predict success in an activity on the basis of personality trait information. Page 2 listed eight person traits and asked the subjects to consider the suitability of such a person for a designated occupation. Subjects were permitted to study the traits for 60 sec before turning the page. On three successive pages they were then asked (without looking back at the stimulus traits) to (a) record how well they thought the person would perform in the occupation mentioned on page 2 (a 21-point scale was used on which −10 was labeled "unsuccessful" and +10 was labeled "successful"), (b) recall as many traits as possible that had described the person, and (c) list additional traits that they thought would be characteristic of the person.

The experimental design was a $4 \times 2 \times 2$ mixed design with eight subjects per cell. The two between-subjects factors were (a) the four stimulus person replications and (b) which occupation (out of an occupation pair associated with a particular stimulus person) was judged by a group of subjects. The within-subjects factor was recall for the judgment-relevant versus judgment-irrelevant traits. This design insured that each subset of four traits was used equally often as a relevant and an irrelevant subset across the eight groups of subjects.

Results

Recall of stimulus traits. The mean number of relevant and irrelevant traits recalled for the four stimulus person replications is displayed in Table 1. In support of the experimental hypothesis, subjects recalled significantly more judgment-relevant traits than judgment-irrelevant traits, $F(1, 56) = 10.57$, $p < .002$. Although no difference appeared for Replication 4, no significant interaction between the trait relevance and stimulus per-

son replication factors was obtained ($F < 1.0$).

Subject-generated descriptions. The second experimental hypothesis was that the traits subjects use to describe the stimulus person would also be judgment relevant. To test this, it was necessary to determine for each subject the number of generated characteristics that were relevant to the judged occupation and the number that were irrelevant (i.e., more relevant to the other occupation in the pair). To do this, the traits that each subject generated were shown individually to two new groups of 10 subject judges. Each group of judges was asked to decide for each trait whether it indicated success or failure for one of the two occupations associated with the person description.

The criterion for determining the relevance of a generated trait to one of the two corresponding occupations was the absolute amount of consensus among the 10 judges as to whether it indicated success or failure for a particular occupation. For example, if 4 judges rated "outgoing" as indicating success as an academician and 6 judges rated it as indicating failure in this profession, while 7 out of 10 judges from the second group rated the trait as associated with success as a sportsman, the trait was scored as being more relevant to sportsman than to academician. This is because there was greater consensus among the judges as to the trait's implications for sportsman (7/10) as compared to academician (6/10). Traits for which the degree of consensus was a tie were not included in the data analyses.

The analysis of subjects' inferred characteristics paralleled the $4 \times 2 \times 2$ mixed design used for analyzing the recall data. The within-subjects factor in this case consisted of the two categories—judgment relevant versus judgment irrelevant—to which the inferred traits had been assigned.

The mean numbers of relevant and irrelevant inferred traits for the four stimulus person replications are displayed in Table 2. The means are generally consistent with the experimental hypothesis. The statistical analysis produced a main effect of borderline significance for occupation relevance, $F(1, 56) = 3.29$, $p < .08$, and no interaction between relevance and person description replications ($F < 1.0$).

Table 2
Mean Number of Judgment-Relevant and Judgment-Irrelevant Inferred Traits: Experiment 1

Stimulus person replication	Trait relevance	
	Relevant	Irrelevant
1	2.19	1.50
2	2.19	1.25
3	1.81	1.88
4	2.25	1.44
M	2.11	1.52

Note. Cell $n = 16$. Scores in each cell could range from zero to as many additional traits describing the person as the subject wished to list.

Because of the borderline significance of the relevance effect, an additional analysis was conducted on the percentage of judgment-relevant items out of each subject's total number of generated implicational associates. This analysis revealed a significant effect for relevance in the predicted direction, $F(1, 54) = 5.83$, $p < .02$.

Discussion

Subjects in Experiment 1 remembered judgment-relevant traits better than judgment-irrelevant traits and used more judgment-relevant characteristics to describe the stimulus persons. This finding is consistent with the notion that an initial judgment thematically organizes an impression. However, before such a conclusion can be drawn, additional evidence is needed. In the present experiment subjects were asked to remember the stimulus traits immediately following their judgment. Consequently, their recall may have represented short-term, as opposed to long-term, memory organization. Before concluding that a judgment organizes an impression, it would be necessary to demonstrate the persistence of the effect over an extended period.

A second type of evidence needed to establish the organizational properties of a judgment on an impression would be to demon-

strate the relevance effect in a memory recognition test as opposed to a recall test. This would have both practical and theoretical implications. Practically, during social interactions people engage in both types of memory tasks. Theoretically, the two kinds of memory tasks depend to varying degrees on different cognitive processes. Several theorists (Anderson & Bower, 1972; Kintsch, 1968) view recall as a two-step process in which information items are first searched for in memory and then compared with the item being sought. A recall test involves both the retrieval and comparison processes. In a recognition test, the items are supplied, making the comparison processes of primary importance. Testing the organizational properties of a judgment using a recognition task would have implications for whether the locus of the relevance effect lay in the retrieval or comparison processes.

It has been shown that an organizing theme is generally better remembered than the individual items of information upon which it is based (Bartlett, 1932; Frederiksen, 1975; Kintsch, Kozminsky, Streby, McKoon, & Keenan, 1975; Mandler & Johnson, 1977). Consequently, a further type of evidence that would help establish the organizational properties of a judgment would be evidence that the judgment (i.e., the organizing theme) is well remembered and unaffected by the relevance of the information upon which it is based or by the passage of time.

To further investigate how a judgment organizes an impression, Experiment 2 examined (a) whether the relevance effect would be present for recognition memory, (b) the persistence of the relevance effect over time, and (c) subjects' memory for an early judgment made about a person.

Experiment 2

Subjects in Experiment 2 were asked to make an occupational suitability judgment about a stimulus person and then to (a) recall the judgment they had made and (b) identify in a recognition test the traits that had been used to describe the person. The independent variables in the 2 × 2 × 2 between-subjects design were (a) whether the memory tests were administered 1 day or 1 week after the occupational judgment, (b) whether the stimulus traits were relevant or irrelevant to the occupational judgment, and (c) a stimulus person replication factor. Several hypotheses were proposed concerning the effects of these factors on subjects' memory for their judgment and the stimulus traits.

As in Experiment 1, it was predicted that subjects would be more accurate when the traits were relevant, as opposed to irrelevant, to an earlier occupational judgment they had made. This was expected to be true both 1 day and 1 week following the judgment.

Several investigators (Bartlett, 1932; Sulin & Dooling, 1974) have found evidence that subjects use a salient theme to help "reconstruct" related events or stimuli. Such reliance on the theme would likely increase with the passage of time as ancillary cues from the experimental setting become less available to assist subjects in remembering the stimulus traits. The assumption that subjects increasingly use the theme to help reconstruct the traits in a memory test led to several additional predictions. The first of these was that subjects' intrusion errors (i.e., traits identified as stimuli that were not part of the original person description) would be judgment relevant. It was further hypothesized that after a week, as compared to a day, (a) subjects would make more mistakes in recognizing the traits, and (b) the relevance effect for both recognition errors and intrusion errors would be greater.

This reasoning also implies that memory for the theme should be superior to memory for individual items of person information. Unfortunately, there is no way of directly comparing memory for an occupational judgment with memory for descriptive person traits. First, each may be of different familiarity to the subjects. Second, in the course of the experiment, subjects are exposed to several traits but to only one occupation. Nevertheless, the question of subjects' ability to remember their occupational judgment is interesting for two reasons. First, there has been little previous research that directly examines memory for an initial impression

judgment. Second, as explained, subjects' ability to remember a judgment is pertinent to the question of whether a judgment thematically organizes an impression. For these reasons, subjects were asked to recall their occupational judgment when they returned for the second experimental session. It was hypothesized that (a) the relevance of the stimulus information would not effect how well the judgment could be remembered and (b) memory for the judgment would be good and unaffected by whether the memory test was administered 1 day or 1 week later.

Method

Subjects. Subjects were 130 male and female undergraduates from Ohio State University who participated in partial fulfillment of an introductory psychology course requirement. Each subject was randomly assigned to one of the eight experimental conditions.

Stimulus materials. The stimulus materials used in Experiment 2 consisted of two occupations—pilot and comedian—and two groups of 22 traits. These traits were selected from the traits generated by subjects in Experiment 1 to describe the stimulus person evaluated as a comedian or a pilot. The 22 traits in one group consisted of traits that the judges agreed implied success as a comedian, but about which there was low agreement as to whether they implied success as a pilot (e.g., talkative and artistic). The second group of traits were selected to be relevant for judging success as a pilot but low in relevance for judging success as a comedian (e.g., accurate and punctual). From each group, 11 traits were randomly selected to serve as a stimulus person description, and the remaining 11 traits were used as foils in the trait recognition test.

Experimental procedure and design. The subjects were tested in small groups of two to four people, and the stimulus materials were presented in booklet format. An introductory page explained that the research concerned people's ability to predict success in a specific occupation on the basis of personality trait information. Page 2 asked subjects to consider the suitability of a stimulus person for one of the two occupations. This person was described by 11 traits and a small photograph of a neatly dressed male. The subjects were permitted to review the traits and picture for 60 sec before being told to turn the page. On the next page subjects were asked to record their judgment of the stimulus person's suitability for the occupation on a 21-point scale anchored by −10 (very unsuccessful) and +10 (very successful). After recording their rating, subjects were scheduled for a second session either the next day or 1 week later. It was implied that the tasks undertaken in the second session would be unrelated to those completed in the first.

During the second session the subjects received a booklet headed by a picture of the stimulus person. They were first asked to recall the occupation on which they had previously evaluated the person. Next, on an evaluation scale identical to the one used in the first session, they were asked to recall their exact evaluation. Finally, on a separate page, the subjects were instructed to identify the 11 traits that had described the person. These traits were embedded among 22 additional traits, 11 of which were the remaining traits relevant to the occupational judgment the subject had made and 11 of which were relevant to the occupation the subject had not seen. Subjects were told to identify exactly 11 traits from the list of 33.

To summarize the design, half of the subjects were scheduled to return for the second session the next day, and half were scheduled for 1 week later. Half of the subjects judged the stimulus person on an occupation for which the stimulus traits were relevant, whereas half judged an occupation for which the traits were irrelevant. Finally, half of the subjects judged a stimulus person described by 11 comedian-relevant traits, and half judged a person described by 11 pilot-relevant traits.

Results

Seven out of the 130 subjects did not return for the second session. Consequently, the reported analyses are based on data from 123 subjects. The stimulus person replication factor (i.e., whether the stimulus person possessed comedian- or pilot-relevant traits) did not interact with either the judgment relevance or time factor for any of the dependent measures and is, therefore, not discussed further.

Recognition accuracy for the stimulus traits. Consistent with the results of Experiment 1, subjects made more recognition errors when the traits were irrelevant, as opposed to relevant, to their occupational judgment (Ms = 2.87 vs. 2.11, respectively), $F(1, 115) = 9.68$, $p < .002$. Furthermore, subjects made more errors in identifying the traits after 1 week as opposed to 1 day ($Ms = 2.92$ vs. 2.00, respectively), $F(1, 115) = 13.8$, $p < .001$. However, the prediction that the effect of trait relevance would be greater after 1 week than 1 day was not supported, $F(1, 115) < 1.0$ for the interaction.

Nature of subjects' intrusion errors. The 22 incorrect foils in the recognition test seen by each subject consisted of 11 traits that were congruent with the traits that had been

used to describe the stimulus person (i.e., relevant to the same occupation) and 11 traits that were incongruent with the description (i.e., relevant to the other occupation). When subjects made a judgment that was relevant to the descriptive traits, there was no reason for them to have intrusion errors from the 11 traits that were incongruent with the description, since these traits were incongruent with the judgment they had made as well. However, in the irrelevant conditions the 11 traits that were incongruent with the description were congruent with the judged occupation. Thus, to the degree that the judgment thematically organized subjects' impressions, there should have been an increase in the proportion of intrusion errors from the 11 description-incongruent traits. Furthermore, this tendency was expected to increase over time if subjects increasingly came to depend on a judgment-based reconstructive process in recognizing the traits.

An examination of subjects' recognition protocols indicated that significantly fewer than 50% of the intrusion errors were from the description-incongruent trait lists (overall $M = 16.2\%$), $F(1, 105) = 172.5$, $p < .001$.[1] Even in the irrelevant condition, only 20.8% of the intrusion errors were from the set of traits that were incongruent with the person description. The predicted tendency for subjects to have an increased percentage of description-incongruent intrusion errors when the occupational judgment they made was irrelevant (as opposed to relevant) to the person-description traits was marginally significant (Ms = 20.8% vs. 11.0%, respectively), $F(1, 105) = 3.64$, $p < .06$. However, this tendency did not increase with time (the Fs for both the time main effect and its interaction with the relevance effect were less than one).

Recall of the occupational judgment. It was hypothesized that subjects' memory for the organizing theme (i.e., the occupational judgment) would be good and unaffected by the passage of time or the relevance of the traits on which it had been based. To test this, an index was constructed that reflected each subject's ability to recall both the judged occupation and the position she or he marked on the evaluative scale. This index was formed by giving a subject a score of 2 for remembering both the occupation and the exact scale position, 1 for remembering only the scale position or the occupation, and 0 for remembering neither.

In support of the hypothesis that neither trait relevance nor the passage of time would affect subjects' ability to recall their judgment, there was no significant difference in the recall index between subject groups who returned 1 week, as opposed to 1 day, later (Ms = 1.72 vs. 1.79, respectively), $F(1, 115) < 1.0$, nor a significant difference between the relevant and irrelevant trait groups (Ms = 1.79 vs. 1.67, respectively), $F(1, 115) = 1.97$, $p < .20$. The interaction between these factors was also nonsignificant, $F(1, 115) < 1.0$.[2]

It would be expected that memory for the judgment would be better than memory for the traits. Although, as has been noted, the two cannot be appropriately compared, it is worthwhile to comment on subjects' excellent ability to remember their judgment. Across all conditions, 96% of the subjects remembered the occupation they had judged, and 77% were able to reproduce their exact rating on the 21-point scale; 97% were able to reproduce their initial rating within one scale position.

Discussion

The results of Experiment 2 provide four types of evidence supportive of the notion that an initial judgment thematically organizes a person's impression of another. First, consistent with the results of Experiment 1,

---

[1] It was necessary to omit 10 subjects from this analysis, since they made no recognition errors.

[2] In addition to the combined index, subjects' memory for their evaluation and judged occupation was analyzed individually. Subjects' ability to remember their evaluation was unaffected by either the relevance manipulation or the time factor. In addition, there was no effect for relevance on subjects' memory for the occupation they had judged. The only significant finding was a difference in subjects' memory for the occupation after a week as compared to a day. Five subjects out of 123 were unable to remember the occupation they had judged, and all five were in the week-delay condition ($p < .04$).

subjects were better able to recognize a set of descriptive person traits when they were relevant, as opposed to irrelevant, to a judgment they had made. Second, this relevance effect persisted over time and was present both 1 day and 1 week after subjects had made their judgment. Third, there was a tendency for an initial judgment to increase the percentage of intrusion errors from a set of traits that had not been part of the person description, but which were relevant to the judgment. Fourth, subjects' memory for the judgment they had made was excellent and not substantially affected by either the passage of time or variations in the relevance of the stimulus traits.

This fourth conclusion must be viewed with caution. Subjects' memory for their judgment was so good that a ceiling effect may have occurred that masked the effect of time and trait relevance on memory for the judgment. Nevertheless, the fact that subjects' memory for their early decision was exceptionally accurate up to a week later is consistent with the idea that the decision thematically organized their impression and that it remains prominently available in memory as part of their impression of the stimulus person. A consequence of this prominence is examined in Experiment 3.

The predictions stemming from the assumption that over time subjects increasingly rely on their judgment to reconstruct the stimulus traits were not supported. While there was a tendency for subjects' judgments to influence intrusion error rates, the passage of time did not strengthen this effect. Furthermore, time did not increase the tendency for judgment-relevant traits to be recognized better than judgment-irrelevant traits.

There are at least two possible explanations why an interaction between the relevance and time factors did not emerge for either type of recognition error. First, subjects may not have relied on a judgment-based reconstructive process to identify the traits. If this were the case, it suggests that the relevance effect is primarily the result of differential encoding of the stimulus traits rather than reconstructive retrieval processes.

A second possible explanation for the failure to find an interaction between relevance and time is a lack of sensitivity of the experimental method. This could have occurred for several reasons. First, as noted, recognition tests are more sensitive to comparison processes than to retrieval processes. The hypothesized effect may have been visible if a recall, rather than recognition, measure had been employed. Second, the person descriptions tended to provide subjects with an alternative theme to help organize and recall the stimuli. For example, when subjects in the irrelevant condition made a judgment about a pilot, all 11 descriptive traits were relevant to the profession of comedian. Thus, it was easy for them to spontaneously conclude that the person would make a good comic and use this decision to organize their impression. Such an occurrence could account for the high rate of description-congruent intrusion errors even when subjects made an irrelevant occupational judgment. In support of this interpretation, four of the five subjects who erred in recalling the occupation remembered having judged an occupation that was congruent with the descriptive traits they had seen (e.g., TV entertainer, showman).

Experiments 1 and 2 used recall and recognition memory to provide evidence about the organizational properties of a judgment on a person impression. These findings have direct implications for people's ability to accurately convey person information to others. First impressions are indeed lasting, at least in their effects on the information people later have available for communication.

There is a second issue of equal concern: If a judgment organizes an impression, it should be reflected in subsequent memory-based judgments that rest on the impression. This was tested in Experiment 3.

Experiment 3

Experiment 3 investigated how organization of an impression by a judgment affects subsequent judgments. The rationale and alternative experimental hypotheses can best be understood in light of the experimental task. This task consisted of describing a stimulus person with all positive, all negative, or

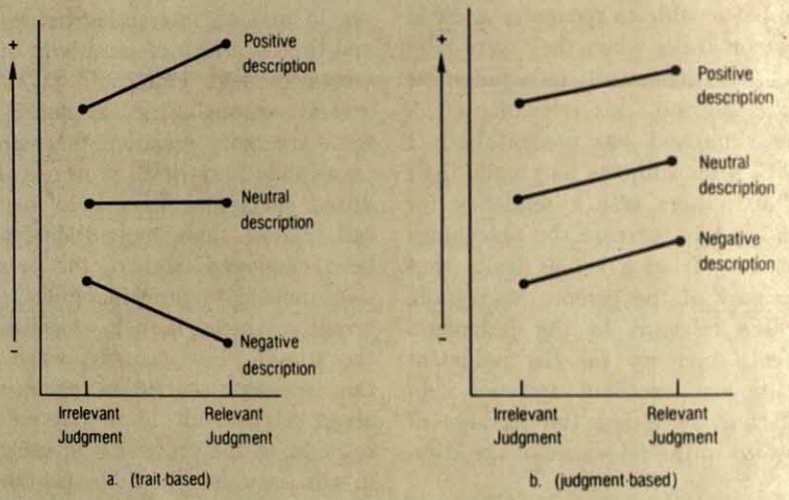


Figure 1. Possible patterns of attribute ratings as a function of judgment relevance and person description valence depending on whether the ratings are trait based (Panel a) or judgment based (Panel b).

all neutral traits and asking subjects to make memory-based ratings of the person's intelligence and friendliness. Before making these ratings, however (and while the traits were still available), subjects were asked to make an occupational judgment about the person. The occupational judgment was relevant to one of the attributes but not the other (e.g., research physicist to intelligence, but not to friendliness, and waiter to friendliness, but not to intelligence). It was expected that subjects' ratings of each stimulus person's intelligence and friendliness would be influenced by the relevance of the occupation they judged to each attribute.

The manner in which an occupational judgment might affect a memory-based attribute rating would depend on the way in which subjects draw on their memory for the person. For example, one way would be for them to recall traits that had been used to describe a person. Experiments 1 and 2 suggest that subjects would best remember traits relevant to their initial occupational judgment. If such memory bias exists, it should have the effect of polarizing subjects' attribute ratings when the occupation is relevant, as opposed to irrelevant, to the attribute being judged.

To understand this prediction, consider the specific case in which a subject is presented with a negatively described stimulus person. The subject is likely to conclude that the person would not be successful, regardless of whether the occupation being judged was research physicist or waiter. However, the occupation judged should affect later memory for the traits. In the case of research physicist, the subject would be likely to remember more intelligence-relevant traits, since intelligence is an important attribute for deciding if someone would be a good physicist, as compared to deciding if that person would be a good waiter. The more intelligence-relevant negative traits a subject can remember, the less intelligent the stimulus person is likely to seem (see Hamilton & Fallot, 1974 for a demonstration of this phenomenon in stimulus-based judgments). Just the opposite would be expected when a subject considers a stimulus person described by positive traits. The increased number of intelligence-relevant descriptive traits remembered following a relevant (as compared to an irrelevant) occupational judgment would be positive, thus making the person seem more intelligent. This interaction between person description valence and judgment relevance that would be expected if subjects based their attribute ratings on memory for the stimulus traits is graphically depicted in Figure 1(a).

It may be that subjects do not draw on memory for descriptive stimulus traits when

making memory-based judgments. Several studies have shown that people rely on their own cognitive responses to stimuli when making judgments (Greenwald, 1968; Lingle & Ostrom, 1979, in press; Petty, Ostrom, & Brock, in press). Lingle and Ostrom (1979), for example, had subjects make pairs of occupational judgments. They concluded that subjects based their second judgment on memory for their first judgment, since second-decision time was not affected by the number of traits that described a stimulus person, but was affected by the similarity of the first occupation to the second. Certainly, in the present Experiment 2, subjects' exceptional accuracy in remembering their occupational judgments suggests that such judgments are readily available in memory to be used as the basis for subsequent decisions.

If subjects were to base attribute ratings on memory for an organizational theme (i.e, the occupational judgment), rather than on memory for stimulus traits, the ratings should reflect the degree to which a particular attribute is stereotypical of the theme (occupation). For example, a subject would seem more likely to judge a person as intelligent if she or he had first judged the person to be a good physicist (as compared to a good waiter), since physicists are stereotypically more intelligent than waiters. One would also be likely to view a negatively described person as more intelligent following an intelligence-relevant occupational judgment, since a bad physicist is still likely to be stereotypically viewed as more intelligent than a bad waiter. Thus, if subjects were to base their attribute ratings on memory for the theme induced by an earlier occupational judgment, the expected pattern of results would be those depicted in Figure 1(b).

A $3 \times 2 \times 2$ within-subjects design was used to test if, and in what way, an initial occupational judgment would influence a subsequent attribute judgment. The three factors varied were (a) person description valence (positive, negative, neutral), (b) relevance of the occupation to the attribute (relevant or irrelevant), and (c) an attribute-rating replication factor (friendliness or intelligence).

Method

Subjects. Subjects were 82 male and female introductory psychology students who participated in partial fulfillment of a course requirement. Of these, 48 took part in the experiment proper, and 34 assisted in preparation of the stimulus materials. Subjects were randomly assigned to different stimulus material presentation orders used to counterbalance the experimental design.

Stimulus materials. Two classes of six occupations were used in the experiment. For one class, friendliness (but not intelligence) was the more relevant attribute for success (i.e., cab driver, airline steward, waiter, baggage porter, telephone solicitor, and shoe salesman); for the other class, intelligence (but not friendliness) was the more relevant attribute (i.e., geologist, research physicist, organic chemist, statistician, medical researcher, and aeronautical technician). These occupations were selected as follows: Two of the authors first chose 30 occupations that they believed to be of the intelligence-relevant class and 29 occupations they believed to be of the friendliness-relevant class from Hopke's (1975) list of occupations. Next, 18 introductory psychology students rated each of these occupations on 8-point scales according to how important they thought knowing about friendliness and intelligence would be for judging a person's aptitude for that occupation. The six intelligence-relevant occupations and the six friendliness-relevant occupations with the largest mean differences between friendliness and intelligence ratings were chosen for use in the experiment.

Three types of stimulus persons (positive, negative, and neutral) were constructed to be used in the experiment. This was done by first selecting 32 traits from three different ranges in Edwards's (Note 1) rescaling of Anderson's (1968) trait list: 250–349 (moderately negative), 350–449 (neutral), and 450–549 (moderately positive). The traits were randomly selected from within each range with the constraint that traits strongly implying anything about either friendliness (e.g., warm) or intelligence (e.g., stupid) were eliminated. Sixteen stimulus persons were then constructed by going through each set of 32 traits twice and randomly selecting (without replacement) eight groups of 4 traits each. To make the stimulus person descriptions of approximately equal relevance for judgments about both the intelligence and friendliness occupations, the 48 generated stimulus persons (3×16) were rated on 7-point friendliness and intelligence scales by a new group of 16 subjects. From each of the three groups of 16 scaled stimulus person descriptions, the 4 descriptions having the smallest absolute difference between their intelligence and friendliness mean ratings were selected to be used in the experiment.

Experimental procedure and design. Subjects were tested in groups of 10 to 20, and stimulus materials were presented in booklet format. The first two pages of the booklet explained that the experiment was

concerned with how people form person impressions from a limited number of descriptive traits and described the experimental task. This task consisted of having subjects make judgments of 12 different stimulus persons. Subjects first saw a list of four traits describing a person and were asked (on the same page) to rate the person on a 7-point scale according to how well they thought the person would perform in a particular occupation. After making this judgment, subjects were asked to turn the page and rate the person (without looking back at the stimulus traits) on 10 different attributes, again using 7-point rating scales. Two of the 10 traits were friendliness and intelligence. The eight attributes rated in addition to friendliness and intelligence were eight traits from Anderson's (1968) trait adjective list, which three of the investigators judged were unrelated to either friendliness or intelligence. These additional attributes were included to conceal the fact that the attributes of intelligence and friendliness were the only ratings of interest. The eight filler ratings were not analyzed.

Each subject judged all 12 stimulus persons, 1 on each of the 12 occupations. The 6 occupations in each class were paired with 2 stimulus persons described by negative traits, 2 described by neutral traits, and 2 described by positive traits. Twelve different rating-scale pages were constructed to be counterbalanced with the stimulus person and occupations. These pages differed only in terms of the order in which the attribute scales appeared. Each was constructed by first randomly assigning the friendliness and intelligence scales to two of the first five positions. The eight additional scales were then randomly assigned to the remaining positions on a page. Thus, for each of the 12 rating-scale pages that a subject saw during the experiment, the friendliness and intelligence ratings always occurred early in the list of ratings, but the order in which the rating scales appeared varied.

Counterbalancing of the 12 stimulus persons, 12 occupations, and 12 rating-scale pages was achieved by a Greco-Latin square. Twelve subjects were required in order for each possible stimulus combination to appear once. Four replications were undertaken by going through the Greco-Latin square four times, bringing the total sample size to 48. Two random orders were used for presenting the 12 stimulus sets.

When a subject initially judged a stimulus person on a friendliness occupation, the subsequent friendliness rating was considered relevant and the intelligence rating irrelevant. It was just the reverse when an intelligence occupation was used. Thus, this was a 3 (stimulus person type) × 2 (relevant or irrelevant occupational judgment) × 2 (friendliness or intelligence rating scale) design.

Results

Occupational suitability ratings. To test whether subjects' attribute ratings were trait based or judgment based, it was first necessary to show that the three levels of stimulus information led to distinctly positive, neutral, and negative occupational judgments on the 7-point rating scales. A repeated measures analysis of variance revealed a strong overall effect for type of stimulus person on subjects' occupation suitability ratings, $F(2, 29) = 223$, $p < .001$. The positive, neutral, and negative stimulus persons received ratings of 5.56, 4.08, and 1.88, respectively. There were no significant differences in the suitability ratings for the two occupation types, $F(1, 47) = 1.57$, nor for the Stimulus Person × Occupation interaction, $F(2, 94) = 1.23$.

Intelligence and friendliness ratings. If subjects' attribute ratings were based solely on memory for the initial stimulus information, occupational relevance should have polarized the ratings. If they were based on memory for the organizing theme, the relevance manipulation should have produced more positive trait ratings. It can be seen in Figure 2 that a strong relevance main effect was obtained, $F(1, 47) = 72.6$, $p < .001$. Even when the thematic judgment was negative (e.g., the person would make a poor research physicist), attribute ratings were more positive following a relevant, as opposed to an irrelevant, thematic judgment. The interaction between relevance and stimulus person valence, as predicted by the memory-for-traits hypothesis, was not significant, $F(2, 29) = 1.67$.

The data analyses did yield two interactions. The first of these was a three-way interaction between relevance, description valence, and attribute rating, $F(2, 94) = 3.50$, $p < .04$. Figure 2 shows that the two trait replications differed primarily in how relevance affected ratings of the negative stimulus person. The second interaction was a two-way interaction between relevance and the attribute ratings, $F(1, 47) = 10.6$, $p < .002$, resulting from the relevance effect being stronger for the intelligence attribute than for the friendliness attribute.

General Discussion

Psychologists have long been interested in the ways impressions are organized. Perhaps

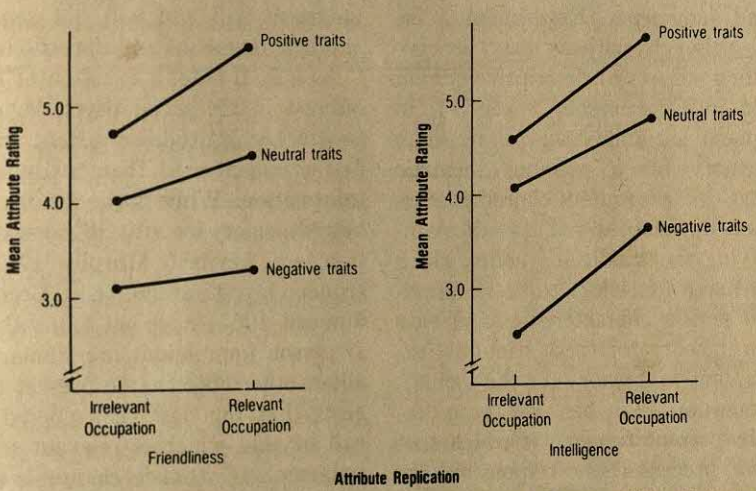


Figure 2. Mean friendliness and intelligence ratings as a function of occupation relevance and person description valence: Experiment 3.

this is because so much social interaction involves transmitting, receiving, evaluating, and acting on information about others. It has generally been assumed that each of these actions depends not only on specific facts that someone knows about another, but on how such facts are organized into a unified impressions. In spite of such interest, there has been little unanimity as to how impression organization should be conceptualized and studied. Asch (1946), for example, conceptualized an impression as a unified gestalt organized by central traits that provide meaning to, and relationships among, peripheral characteristics. More recent theorists (e.g., Cantor & Mischel, 1977; Markus, 1977) have viewed impression organization as the formation of prototypes or cognitive schemata that develop out of stimulus cues or experiences when these cues converge on a common concept.

The approach exposited in the present article conceptualizes impressions as being organized around a central theme that orders the availability in memory of impression-related cognitions. While stimulus configurations may be capable of imparting such organization (as when they are highly interassociated), the present studies investigated the organizational consequences of a judgment made when information about a person is initially received. The importance of investigating the thematic organizational properties of a judgment, as opposed to some other organizing factor, arises from the fact that people often receive information about others in a judgment context. From national elections to job interviews to psychology experiments, people perceive, organize, and encode information about others within a judgment context.

In support of the hypothesis that an early judgment thematically organizes an impression, Experiments 1 and 2 found that (a) the judgment itself was extremely well remembered up to a week after it had been made, (b) judgment-relevant information was better remembered than judgment-irrelevant information (again, up to a week later), and (c) the characteristics subjects used to freely describe the person tended to be judgment relevant. Each of these properties is consistent with the properties that research in cognitive psychology has established as characteristic of an organizing theme in prose or discourse.

While Experiments 1 and 2 investigated the general availability of person information within the context of a recognition or recall test, Experiment 3 examined what subjects access from memory in a judgment situation. Although integration theorists (e.g., Anderson, 1974) have developed algebraic models capable of describing how people combine information cues in a stimulus-based decision, there has been little research investigating what thoughts people access when they make

memory-based judgments. Experiment 3 indicated that although subjects may be capable of recalling subsets of descriptive stimuli (as indicated in Experiments 1 and 2), in certain judgment situations they draw more on their memory for a previous thematic judgment and its associated characteristics. One important, but unanswered, question concerns identifying the conditions under which subjects will make the effort to recall previously learned person characteristics and base a decision on memory for these stimulus features (rather than on memory for the theme).

Another question that has not been answered by the present research is which specific cognitive processes are responsible for the organizing properties of an early judgment on an impression. The fact that the relevance effect was obtained for recognition as well as recall memory suggests that encoding processes may have been partially responsible. This is also indicated by the finding that the relevance effect did not increase with time. If encoding processes do play a role, several mechanisms may be involved. For example, it may be that while considering the person descriptions, subjects generate and store in memory characteristics that are jointly associated with both the descriptive traits and the judgment they have to make. Several studies have demonstrated that people's internally generated thoughts are stored in memory and well remembered (Dosher & Russo, 1976; Greenwald, 1968). These inferred characteristics may later be recalled and serve both as a basis for subjects' descriptions and as a cue for remembering the stimulus traits.

An alternative explanation of the relevance effect could be that the judgment-relevant stimulus traits are processed more deeply (Craik & Lockhart, 1972; Craik & Tulving, 1975) and therefore are remembered better than judgment-irrelevant traits. Memory for these traits might then have served as the basis for subjects' judgment-relevant free descriptions. These, as well as other possible mechanisms, need to be explored by further research in order to establish the conditions under which an early (and perhaps a later) judgment will and will not serve as an organizing theme for an impression.

As a final note, a question of long-standing interest in the social psychological literature is whether attitude-consistent information is better remembered than attitude-inconsistent information. While some studies have found better memory for attitude-consistent information (e.g., Levine & Murphy, 1943), most subsequent investigations (e.g., Greenwald & Sakumura, 1967) have not found this to be true. If person impressions are similar to attitudes about other objects, the present research suggests that the best remembered information will be that which is relevant to an attitude judgment. If attitude change is equated with the reaching of a new attitude judgment, information should be best remembered when it is relevant to attitude change, not associated with attitude constancy.

Reference Note

1. Edwards, J. D. *Revised likableness ratings of 554 personality-trait adjectives.* Unpublished manuscript, Ohio State University, 1967.

References

Anderson, J. R. Memory for information about individuals. *Memory & Cognition,* 1977, *5,* 430–442.

Anderson, J. R., & Bower, G. H. Recognition and retrieval processes in free recall. *Psychological Review,* 1972, *79,* 97–123.

Anderson, J. R., & Hastie, R. Individuation and reference in memory: Proper names and definite descriptions. *Cognitive Psychology,* 1974, *6,* 495–514.

Anderson, N. H. Likableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology,* 1968, *9,* 272–279.

Anderson, N. H. Cognitive algebra: Integration theory applied to social attribution. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 7). New York: Academic Press, 1974.

Asch, S. E. Forming impressions of personality. *Journal of Abnormal and Social Psychology,* 1946, *41,* 258–290.

Bartlett, F. C. *Remembering: A study in experimental and social psychology.* Cambridge, England: Cambridge University Press, 1932.

Bransford, J. D., & Johnson, M. K. Considerations of some problems of comprehension. In W. Chase (Ed.), *Visual information processing.* New York: Academic Press, 1973.

Cantor, N., & Mischel, W. Traits as prototypes: Effects on recognition memory. *Journal of Personality and Social Psychology,* 1977, *35,* 38–48.

Craik, F. I. M., & Lockhart, R. S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior,* 1972, *11,* 671–684.

Craik, F. I. M., & Tulving, E. Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General,* 1975, *104,* 268–294.

Dooling, D. J., & Mullet, R. L. Locus of thematic effects in retention of prose. *Journal of Experimental Psychology,* 1973, *97,* 404–406.

Dosher, B. A., & Russo, J. E. Memory for internally generated stimuli. *Journal of Experimental Psychology: Human Learning and Memory,* 1976, *2,* 633–640.

Frederiksen, C. H. Representing logical and semantic structure of knowledge acquired from discourse. *Cognitive Psychology,* 1975, *7,* 371–458.

Greenwald, A. G. Cognitive learning, cognitive response to persuasion, and attitude change. In A. Greenwald, T. Brock, & T. Ostrom (Eds.), *Psychological foundations of attitudes.* New York: Academic Press, 1968.

Greenwald, A. G., & Sakumura, J. S. Attitude and selective learning: Where are the phenomena of yesteryear? *Journal of Personality and Social Psychology,* 1967, *7,* 387–397.

Hamilton, D. L., & Fallot, R. D. Information salience as a weighting factor in impression formation. *Journal of Personality and Social Psychology,* 1974, *30,* 444–448.

Hopke, W. E. *The encyclopedia of careers and vocational guidance* (Vol. 2). Chicago: Ferguson, 1975.

Kintsch, W. An experimental analysis of single stimulus tests and multiple-choice tests of recognition memory. *Journal of Experimental Psychology,* 1968, *76,* 1–6.

Kintsch, W., Kozminsky, E., Streby, W. J., McKoon, G., & Keenan, J. M. Comprehension and recall of text as a function of content variables. *Journal of Verbal Learning and Verbal Behavior,* 1975, *14,* 196–214.

Levine, J. M., & Murphy, G. The learning and forgetting of controversial material. *Journal of Abnormal and Social Psychology,* 1943, *38,* 507–517.

Lingle, J. H., & Ostrom, T. M. Retrieval selectivity in memory-based impression judgments. *Journal of Personality and Social Psychology,* 1979, *37,* 180–194.

Lingle, J. H., & Ostrom, T. M. Principles of memory and cognition in attitude formation. In R. Petty, T. Ostrom, & T. Brock (Eds.), *Cognitive responses to persuasion.* Hillsdale, N.J.: Erlbaum, in press.

Mandler, J. M., & Johnson, N. S. Remembrance of things parsed: Story structure and recall. *Cognitive Psychology,* 1977, *9,* 111–151.

Markus, H. Self-schemata and processing information about the self. *Journal of Personality and Social Psychology,* 1977, *35,* 63–78.

Petty, R. E., Ostrom, T. M., & Brock, T. C. *Cognitive responses to persuasion.* Hillsdale, N. J.: Erlbaum, in press.

Picek, J. S., Sherman, S. J., & Shiffrin, R. M. Cognitive organization and coding of social structures. *Journal of Personality and Social Psychology,* 1975, *31,* 758–768.

Potts, G. R. Information processing strategies used in the encoding of linear orderings. *Journal of Verbal Learning and Verbal Behavior,* 1972, *11,* 727–740.

Rumelhart, D. E. Notes on schema for stories. In D. Bobrow & A. Collins (Eds.), *Representation and understanding: Studies in cognitive science.* New York: Academic Press, 1975.

Schank, R. C. *Conceptual information processing.* Amsterdam: North-Holland, 1975. (a)

Schank, R. C. The structure of episodes in memory. In D. Bobrow & A. Collins (Eds.), *Representation and understanding: Studies in cognitive science.* New York: Academic Press, 1975. (b)

Sulin, R. A., & Dooling, D. J. Intrusion of thematic ideas in retention of prose. *Journal of Experimental Psychology,* 1974, *103,* 255–262.

Thorndyke, P. W. Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology,* 1977, *9,* 77–110.

Received November 4, 1977 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 5, 688-714

Effects of Preparatory Information About Sensations, Threat of Pain, and Attention on Cold Pressor Distress

Howard Leventhal, Donald Brown, Saya Shacham, and Gretchen Engquist
University of Wisconsin—Madison

Pain experience is conceptualized as a combination of stimulus sensations (e.g., aching) and emotional distress. In Experiment 1, less distress was reported to cold pressor stimulation by subjects first told about stimulus sensations than by subjects who were uninformed or were told about symptoms of bodily arousal (e.g., tension). Adding a pain warning to sensation information blocked distress reduction, presumably by eliciting an emotional interpretation of the stimulus. In Experiment 2, subjects attending only to hand sensations reported less distress than subjects attending to their bodies. This decrease in the power of the stimulus to provoke emotion is presumably mediated by a schema of hand sensations formed by attention. In Experiment 3, subjects attending to hand sensations early in the immersion and distracting themselves later reported the same low levels of distress as did subjects who attended to hand sensations throughout. Subjects distracted throughout and subjects attending to hand sensations later showed no distress reduction. Therefore, stimulus schematization must precede distress reduction. Implications for distress control are discussed.

A number of controlled laboratory (Johnson, 1973; Staub & Kellet, 1972) and field (Cohen & Lazarus, 1973; Egbert, Battit, Welch, & Bartlett, 1964; Johnson, Kirchoff, & Endress, 1975; Johnson & Leventhal, 1974; Johnson, Morrissey, & Leventhal, 1973; Sime, 1976; Wolfer & Davis, 1970) experiments have demonstrated the value of preparatory information for the reduction of distress during exposure to noxious stimulation. Despite these positive findings, our picture as to how preparation reduces distress is not especially clear. We do not know which components of the preparation package affect particular indicators of distress, and we lack a detailed view of the process underlying distress reduction. Given these limits, it is not surprising that a variety of specific, often apparently competing, hypotheses are offered to account for the distress-reducing effects of preparation. For example, preparation is said to reduce distress if it stimulates moderate fear levels and preparatory behaviors (Janis, 1958, 1967); if it provides specific, accurate expectations about stimulus properties (Epstein, 1973; Janis, 1958; Johnson, 1973, 1975; Johnson & Leventhal, 1974); if it provides a neutral label to account for diffuse body reactions that would otherwise be labeled as fear (Nisbett & Schachter, 1966; Ross, Rodin, & Zimbardo, 1969); and/or if it enhances the individual's sense of control over the stressor (Geer, Davison, & Gatchel, 1970; Glass & Singer, 1972; Weiss, 1971).

None of the specific hypotheses appears adequate to account for the empirical find-

This research was supported by Grant SOC 75-02482 from the National Science Foundation to the first author. Experiment 1 was completed in partial fulfillment of the requirements for the Doctor of Philosophy degree awarded to Donald Brown by the University of Wisconsin—Madison. Experiments 2 and 3 were completed in partial fulfillment of the requirements for the Master of Science degree awarded to Saya Shacham by the University of Wisconsin—Madison.

Donald Brown is now at the Department of Psychology, Center College, Danville, Kentucky. Gretchen Engquist is now at the Department of Psychology, University of Virginia, Charlottesville, Virginia.

Requests for reprints should be sent to Howard Leventhal, Department of Psychology, 1202 West Johnson Street, Madison, Wisconsin 53706.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3705-0688$00.75

ings. Thus, moderate levels of fear prior to stressor exposure are reported to innoculate against distress during exposure in some studies (Janis, 1958) but not in others (Cohen & Lazarus, 1973; Johnson, Leventhal, & Dabbs, 1971; Sime, 1976; Johnson, Rice, Fuller & Endress, Note 1). Perceived control is found to reduce distress from noxious stimulation in some studies (Geer et al., 1970; Staub, Tursky, & Schwartz, 1971, Experiment 2) but not in others (Brady, Porter, Conrad, & Mason, 1958; Pervin, 1963; Staub et al., 1971, Experiment 1). When control reduces distress, it does not always do so during actual exposure to the stressor (Glass & Singer, 1972). While no single study can possibly untangle the complete set of factors underlying these apparent inconsistencies, the three experiments reported here are designed to clarify at least one aspect of the problem, the role of stimulus interpretation or stimulus coding (e.g., Holmes & Houston, 1974; Lazarus, 1966; Leventhal, 1970; Sokolov, 1963) in distress reduction. This focus should not be seen as a rejection of the importance of prior fear level or of coping as important contributors to distress reduction; all of these factors are important elements in a complete picture of distress reduction. It is our assumption, however, that they can be better understood if we clarify the way the coding or interpretation of the stressor influences distress during impact.

The general thesis underlying the present experiments is that distress during stressor impact consists of both informational (sensory, stimulus attributes) and emotional (subjective distress or fear) components, which are usually integrated into a common experience (Beecher, 1959; Leventhal, 1970, in press; Leventhal, & Everhart, in press). Three questions are raised respecting the effect of coding or interpretation on the distress experience: (a) Can interpretation actually alter the total experience of stressor impact? (b) What aspects of the experience are actually altered (e.g., the sensory information or the emotional distress)? (c) How does the alteration take place? The three experiments reported here show that distress on impact can indeed be reduced by preparation. Second, they suggest that emotional distress (or suffering) is the component of the experience most likely to be altered by preparation. Third, they suggest that this change happens because identifying or focusing on the sensory properties of the stimulus in a nonthreatening setting leads to the formation of a template or schema of the stimulus that causes it to lose the power to provoke emotional reactions (Sokolov, 1963).

All three investigations study distress induced by cold pressor stimulation. The cold pressor test seems ideally suited for the study of coding or interpretation of noxious stimulation because it provides a rich array of sensory features (Mountcastle, 1968; Wolf & Hardy, 1943); has a relatively slow onset of distress induction, providing time for the impact of mental operations (Wolf & Hardy, 1943); can be affected by psychological manipulations such as set (Barber & Hahn, 1962) and hypnotic analgesia (Hilgard, 1969, 1971); and is not so threatening an experience as to prevent manipulating threat level while holding physical stimulation constant (e.g., Teichner, 1965).

Experiment 1

Experiment 1 compared the level of distress reported during a cold pressor test of groups of subjects given three different types of preparatory information: (a) information on the distinctive features or *sensory* properties of the noxious stimulus; (b) information on the individual's likely overt and covert emotional or *arousal* reactions to the stimulus; and (c) information on the potential strength, *painfulness,* or *magnitude* of the noxious stimulus. Both sensation information and arousal information have been used to reduce distress and fear behaviors on impact with noxious stimulation, and both are presumed to do so because they affect expectations or interpretations of the stimulus. Magnitude or threat information, on the other hand, is expected to limit the operation of both sensation and arousal information. Experiment 1 compared six conditions, sensation information, arousal information, and control, each presented with or without high-

magnitude pain warnings. The specific reasoning and hypotheses for each of the conditions are presented in the following sections.

Sensation Information

Sensation information prepares the individual with a detailed preview of the tactile, thermal, and visual changes that he or she will experience during stressor impact. Johnson (1973) gave subjects sensation information about ischemic pain ("to feel pressure and sensations such as tingling and aching, followed by numbness, . . . arm and hand will be . . . pale and blotched") and in two experiments found substantially lower levels of reported distress for these subjects in comparison to control subjects given information on experimental procedures. There were no treatment differences, however, for reported strength of stimulus sensations; preparation affected only the emotional component of experience.

Beneficial effects of sensation information were also suggested, though less clearly so, in a study by Neufeld and Davidson (1971) and in a study by Staub and Kellet (1972), although in the latter study the investigators did not expect to find the reduction in distress reported for electric shocks by sensation-informed subjects, and it occurred only when the sensation-informed subjects were reassured that the shock was harmless. Finally, reductions have been observed in behavioral signs of distress (e.g., gagging and heart rates) for patients given sensation information the night before an endoscopic examination—a diagnostic test of the upper gastrointestinal tract in which the patient must stay awake and vary his or her position after swallowing and admitting a fiber optic tube (Johnson & Leventhal, 1974; Johnson et al., 1973). Similar effects in emotional distress reduction have been found during cast removal with children (Johnson et al., 1975).

Two hypotheses have been offered to account for the reduction of distress by sensation information. The first argues that the information provides the subject with accurate expectations against which to check stressor impact (Johnson, 1973, 1975). When expectations are accurate, uncertainty, surprise, and startle are minimized, and the individual will believe that the situation is under control. The second hypothesis takes the more inclusive view that sensation information alters the way the noxious stimulus is processed (i.e., it changes the schemata or codes that are integrated with the noxious input). If the subject processes the noxious input in terms of its informational properties and forms a schema of its sensory features, the input will lose the power to stimulate attentional and emotional responses, and there will be a gradual decline or habituation of the emotional (distress) component of the experience. This analysis suggests that the accuracy of the sensation information per se is not the critical factor in producing distress reduction. The critical factor is the schematization or coding of the input as a set of objective features, such as coldness, numbness, and pins and needles, rather than the schematization of the input as pain, fear, and uncertainty as to whether the skin will break and bleed, the flesh will be damaged, and so on (Engel, 1959; Leventhal, 1970, 1975, in press; Leventhal & Everhart, in press).

If the accuracy hypothesis is a sufficient explanation for the beneficial effects of sensation information, one might expect that any accurate information, whether it describes the sensory properties of the noxious stimulus or its painful magnitude, will reduce surprise and distress. Thus, warnings about the painfulness of the stimulus should be distress reducing, as is information about its sensory features. The processing interpretation, however, suggests that warning information is precisely the kind of preparatory input that might encourage encoding a stimulus as a threatening event (i.e., integrating the stimulus with emotional and pain memories that stimulate expectations and uncertainty about possibly harmful outcomes). This type of processing can occur regardless of the accuracy of the warning. The processing analysis is supported by Epstein's (1973) penetrating review of studies of the effects of prior warnings on subjects' reactions to the impact of noxious events. He points out that distress and emotionality are not reduced by high-

magnitude information (warnings that a stimulus will be painful), even if this information is accurate. Epstein found that subjects who expected the 10th of a series of tones to be very severe were more anxious than subjects who expected the tone to be of moderate or mild volume, and he found no evidence that the accurate information about stimulus magnitude reduced distress on impact (as measured by heart rate and skin conductance; Epstein & Clarke, 1970). To the contrary, the evidence suggested that magnitude information that established "an accurate expectation can amplify reactivity to a sufficiently threatening stimulus, at least initially" (Epstein, 1973, p. 103). The Staub and Kellet (1972) findings are consistent with this interpretation.

The accuracy and processing notions lead to quite different expectations, therefore, respecting the effects of preparation on distress to cold pressor pain. The accuracy hypothesis suggests that accurate information on sensations and magnitude of a stressor should reduce distress on impact because both types of information limit uncertainty and define anticipated harm. The processing hypothesis suggests that magnitude and sensation information will have different effects; sensation information will permit the schematization of the noxious stimulus and distress reduction, whereas magnitude information will establish a set for integrating the stimulus with emotional memory schemata that will enhance distress reactions. Thus magnitude information will cancel the effect of sensation information.

If we apply the reasoning described above to cold pressor stimulation, in which stimulus sensations and distress gradually become more intense, we can anticipate the following: If the accuracy hypothesis is true, (a) the group receiving sensation information and the high-magnitude pain warning will show the lowest levels of distress, the unprepared control group will show the highest levels of distress, and the sensation-only and pain-warning-only conditions will fall in between; and (b) the effect on emotional distress should appear early, or as soon as stimulation confirms the preparation. On the other hand, if distress reduction involves the more complex process of coding or schematizing the objective features of the stimulus—a process that can be disrupted by a pain warning—one would expect (a) lower distress ratings with sensation information and no pain warning, with the remaining conditions more or less the same, and (b) condition differences toward the latter part of the exposure period (i.e., after schematization takes place).

Arousal Information

Arousal information involves descriptions of emotional behaviors—the objective and subjective signs of arousal such as heart beating, hand sweating, tenseness, and so forth—which allow individuals to anticipate their overt and covert reactions to a noxious stimulus. Stimulated by Schachter's (Schachter, 1964; Schachter & Singer, 1962) revival of cognition–arousal theories of emotion, investigators have developed a variety of experimental paradigms in which they give subjects arousal information prior to exposure to a noxious stimulus. This information is usually presented along with misattribution information (i.e., information incorrectly linking the arousal signs to a neutral source). For example, Nisbett and Schachter (1966) gave subjects a placebo pill before exposing them to a series of increasingly painful electric shocks. The subjects were told that the pill would cause changes (misattribution) in their emotional reactions (arousal information), for example, heart rate, hand sweating, trembling, and so forth. Because the subjects were led to assume that their arousal behaviors were caused by the neutral pill rather than by the painful shock, they were expected to, and did, tolerate higher levels of shock.

Although the misattribution studies are interpreted as demonstrating that the neutral cognitive label redefines the subjects' emotional state they feel less frightened when they associate their arousal with a neutral pill than with a threatening shock, it is unclear whether the effect of the manipulation is due to the misattribution to the neutral source or to informing the subjects about their arousal behaviors (Leventhal, 1970, p.

172; 1974). Indeed, the major difference between misattribution and control subjects is the receipt of information on arousal behaviors by the former and the receipt of information on a series of irrelevant signs (suggestions that they will feel numbness, itching, etc.) by the latter. Consistent with this, Calvert-Boyanowsky and Leventhal (1975) found that arousal information per se was important in reducing subjects' fear-motivated avoidance behavior (i.e., arousal-informed subjects worked longer at puzzles to win money than at puzzles to avoid shock, whether their arousal was attributed to a "neutral" noise or to a threatening shock). They concluded, therefore, that arousal information and not the source of attribution (neutral noise vs. threatening shock) is the factor responsible for reduction in fear-motivated avoidance behavior.

The effects of arousal information are similar in some respects to those for sensation information, since arousal information appears to reduce avoidance behavior and is less effective in doing so when accompanied by warnings of high magnitudes of danger (e.g., Nisbett & Schachter, 1966). But there is little evidence to suggest that arousal information actually reduces distress on stressor impact. For example, while Calvert-Boyanowsky and Leventhal (1975) found that arousal information reduced the subjects' avoidance behavior, it did not change their subjective emotional state. Nisbett and Schachter (1966) also seem to have found much more substantial differences for avoidance behavior (shock tolerance) than for reports of distress, holding levels of shock constant. These findings raise the question, therefore, as to whether arousal information will actually reduce distress during cold pressor impact, and will do so only in the absence of high-magnitude pain warnings, or whether the experience of distress during cold pressor stimulation will be as severe for subjects with arousal information as for subjects without it.

It was our expectation that arousal information would not reduce distress on cold pressor impact, since it does not assist the subjects in forming a schema of the stimulus features, nor does it assist them in blocking out awareness of stimulus sensations. What seemed likely was that arousal information would prevent the subjects from becoming frightened by their own fear reactions, reducing, therefore, signs of distress-motivated avoidance.

Method

Design and Subjects

The experimental design crossed magnitude information (pain message vs. no pain message) with three information conditions: (a) sensory information, (b) arousal information, and (c) a control group given information on the procedure. Judgments or trials were a within-subjects factor. The magnitude information was included in a single sentence that informed the subjects that they would experience strong pain while their hands were in ice water. The reference to pain was omitted in the low-magnitude conditions. The sensory information message described the physical sensations the subjects would feel during and after the immersion of their hands in the water. The arousal information described specific arousal reactions that would occur during this same time period, and the control message described general details of the experimental procedure.

A total of 50 males were included in the analysis of the study (see Figure 1). While we know of no important sex differences in response to cold pressor or to the experimental treatments, the quit rate is slightly higher for female subjects, and to maximize experimental time we decided to use male subjects only in all three studies. These subjects were recruited from the introductory psychology pool and had completed and met the criteria for participation set by a questionnaire. The questionnaire was used to disqualify subjects who reported medical problems such as hypertension or other circulatory problems, rheumatic fever, heart disease or heart abnormalities, and asthmatic attacks or being on medication for asthma. Reports of any form of convulsion, prior frostbite, taking psychoactive drugs within the past 12 hours, or special medical problems requiring a physician's care (e.g., stomach ulcers) were also grounds for disqualification. One subject was disqualified because of a broken arm.

Experimental Setting

The subject was seated in a comfortable chair in a room containing little visible equipment. There was a low platform that held the ice water tank, three large metal filing cabinets, and a one-way mirror that was almost completely covered by a series of 8¼ × 11-inch color prints. A portable cart held the materials for attaching heart rate electrodes and the questionnaires for the initial portion of the experiment.

All equipment (a Grass polygraph used to record heart rate and skin temperature, a Panasonic videotape recorder used to record subjects' facial expressions, and a Sony tape recorder used to present instructions) was in an adjacent room. The two rooms were connected by an intercom, and the experimenter could view the subject through the spaces between the prints that covered the one-way mirror.

Procedure

Subjects were randomly assigned to the six experimental conditions, with the restriction that a complete replication of the design (a subject in each of the six cells) was run before beginning a second set of six. There were four experimenters, each aided by an assistant experimenter. The assistant placed the appropriate instruction folder on the subject's table, randomized and assigned the subjects to conditions, and gave the experimenter a code letter that indicated which instruction tape to play to the subject. The further precaution was taken of assigning new code letters to the experimental tapes halfway through the study.

The subject was conducted to the experimental room, where he donned earphones and listened to a brief introduction that stated that the study was one of a series of investigations on "psychological and physiological reactions to cold temperature" and advised the student that he could "quit the experiment now or at any time . . . without forfeiting (his) experimental credit."

Informational Conditions

The instructions for each condition were as follows:

(*Sensory information*). The cold temperature treatment that we mentioned will involve submerging your hand (to the wrist) in an ice bath. When you put your hand in the water, the first sensation will be one of extreme coldness. The feeling of coldness will last for a short period of time (around 20 or 30 seconds), and then you will begin to feel a number of different sensations. *The first of these is a sensation of pain, which will begin to get very strong around this time.* (This last sentence was eliminated in the low-magnitude message.) Along with this, you will begin to get a feeling of strong pressure on your hand. You may notice that the feeling of discomfort is not spread evenly around your hand but rather is concentrated in certain areas. For instance, you may begin to feel a tingling sensation in your fingers which seems to bite or burn. Your whole hand may throb after some additional time, and the joints of your fingers will begin to feel somewhat stiff.

After a while, the strong sensations will begin to fade. At this time you will feel a pinpricking sensation or a feeling that your skin is being pulled tightly across the back of your hand. This sensation will fade in your fingers and lower hand until you can feel only numbness. The prickly feeling will remain only in a ring at the point where your hand enters the water.

After your hand is removed from the water solution, you will be aware of a new set of sensations. First, there is a prickly feeling which is similar to a limb waking up. It starts in your fingers and moves slowly up the sides of your palm. You will also detect nervous twitching in the muscular pads of your palm and in your fingers.

(*Arousal information*). The cold temperature treatment that we mentioned will involve submerging your hand (to the wrist) in an ice bath. This experience produces many feelings and reactions in addition to the sensations you will feel in your hand.

When you first put your hand in the water, you will feel a sense of apprehension or anticipation. Almost immediately your whole body will begin to react to the temperature change. *One of the first feelings you will notice is the sensation of pain, which will begin to get very strong about this time.* (This sentence was omitted in the low-magnitude message.) You'll also be aware of additional feelings and reactions. Some of these will be similar to those feelings you have when you are experiencing an emotion such as fear or excitement. You may even be experiencing some of these reactions right now, such as butterflies in the stomach. You may notice that your other hand has begun to sweat. Along with this, you will feel yourself becoming more alert or awake. Generally, you will feel your whole body is exerting a great deal of effort. Your facial muscles, in particular, will show an increase in tension. You might feel your forehead raise and wrinkle. Tension will sometimes spread to other parts of your body—your arms, shoulders, and chest. After a few minutes, this muscle tension may cause a feeling of weakness in various joints and muscles in the legs and chest. After a while your emotional reactions and your feelings of tension will begin to fade.

Once you have removed your hand from the water, you will probably feel a lot of pressure being removed from your body muscles—this is relief. This relief may be accompanied by a surge of warmth and by the feeling of tension rapidly leaving your chest and arms.

(*Procedural information control*). The cold temperature treatment that we mentioned will involve an ice bath. This bath consists of a small bucket of ice-cold water. A cylindrical wire screen is used to keep the ice out of the center area of the bucket. Considering your normal body temperature of 98°, ice temperature represents a large environmental change. *For this reason, after your hand has been*

in the water a short period of time, the sensation of pain will begin to get very strong. (The last sentence was omitted in the low-magnitude message.)

The experimenter is presently preparing to take the physiological measures we need. He has already attached some measuring electrodes. These electrodes are part of a recording apparatus and are connected to a penlike oscillograph located in the next room. Before the next part of the procedure, the experimenter will take another physiological measure—blood pressure. The entire time your hand is in the water, the experimenter will be next door operating all of the recording apparatus, since it is necessary for him to adjust the equipment occasionally as your physiological reactions reach a peak and begin to fade, and he must make small notations on the settings which are being used for recording this rise and fall in your physiological reactions.

After you remove your hand from the water, the electrodes will remain attached for a short period of time. Finally, the experimenter will have some questions to ask you.

Each message terminated with the following instructions: "Your hand will be in the water for a 6-minute period unless you decide to remove it before that time. Now let's go over that again so that you will remember it clearly." All subjects were given a written copy of the appropriate message for their group.

Instructions on safety. At the close of the instructions specific to his condition, the subject was given the following information, which stressed that the experiment was harmless:

Some people who have participated in this experiment have expressed concern about injury to their hand. We want to assure you that under these conditions even ice-cold water cannot and will not cause any damage during the period of time you will be exposed to it. However, as a safety factor, we are monitoring skin temperature and will make certain that your hand does not reach freezing temperatures. There is absolutely no danger of any damage to your hand.

Measures

Mood ratings. Subjects were next instructed to fill out a Mood Adjective Checklist (MACL), because "your emotional feelings or moods at the moment may be related to some of the measures of physiological activity which we are taking. For this reason we would like you to fill out the mood and feelings rating scale located under the manila folder. We'd also like you to fill out the attached questions on your expectations at this point." When the MACL was completed, the experimenter again entered the room to take the subject's blood pressure. Two TV cameras and some additional TV equipment were on one of the cabinet shelves. One of the cameras was live and transmitting to the next room, so that the experimenter could observe the subject's behavior and score expressive reactions. None of the expressive ratings are reported, since very little expressive behavior was observed.

Sensation and distress ratings. After the experimenter took the subject's blood pressure, the ice bath and table were moved to the subject's right, and the experimenter displayed a sensation and a distress scale, which the subject was to use to make ratings during *and* following the immersion of his hand in the cold water. When the experimenter departed, the subject heard the taped instructions for rating sensation and distress during and after immersion of his hand in the cold water. He was told that sensation was "the physical intensity of what you will be feeling," and distress was "the amount of upset or distress the sensations cause." Each scale was printed in the vertical orientation on a separate $8\frac{1}{2} \times 11$-inch page. The points along the sensations scale were marked with numbers, and the points along the distress scale with numbers and verbal labels of distress magnitude. Instructions for rating sensation and distress were the same as Johnson's (1973).

Following the instructions, the subject was told to put his hand in the water and to make sensation and distress judgments whenever the tape indicated. The signals for making the ratings were given 10, 35, and 60 sec after the subject immersed his hand in the water, every 30 sec thereafter for 4 additional judgments, and then once every minute until a 6-min. period had elapsed (10 judgments in all).

The subject was then asked to remove his hand from the water and to continue to make sensation and distress judgments when asked. A total of 6 postimmersion judgments were made beginning 15 sec after the subject removed his hand from the water and continuing at 30-sec intervals for a $2\frac{1}{2}$-min. period. The experimenter returned to take the last blood pressure reading, and the subject completed another MACL.

Postexperimental questionnaires. Following the completion of the MACL, the subject was taken to another room to complete a series of questionnaires that asked about each of the sensations, arousal cues, and procedural events mentioned in the instructions along with questions on other sensations and feelings not mentioned. All items used 20-point Likert scales.

A final MACL was administered, along with a self-esteem questionnaire and a postexperimental questionnaire that asked about thoughts, feelings, and methods the subject used to adapt to stress during the immersion period. Halfway through the questionnaires, the subject was offered his choice of coffee, tea, coke, or milk for a short break. A thorough debriefing completed the procedure.

Results

Manipulation Checks

The postexperimental questionnaire checked on the effectiveness of the manipulations. For each of the 12 sensations mentioned in the sensory message, each of the 12 arousal signs mentioned in the arousal message, and each of the various events mentioned in the procedural information message, the subject was asked to make ratings with respect to the following four questions: (a) How strongly was it expected? (b) How strongly did he remember being told to expect it? (c) How strongly was he aware of the sensation, sign, or event? and (d) How closely did his experience match his expectations? Several other items were included on the questionnaire, and they will be mentioned when results are reported.

Information manipulations. Overall means of subjects' responses to the 12 sensory and the 12 arousal questions are presented in Table 1. Compared to subjects in the other conditions, subjects in the sensory information condition recalled being exposed to the sensory information, $F(2, 44) = 69.26$, $p < .001$, and reported stronger expectations that they would experience these specific sensations, $F(2, 44) = 5.54$, $p < .05$. Subjects given arousal information recalled being exposed to that information, $F(2, 44) = 45.50$, $p < .001$, but when compared to other subjects they did not report significantly stronger expectations that they would experience these arousal symptoms, $F(2, 44) < 1$, $p > .25$. Finally, subjects in the high pain-magnitude conditions recalled being told to expect strong pain, $F(1, 44) = 9.10$, $p < .005$, and reported stronger expectations that they would feel pain than did subjects in the no-pain condition, $F(1, 44) = 7.21$, $p < .01$.

The results are very clear in showing that subjects remember what they have been told. But they also show that what subjects have been told is not always equivalent to what they expect. The high-magnitude pain warning clearly produced differences in expectations for experiences of pain. The sensory information was also reasonably effective in producing differences in expectations regarding these experiences. The one exception to this was the relatively high expectation of sensations for the subjects in the high-magnitude pain warning control condition. The overall mean of the 12 items measuring expected sensations for this group was high be-

Table 1
Mean Responses by Condition for Questions on Sensation Information and Behavioral Arousal Information

Warning	Information	Remember	Expect	Awareness	Match
		Experimental condition			
		Sensation information questions			
High pain warning	Sensation	14.682	12.773	13.000	14.180
	Arousal	2.377	9.195	12.140	13.180
	Control	3.875	11.125	12.180	13.190
No pain warning	Sensation	13.404	11.980	10.270	13.050
	Arousal	4.586	8.838	11.130	12.940
	Control	2.111	9.040	10.270	14.840
		Behavioral arousal questions			
High pain warning	Sensation	2.830	7.568	6.977	14.190
	Arousal	13.455	7.091	6.455	12.690
	Control	1.568	8.102	8.170	14.140
No pain warning	Sensation	3.091	5.687	6.545	13.510
	Arousal	10.566	6.828	8.020	11.700
	Control	2.404	6.054	5.323	15.530

Note. Scores can range from 1 to 20.

cause the high-magnitude pain warning produced a pronounced elevation in three of the expectation items: (a) expected coldness, (b) expected stiffness of the fingers, and (c) expected pinpricking sensations. Each of these questions refers to a sensation known to be associated with severe cold, and the pain warning seems to have alerted subjects to expect these symptoms. Finally, the data suggest that arousal information, although clearly remembered, failed to provide any strong expectation or set for the subjects.

There are two reasons for caution in interpreting the effects of the instructions on subjects' expectations. First, the responses are retrospective reports given after the subjects had been exposed to the cold water stimulus, and we do not know whether the reports at this point would correspond precisely to reports given prior to exposure. For example, it is possible that expectations immediately prior to exposure are more precise, clear, or certain in informed than in uninformed subjects. Second, there is no reason to assume that expectations will have the same effect on later behavior if they are based in one instance on information from the experimenter (sensory informed) and in another instance on information provided by the subject himself (e.g., high-magnitude or pain warning control subjects).

Table 1 also indicates that the preparatory information did not produce suggestion effects, since there were no significant differences between conditions on awareness of sensation or arousal signs. There also were no differences between conditions in reported matches between experience and expectations. But we cannot assume that subjects in different conditions were equally attentive to the sensory experiences and arousal signs, since the question merely asks if the signs were noticed. Similarly, we do not know if the reported match of experience and expectations in unprepared subjects was a postexperimental conclusion saying, "Yes, I guess that was what I did expect," and for prepared subjects was a conclusion during immersion saying, "Yes, I'm feeling just what they told me to expect."

Response to Stressor Impact

Four major dependent measures were obtained during exposure to the noxious cold pressor stimulus: distress ratings, strength-of-sensation ratings, finger temperature, and heart rate. Though there were changes in heart rate over time, the average 2 beats/minute increase during instructions and the average 8–12 beats/minute decrease during immersion being statistically significant, there were no significant treatment differences. Thus, the heart rate data will not be reported or discussed further. Reports of trial effects are omitted, since they are all highly significant.

Distress ratings. Figure 1 presents the 10 distress ratings for the sensation, arousal, and control conditions in high-magnitude pain warning conditions, and Figure 2 presents the same data for the low-magnitude (no pain warning) conditions. It is immediately obvious that distress increased quite sharply during the first minute and a half (the first four judgments) of exposure and then leveled off. But distress reporting was not identical across the three conditions; the sensory information group showed a precipitous drop in reported distress, beginning with Judgment 6 ($2\frac{1}{2}$ min.), in the no pain warning conditions. The decline in distress is far smaller and begins somewhat later, at Judgment 7 (3 min.), in the high-magnitude pain warning condition. The difference between the three types of condition is reflected in a highly significant interaction between treatments and judgment trials, $F(18, 396) = 3.44$, $p < .001$.

It is obvious from the pattern of findings that treatment effects became manifest only during the final 4 min. of immersion; all groups acted similarly during the first 2 min. of cold water impact. Separate planned comparisons were run on Judgments 5 through 10. Subjects in the sensory information groups reported significantly less distress than did subjects in the control conditions, $F(1, 229) = 24.49$, $p < .001$, and the sensory-informed subjects also reported significantly less distress than did subjects in the arousal information groups, $F(1, 220) = 9.76$, $p < .001$. The level of distress reported in the arousal information groups did not differ from that in

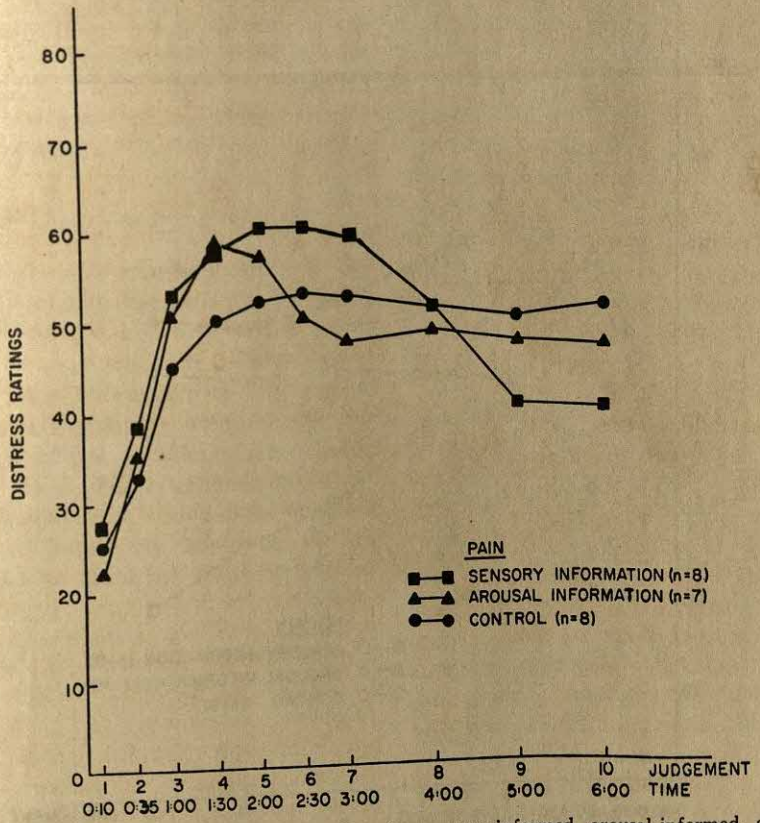


Figure 1. Ratings of distress during immersion for sensory-informed, arousal-informed, and control conditions with high-magnitude pain warning—Experiment 1.

the control conditions. These tests show that sensory information significantly reduced distress during the final judgments (last 4 min. of immersion) in relation to both control (procedural) and arousal information, while arousal information had no substantial affect on distress reduction.

The final critical comparison is between the sensory information group with no pain warning and the sensory information group with the high-magnitude pain warning. If the pain warning blocked distress reduction by sensation information, it would support the processing notions and lead us to reject the simpler version of the accuracy hypothesis. A separate analysis of variance of the last judgments, 5 through 10, for these two conditions (sensation – high pain vs. sensation – no pain) yielded a nearly significant main effect, with less overall distress reported in the no pain warning – sensation group, $F(1, 15) = 4.29$, $p < .06$. The reduction of distress in the sensory information group with no pain warning (M decrease $= 34.12$) was greater than that in the sensory information group with the high pain warning ($M = 20.97$), but the overall interaction of pain warning with trials was not significant, $F(5, 75) = 1.84$, $p < .15$. Given the virtual identity in rated distress for the first five trials, it appears that the pain warning interfered with distress reduction in the final 3½ min. of exposure.

Sensation ratings. Sensory information was expected to change distress ratings but to have no effect on the reported strength of sensations. The data showed, however, a greater drop in the rated strength of sensations in the sensory information groups than in the control groups and low ratings in the arousal information – no pain warning condition. This produced a significant Condition × Trials interaction, $F(18, 396) = 3.98$, $p <$

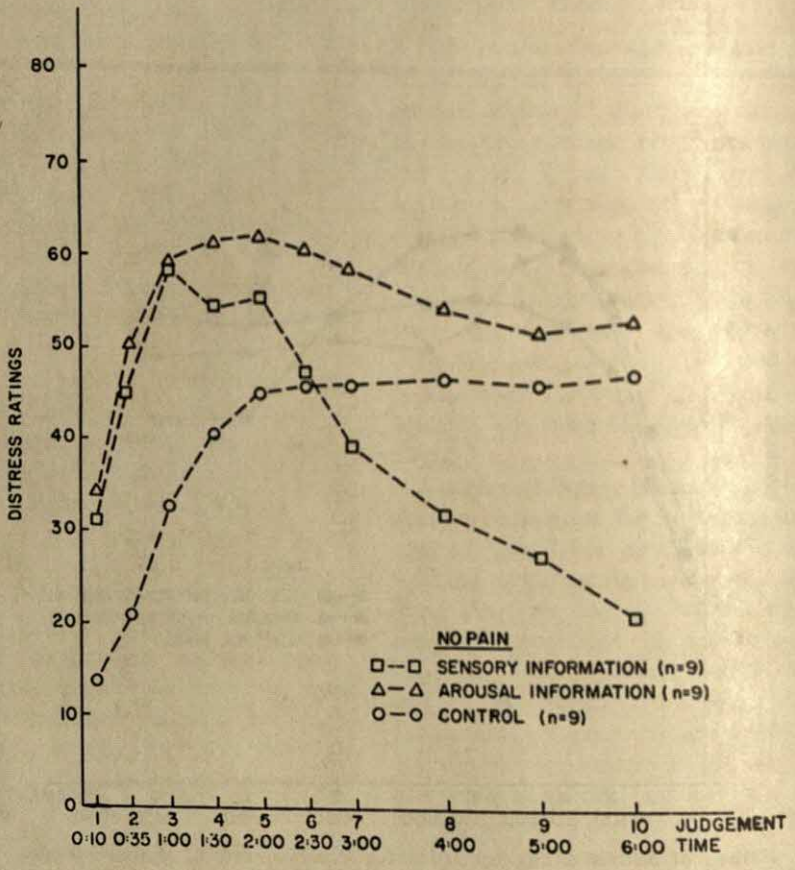


Figure 2. Ratings of distress during immersion for sensory-informed, arousal-informed, and control conditions with no pain warning—Experiment 1.

.001. Despite this effect, the condition differences were smaller than those for distress ratings, and there was no difference between the sensory information – high-magnitude pain warning and sensory information – no pain warning conditions. It is possible that both distress and sensation ratings changed in similar ways because subjects failed to discriminate sharply between them, but the more substantial impact of sensation information on the distress ratings suggests that the differentiation was made by at least some of the subjects. Separate figures are not reported for the sensation data because of their basic similarity to the distress findings.

Finger temperature. Figure 3 presents the finger temperature readings for all six experimental groups at 10 different times during the final 3 min. of cold water immersion. Finger temperatures for the first 3 min. were excluded from the figure and analyses becaus Teichner's (1965) findings showed the latenc of vasodilation to be in the neighborhood 240 sec. It is immediately obvious from Fi ure 3 that the no pain warning – sensory i formation condition shows a distinct rise skin temperature. The difference between t 3- and 6-min. finger temperature was co puted for subjects in all six conditions, a a planned comparison showed significant greater increases in skin temperature for su jects in the sensory information – no pa warning conditions than in the other five co ditions, $F(1, 342) = 41.31$, $p < .001$, a greater increases in skin temperature for t sensory-informed – no pain warning group comparison to the sensory-informed – hig magnitude pain warning group, $F(1, 342)$ 15.83, $p < .001$. The temperature data pr vide support for the hypothesis that senso

information reduced distress only in the absence of pain warnings—a prediction following from the view that processing the stressor input as objective rather than as subjective emotional information is the critical factor in distress reduction.

Postimmersion Questionnaire Responses

The postimmersion questionnaire was designed to help clarify the processes underlying the treatment effects. Significant results were found for one of two sets of questions, a 3-item scale that attempted to assess subjects' strategies for coping with the distress experience (Kanfer & Goldfoot, 1966). This scale focused on the subject's tendency to develop an "objective set" toward the noxious stimulus: (a) "While my hand was in the water, I tried to think of it as something separate from me, so that I was not actually experiencing the discomfort, but rather, my hand was"; (b) "When my hand was in the water, I tried to think of the whole experience as something which would be informative and interesting to observe"; (c) "While my hand was in the water, I tried to concentrate on all the sensations that it was experiencing—sensations such as tingling, pinpricking, etc." As can be seen in Table 2, subjects receiving sensory information gave significantly higher ratings to these items than arousal information or control subjects, $F(1, 44) = 4.38$, $p < .05$. This did not happen because sensory-informed subjects agreed to Item 3, since the second item (informative and interesting to observe) was the best single discriminator of the groups, $F(1, 44) = 4.90$, $p < .04$.

Four other significant findings suggested that both sensation and arousal information

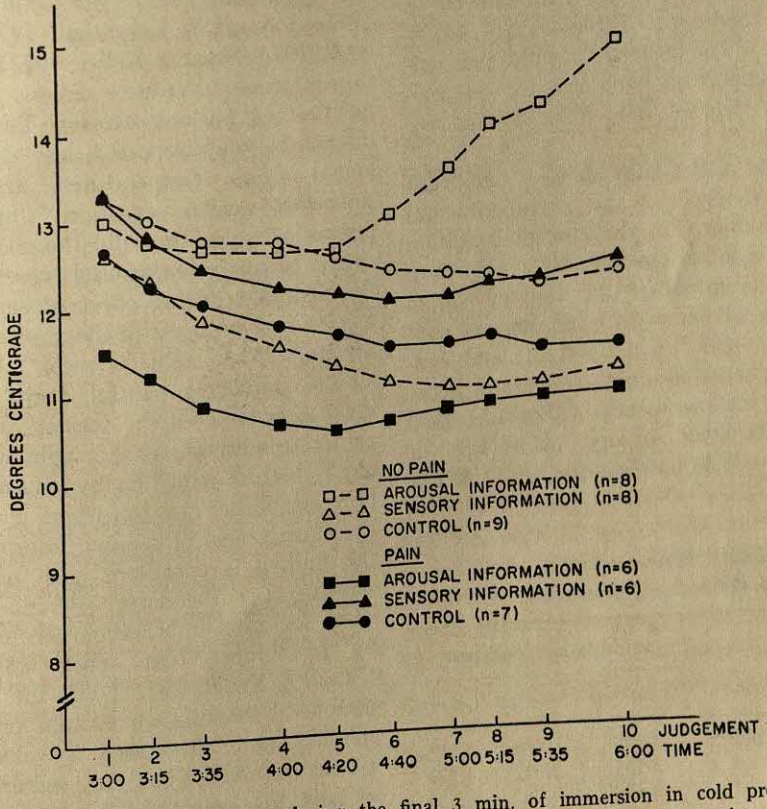


Figure 3. Skin temperature readings during the final 3 min. of immersion in cold pressor for sensory-informed, arousal-informed, and control conditions, both with no pain warning and pain warning—Experiment 1.

Table 2
Self-Ratings on Objective Set Items

Warning condition	Information condition		
	Sensation	Arousal	Control
High pain warning	7.417	5.048	6.458
No pain warning	7.074	5.148	5.074

Note. Scores can range from 1 to 15.

affected emotional states. First, when asked if they had "thoughts and worries about how the cold water might damage your hand," subjects in the arousal information condition reported significantly *less* worry than did subjects in either the sensory or the control group, $F(1, 44) = 4.57$, $p < .05$ (see Table 3). Second, on a postexperimental questionnaire measure of avoidance motivation, subjects in the arousal condition reported less desire to remove their hand from the cold water during the immersion period than subjects in the control conditions (Table 4), $F(1,44) = 4.32$, $p < .05$. The postexperimental questionnaire data support the hypothesis that arousal-informed subjects will show reduced avoidance motivation.

Two mood scales also showed significant treatment differences on the postexperimental measures: Subjects in the control condition were more tense and anxious ($M = 6.5$) than were subjects in either the *sensory* ($M = 2.2$) or arousal ($M = 3.2$) information conditions, $F(1, 43) = 8.26$, $p < .01$, and control subjects reported a greater sense of hopelessness and defenselessness ($M = 4.5$) than did subjects given sensory ($M = 1.6$) or arousal ($M = 2.3$) information, $F(1, 43) = 7.85$, $p < .05$. These differences appeared only in the high-pain warning conditions and only for the postexperimental (post-cold pressor) measures; there were no significant differences on the mood scales given prior to the cold pressor treatment.

Table 3
Means of Subjects' Reports of Thoughts and Worries

Warning condition	Information condition		
	Sensation	Arousal	Control
High pain warning	5.625	1.857	9.125
No pain warning	3.111	2.222	4.222
M	4.294	2.062	6.259

Note. Scores can range from 1 to 15.

Table 4
Means of Subjects' Reported Desire to Remove Their Hand from the Water

Warning condition	Information condition		
	Sensation	Arousal	Control
High pain warning	13.50	8.86	15.38
No pain warning	10.11	8.89	12.44
M	11.71	8.88	13.82

Note. Scores can range from 1 to 20.

Discussion

The results lend clear support to past findings (Johnson, 1973, 1975; Johnson et al., 1975; Johnson & Leventhal, 1974; Johnson et al., 1973; Staub & Kellet, 1972) that sensation information reduces distress during contact with a noxious stressor. The sensation-informed group reported less distress, had higher finger temperatures, and reported somewhat weaker sensations from the cold pressor stimulus than did the control groups or the groups given arousal information. But the distress-reducing effect of sensory information was strongest in the absence of a pain warning. When subjects were told that the stressor was very painful, there were small differences between the sensory, arousal, and control conditions. That a pain warning can block distress reduction by sensory information establishes an important limitation on the effectiveness of sensory information. And the finding is consistent with prior studies showing (a) increased tolerance of shock when sensation information was accompanied by reassurance that shock was harmless (Staub & Kellet, 1972); (b) the blocking of peripheral vasodilation to cold pressor when subjects were threatened by electric shock (Teichner, 1965); and (c) increased reports of pain to electric shock when the subjects were given a pain warning (Hall & Stride, 1954).

The findings suggest that a simple accuracy hypothesis does not readily account for the pain-reduction effects of sensory information and indicate that different kinds of preparatory information alter the way the noxious stimulus information is processed to change the subject's experience of the stressor. Sensory information leads to an objective, nonaffective experience of the stimulus; magnitude information, or a pain warning, leads to an emotional experience of the stimulus. When objectively processed, the noxious input is coded or categorized in terms of specific sensory features, such as coldness, numbness, pins and needles, and so forth, and emotional distress reactions habituate. When emotionally processed, the information is encoded or integrated in an emotional schema or pain memory (Engel, 1959; Leventhal & Everhart, in press), and the stimulus and coding continue to stimulate distress.

The hypothesis that the pain warning activates an emotional pain memory for the encoding of the noxious experience of the cold pressor suggests that the warning might convert each of the sensation cues into a series of threat cues; Epstein (1973) offers a similar interpretation of the effects of pain warnings for subjects' responses to successive numbers in response to a loud noise presented during a sequential countdown procedure. Correlations between heart rates and the number of sensations expected by subjects (measured by the 12 items in Table 1) support this expectation in the present data. Sensation-informed subjects who received a pain warning showed higher heart rates the more sensations they expected ($rs = .16$ to $.62$ for the nine heart rates recorded during immersion), whereas sensation-informed subjects who *did not* receive a pain warning showed lower heart rates the more sensations they expected ($rs = -.46$ to $-.67$ for the nine heart rates recorded during immersion).

Other features of the data support the idea that sensation information leads to an objective or nonthreatening schematization of the stimulus, which then facilitates habituation of distress. First, the subjects given sensation information clearly adopted a more objective, detached set toward the exposed hand. While this finding held whether or not the subjects received a pain warning, the sensation-informed subjects given a pain warning also reported considerable worry about the noxious event; the objective schematization would be in competition, therefore, with an emotional representation of the experience that would be distress stimulating. Second, it is also clear that sensation information produces its effects gradually; the conditions all reached approximately equal peaks of distress at $1\frac{1}{2}$ min. after immersion, and differences between conditions were clear only during the latter $3\frac{1}{2}$ min. of exposure. And while the data were not as clear as we might like, it is reported distress rather than reported strength of sensations that shows the more pronounced decline during this final period, and increased hand temperature (for sensation-informed – no pain warning subjects) follows approximately 2 min. after the beginning of the decline in distress reports. It would appear, therefore, that the acquisition of an objective schema of the noxious stimulus facilitates the habituation or reduction of distress, and this in turn reduces the sympathetic discharge that sustains peripheral vasoconstriction (Teichner, 1965).

Arousal information had relatively little impact on reports of stimulus strength or distress during stimulus impact. It did, however, sharply reduce past experimental reports of worries, thoughts of harm, and reported desire to remove one's hand from the water. Thus, being aware of and expecting a host of bodily reactions seems to have reduced the level of fear and avoidance motivation stimulated by the noxious stimulus sensations and distress. And these effects occurred even though all subjects, arousal-informed or not, expected these responses. Thus, being told about them by the experimenter appears to be a form of reassurance, perhaps, because it suggests that one is responding as do other subjects and need not be threatened or worried by one's reactions. This outcome is compatible with Calvert-Boyanowsky and Leventhal's (1975) interpretation of the misattribution studies.

Experiment 2

Experiment 1 suggested that sensory information reduced distress because it altered the way subjects processed the noxious input; it prepared the subjects to generate an objective schema of the stimulus features, which facilitated the habituation of distress. An alternative interpretation of this effect is that the act of analysis itself and its accompanying "distancing" of the self from stimulation, rather than schematization of the noxious event per se, is the critical factor in distress reduction. Analysis of features may create an attentional set inconsistent with the holistic, integrative set that seems to be necessary for the integration of stimulus perception and bodily reactions into an emotional experience (Blitz & Dinnerstein, 1971; Krueger, 1928; Leventhal, in press; Schachter & Singer, 1962). Emotional experience is described as holistic at a phenomenal level (Krueger, 1928), and experimental data show that analytic sets disrupt emotional experiences associated with positive, humorous stimulation (Cupchik & Leventhal, 1974; Leventhal & Cupchik, 1976).

The considerations discussed above suggested a simple study in which subjects were assigned to one of three conditions: (a) an attention set, in which subjects were asked to monitor the site of impact (i.e., the experiences they were having in their hand while it was in the cold water; (b) an attention-to-hand-and-body set, where subjects were asked to monitor both their hand and their body reactions to the stressor; and (c) an uninstructed control group.

The comparison of attention to hand and body to the control condition and to the attention-to-hand-alone condition would allow us to evaluate a variety of alternative hypotheses. If subjects in the combined attention-to-hand-and-body condition reported less distress than subjects in both control and attention-to-hand-only conditions, it would support the idea that focusing attention on all details of the situation is distress reducing (i.e., that an analytic set and coding of specific features is distress reducing). This outcome also would seem plausible if one believed that a combination of sensation and body information would allow for maximal preparation.

If objective schematization of the input is the critical factor in distress reduction and habituation, it would be essential for the subject to monitor the site of cold pressor impact, and it would be most efficient for schema formation if he paid exclusive attention to the objective features of stimulation to the hand. Assuming that the formation of an objective schema of the input is necessary for habituation also suggests that attention to the hand should produce a decline in distress during the latter 3½ to 4 min. of exposure to the noxious stimulus. This hypothesis assumes that sensation information affected distress in Experiment 1 because it encouraged schematization of objective stimulus features by focusing attention on them. This reasoning is supported by Epstein's (1973) finding that habituation to simple tones occurs only when the subject directly attends to them. When attention is directed elsewhere, there is no habituation to the tones.

Method

Subjects and Design

A total of 68 right-handed males volunteered for this study. All were enrolled in introductory psychology classes at the University of Wisconsin–Madison and received class credit for participation. Five were disqualified because of health problems; six asked to be excused from participation during the cold pressor task; and one subject was dropped because of extremely atypical behavior before the experiment. The remaining 56 subjects were randomly assigned to the three conditions: (a) attention to hand only, (b) attention to hand and body, and (c) control.

Procedure

The subject was seated comfortably in an armchair and heard the introductory instructions over a loudspeaker. These instructions explained the procedure for the attachment of a thermistor measuring hand temperature and assured the subject that the cold pressor treatment would be harmless. Each subject completed a medical questionnaire, to screen out people with health problems. The subject then read and signed a voluntary consent form that indicated that he could quit the experiment without losing credit for participation. All subjects initially agreed to take part in the study. The thermistor was

then attached to the middle finger of the subject's right hand, and the experiment proceeded as follows.

Stress task. Two insulated water tanks were located to the right of the subject—one with water at room temperature and the other with cold water. The room-temperature tank (22 °C) was used to be sure of a common initial hand temperature for all subjects. The cold water was maintained at a constant temperature of 7 °C by a circulating pump that sent the water through an auxiliary ice bucket. Before leaving the room, the experimenter asked the subject to immerse his right hand in the room-temperature tank. The remainder of the instructions were heard over a loudspeaker.

Distress ratings. The instructions for rating distress were identical for all the groups. Subjects were told:

> While your hand is in the cold water, you will be asked to record your judgment on the scale of distress. The distress scale . . . refers to the amount of upset or distress the sensations cause. Zero . . . means you are feeling no upset or bother at all. One hundred means the maximum amount you can imagine of the upset or distress you will experience.

The subject was told to make his judgments when the word *judgment* was said in the tape-recorded instructions. He was told to "say the number on the scale which most accurately describes how you feel at that moment. . . . Each judgment is independent of previous ones. That means your ratings may go up, down, or stay at the same level."

Attention instructions. At this point, the experimenter randomly selected a ballot to assign the subject to one of the three conditions and played back the appropriate instructional tape for that condition. In the attention-to-hand-only condition, the subject heard the following instructions:

> During the time your hand will be in the colder water, you will feel many sensations in your hand. The sensations will change over time. We would like you to pay attention to the sensations or feelings in your hand only, so that you will be able to describe each of the specific feelings and sensations that you experienced. I will ask many questions about the different sensations in your hand afterwards, and this is the most important part of the study. So ignore everything else and pay close attention to the sensations in your hand.

In the condition for attention to hand and body, subjects heard the same instructions, with the exception that the phrase ***hand and body*** was inserted for ***hand***. The control group did not get any of the manipulation instructions.

Distress ratings. The recording then asked the subject to move his hand to the cold water tank and called for the distress judgments. The subject made 13 distress ratings—9 during the 6 min. of the cold pressor task and 4 after removing his hand from the water. The first judgment was requested 10 sec after immersion in the cold water; 2 additional judgments were then requested at 25-sec intervals, and the interval was expanded to 50 sec for the remaining judgments. Six minutes after he placed his hand in the cold water, the subject was instructed to remove it, and 4 additional judgments were requested at 30-sec intervals. The subject made his ratings by calling out a number from 0 (not at all) to 100 (maximum) on the 100-point distress scale.

Distress judgments were the only judgments called for in this experiment and in Experiment 3. Interviews with pilot subjects showed that it was more difficult for them to make the distinction between distress and sensation than in Experiment 1, a difficulty that could stem from the absence of sensation instructions that helped the subjects in Experiment 1 focus more clearly on particular sensory features. Use of a pain rating was rejected, since mentioning the word *pain* prior to cold water immersion might obliterate distress reduction and treatment differences.

Finger temperature. Finger temperature was recorded on a meter located in the control room. The first temperature recording was made while the subject's hand was in the room-temperature tank, and three additional recordings were made, one at each of the distress ratings.

Postexperimental questionnaires. At the close of the procedure, the experimenter removed the thermistor and gave the subject two questionnaires. The first was a sensation and feature checklist asking the subject to rate (on a 21-point scale) the degree to which he noticed various sensations in his hand ("hand throbbing" or "tingling, biting sensations in fingers"), sensations in his body ("speeded-up breathing" or "feeling as if your body is exerting a great deal of effort"), and various features of the environment (e.g., instructions, procedure, and room).

The second questionnaire included direct questions about the percentage of the time the subject focused his attention to his hand and to his body. The subject made these judgments with reference to each of three time periods—the beginning, middle, and end of the cold pressor task. Thus, there were three questions for attention to hand and three for attention to body. The second questionnaire also included a question inquiring about the level of tiredness and tenseness that the subject felt toward the end of the immersion episode. At the conclusion of the session, the subject was debriefed, and the experimenter answered any of his questions about the study.

Results

Distress Ratings

Separate analyses were conducted on the 9 impact ratings given during the exposure to the cold water and the 4 recovery ratings given after the hand was removed from the

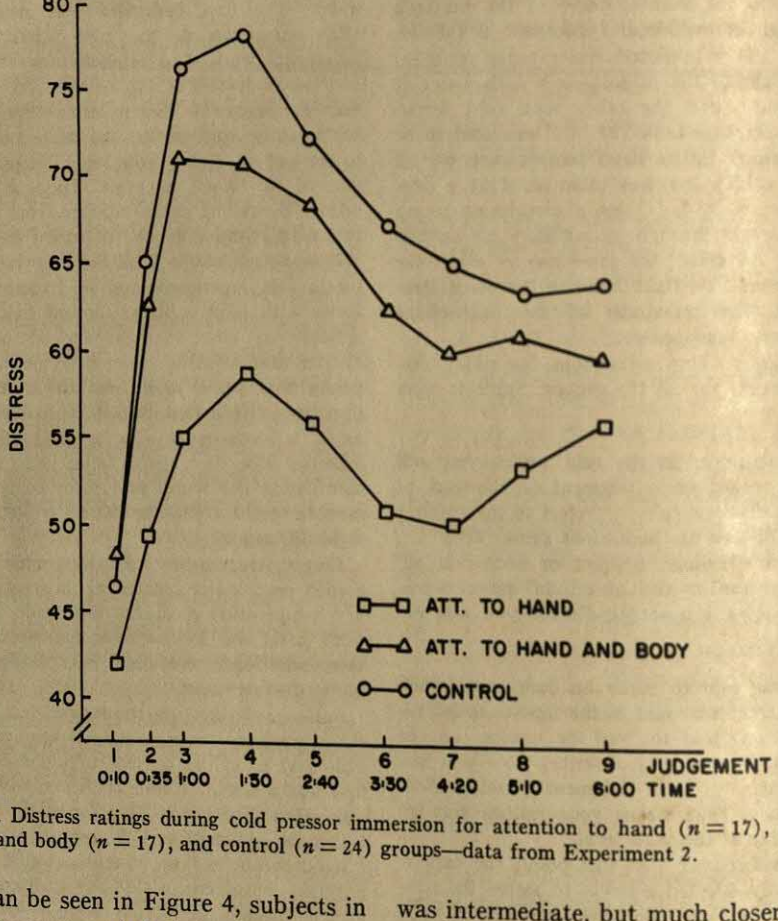


Figure 4. Distress ratings during cold pressor immersion for attention to hand ($n = 17$), attention to hand and body ($n = 17$), and control ($n = 24$) groups—data from Experiment 2.

water. As can be seen in Figure 4, subjects in all three groups reported sharp increases in distress during the first minute of immersion (Judgments 1 to 3), which peaked at Judgment 4 and then leveled off and dropped slightly during the remaining time the hand was in the water. As can also be seen from the figure, the groups began at the same level, but the rate of distress increased very much more rapidly in both the control and attention-to-hand-and-body conditions, so that by the second judgment these two groups were at substantially higher levels of reported distress than the attention-to-hand subjects, and the difference remained at relatively the same magnitude throughout the 6-min. period.

The component of the main effect that tests the hypothesis that attention-to-hand subjects are less distressed ($M = 52.99$) than control subjects ($M = 66.68$) is statistically significant, $F(1, 55) = 4.45$, $p < .05$. The attention-to-hand-and-body group ($M = 61.96$) was intermediate, but much closer to the control than to the hand-only condition, and separate tests show that it did not differ significantly from either the hand-alone, $F(1, 55 = 2.12$, *ns*) or the control condition, $F(1, 55 = .58$, *ns*). Finally, there is no significant Treatment × Trials interaction.

Recovery Ratings

Although distress declined sharply during the recovery phase, there were no treatment differences.

Finger Temperature

There were no significant treatment differences or interactions of treatments with trials. The failure to find effects may be due in part to the change in water temperature (2 °C in Experiment 1, 7 °C in Experiment 2), difficulties with the temperature meters and elec

trodes, and factors related to the time of year at which the data were collected.

Postexperimental Questionnaire Reports

Awareness. Subjects estimated the amount of attention they paid to their immersed hand during the beginning, middle, and end of the 6-min. immersion in the cold water and the amount of attention they paid to their body reactions during each of these same three periods. These ratings were made on 11-point scales ranging from 0% to 100% of attention. Subjects also completed a checklist on which they rated the degree to which they felt each of 35 sensations (1 = not at all aware of it, 21 = very strongly aware of it). Eight of these items related to the hand only, and 11 to the body reactions only. The remaining 16 items focused primarily on attention to situational factors such as the procedure.

Significant treatment effects appeared on both measures: Subjects in the attention-to-hand-and-body condition reported paying a greater percentage of attention to their body reaction ($M = 32.05\%$) than did subjects in either the attention-to-hand-only condition ($M = 17.35\%$) or the control condition ($M = 14.82\%$), $F(2, 100) = 5.68$, $p < .01$; the control and attention-to-hand-and-body conditions were not significantly different. Subjects in the attention-to-hand-and-body condition also checked greater awareness of specific body reactions ($M = 79.7$); the control group was intermediate ($M = 59.5$); and the least aware were subjects in the hand-only condition ($M = 39.4$), $F(2,47) = 5.77$, $p < .01$. Each of the groups was significantly different from the others at the .05 level by t test. There were no significant treatment effects for the measures of attention to the hand.

Fatigue. On three 21-point scales, subjects responded to the following items: "Toward the end of the time your hand was in the water, did you feel (a) tired, (b) tense and jumpy, (c) detached and daydreamy?" Subjects instructed to attend to their hands only reported lower levels of tiredness ($M = 2.86$) than was reported by subjects instructed to attend to hand and body ($M = 6.4$) or by control subjects ($M = 5.6$), $F(2, 42) = 3.94$, $p < .05$.

Discussion

Attention to the site of stressor impact, the hand only, reduced levels of reported distress during cold pressor. Instructions to attend to both hand and body did not, however, produce statistically significant effects. Thus, the mere act of selectively deploying attention and analyzing the stimulus field is not sufficient for distress reduction; focusing of attention on the site of stressor impact is critical for distress reduction, through a habituation-type process.

While the data are generally supportive of a processing interpretation (i.e., that lower levels of distress and reduced tiredness are due to the formation of an objective schema of stimulus properties and the habituation of distress), there is one feature of the data that is not fully consistent with this interpretation: The differences between conditions should have been greatest toward the 3rd and 5th minutes of the immersion period, rather than during the 1st minute, as shown in Figure 4. The data suggest that the conditions differentiated very rapidly, and this was reflected in the analysis, which found a main effect between conditions but no Treatment × Trials interaction. A second problem was the failure to detect differences between conditions in the measure of the percentage of time subjects reported that they attended to their hand. Indeed, control subjects reported having attended as much to their hand as did attention-instructed subjects. Postexperimental interviews suggested that this attention was not the same. For the control subjects, the attention seemed automatic, involuntary, and stimulus engendered; they could not ignore the distress of the stimulus impact. For the instructed subjects, attention seemed self-engendered, controlled, or voluntary, a deliberate focusing of awareness on stimulus properties. Indeed, a comparison of findings and impressions from the postexperimental interviews of Experiments 1 and 2 suggests that sensation information may be superior to direct attentional instructions, in reducing distress, since the former defines the stimulus features to be monitored and provides schemata to be used in coding the input (Bruner, 1957).

Experiment 3

Experiment 3 was designed to obtain better evidence with respect to the effects of the attentional instructions on the cues subjects monitor and to provide a better test of the hypothesis that schematization of the stimulus preceded distress reduction and habituation. Evidence on the deployment of attention was obtained by introducing a specific distractor to direct attention in the control condition. Evidence was obtained for the assumption that objective schema formation precedes reductions in distress by systematically varying the deployment of attention during the early and later parts of the cold water immersion. The strength of sensations and distress from the cold pressor stimulation begin at relatively low levels and gradually build in intensity. It should be relatively easy to direct attention to and schematize stimulus features at this time. As the stimulus becomes more intense, the schema should be reinforced by repetition of the features detected earlier. On the other hand, subjects instructed or motivated (by threat) not to attend to the stimulus may have difficulty deliberately directing their attention to stimulus features if they are first instructed to monitor the stimulus later in exposure. Delaying attention and permitting the stimulus to build to maximum strength, to be integrated with pain memories, and to be experienced as a distressing event will make it difficult to monitor and schematize, since the pain and distress may block attention to specific features.

The assumptions mentioned above are generally compatible with empirical observations made on studies of behavior and neural habituation in the cat (Groves & Thompson, 1970). Habituation occurs more rapidly to stimuli that increase gradually in strength early in a series of stimulus exposures, and habituation is relatively slow and sometimes fails to occur for strong stimuli. To test the assumption with cold pressor, a four-group experiment was designed varying attention to and distraction from the hand at different times during immersion in ice water. In one condition subjects were instructed to monitor the hand throughout the immersion period (attention–attention). In a second condition (attention–distraction) they were instructed to monitor the hand for the first half of the immersion period, during the time of increasing strength of sensations when schema formation is critical, and they were distracted by slides in the second half. In a third group (distraction–attention) subjects were instructed to attend to and monitor the hand for the last half of the immersion period after the input had presumably been classified and responded to as painful, thus blocking the positive effects of monitoring. A fourth group (distraction–distraction) was distracted for the entire immersion period. The hypothesis under test was that monitoring throughout the immersion period and monitoring during the first half of the immersion period would be equally effective for schema formation and would facilitate habituation during the second half of exposure. Monitoring during the last half of immersion and continual distraction should be equally ineffective for distress reduction.

Method

Subjects

Seventy-six male undergraduates at the University of Wisconsin served as subjects. They were selected from a larger pool of subjects, and before being permitted to participate, they were screened and excluded if they reported a health problem (e.g., a history of diabetes, rheumatic fever, heart abnormality, or high blood pressure) or if they were left-handed. An incentive of $2 per hour was offered to encourage participation. Nine subjects quit during exposure to the cold water treatment (they did not cluster in any particular condition), and one refused to participate when told about the noxious stimulus. Two subjects were dropped, one because he was hard of hearing and did not hear all of the instructions and another because he meditated during the experiment. The 64 subjects who remained were randomly assigned to one of the four conditions with the restriction that a subject was assigned to each of the four conditions before assignment was begun to another replication. This procedure was repeated eight times for each of the two male experimenters.

Distraction Materials

In postexperimental interviews for Experiment 2 subjects suggested that it was difficult to distract

themselves from their hand in the absence of an alternative stimulus. For this reason, 48 color slides of landscapes and people were used as distractors throughout the nonattention condition of Experiment 3, and 24 of the slides were used as distractors for the late and early nonattention (distraction) periods of the early and late attention conditions.

The 48 slides were selected from a larger set of 250. The experimenter first eliminated those slides suggesting pain or coldness (e.g., winter landscapes, scenes including water, etc.). The 60 slides remaining after this elimination were then shown to a group of six males and six females who had previously been exposed to the cold water stimulus and who rated them as to how much they thought the slide could distract them from the (cold pressor) pain. The 15 slides having the highest mean score (none scored less than $M = 4.75$ on a 7-point scale where 1 = attends to the pain and 7 = distracts from the pain) were selected as distraction stimuli and were used as standards for the selection of 33 additional distractors from yet another set of slides. The additional slides were matched to the first 15 in subject matter and amount of detail. We can best characterize these slides by saying that they were sufficiently interesting to attract attention and prevent deliberate monitoring of the hand, yet not so interesting as to provoke a strong alternative emotional response that would compete with the emotional distress evoked by the noxious stimulus.

Procedure

The procedure was virtually identical to that in Experiment 2. One major change was to raise the temperature of the laboratory to 26 °C and the temperature of the water used to establish a common baseline to 30 °C. It was hoped that these changes would facilitate recording skin temperature change.

The subjects in each of the four conditions went through the same seven steps in the following order: (a) completion of the medical screening questionnaire and reading and signing the consent form; (b) listening to an overall description of the aim and procedure of the study; (c) immersing the right hand in the warm (30 °C) water; (d) hearing an explanation of the distress rating scale and the rating procedure; (e) receiving attention and/or distraction instructions appropriate for the experimental condition; (f) removing the hand from the warm water and immersing it in the cold water (7 °C) for 5 min. and 15 sec; (g) removing the hand from the cold water tank and making judgments of distress at specified times during a 2-min. postimmersion period. The exact times for judgments and so forth are presented later on.

Attention instructions. The attention–distraction instructions for each condition were presented after the description of the distress rating procedure. Subjects assigned to the attention–attention condition heard the same attentional instruction as that given in Experiment 2. The instructions for the distraction–distraction group differed from those in Experiment 1, since slides were used as distractors in the current study. The instructions were as follows:

> During the time your hand will be in the cold water, you will see landscape and art slides. We would like you to pay attention to these slides, so you will be able to answer questions about them afterwards. This is the most important part of the study, so ignore everything else and pay close attention to the slides.

The instructions were modified somewhat when combined for the attention–distraction condition, and their order was reversed for the distraction–attention condition. The instructions were as follows:

> The time your hand will be in the colder water is divided into two parts. In the first part, we would like you to pay close attention to the sensation in your right hand so that you will be able to describe each of the specific feelings and sensations that you experienced. You will be asked many questions about the different sensations in your right hand afterwards. In this part try to ignore everything else and pay close attention to the sensation in your right hand.
>
> In the second part, you will see landscape and art slides. In this part we would like you to pay attention only to these slides, so you will be able to answer questions about them afterwards. Try to ignore everything else and pay close attention to the slides. It is very important that you follow the instructions, so pay close attention to the sensations in your right hand in the first part and try to ignore them and pay attention to the slides in the second part.

Timing of procedures. Tape-recorded instructions told the subject when to shift his hand from the warm to the cold water tank and also told him when to make the distress ratings. The subject made 12 judgments of distress during the 5 min. and 15 sec that his hand was immersed in the cold water; the distress scale was the same as that used in Experiments 1 and 2.

The immersion in the cold water was divided into two equal periods of 2 min. and 35 sec each, separated by a 5-sec interval in which the instructions appropriate for the last 2-min. and 35-sec period were repeated. Because it was necessary to repeat the instructions in the conditions in which subjects switched the focus of their attention, the instructions were also repeated in the attention–attention and distraction–distraction conditions. Six distress judgments were made in each 2-min. and 35-sec period. The first was requested 15 sec after the start of each period (15 sec after the instructions to immerse the hand in the water or 15 sec after completion of the between-periods instructions). The second judgment followed 15 sec later. A 30-sec interval was used between the remaining four judgments. To further en-

sure that subjects followed the appropriate procedure, two brief reminders were given during each period, one 45 sec after its start and the second a minute later.

Two different sets of 24 slides were used during each of the distraction periods. The slides were exposed for 5 sec each, and the between-slides interval equaled the time taken for the slide to change. The screen was darkened during the first of the three repetitions of the instructions and for 5 sec during each of the six ratings. The subject was told to remove his hand from the cold water and make four additional distress judgments at 30-sec intervals at the end of the second period.

Postexperimental questionnaires. After the 16 judgment trials, the thermistor was removed, and the subject completed two questionnaires. The first was the sensations and features checklist used in Experiment 2. The second questionnaire included modified versions of the direct questions asking for estimates of the deployment of attention and the questions about tiredness and tenseness. The attention questions had the subject estimate the percentage of time he attended to his hand, his body, the slides, and other features of the setting. An 11-point scale was used for each of the four foci of attention (in 10% intervals from 0% to 100%). The subject was told that his four estimates had to sum to 100%. The estimates were made twice, once for each of the two periods of the cold pressor.

Results

The data for the four conditions were analyzed as a 2 × 2 factorial design with a third repeated measure factor. The first factor was the attentional set (attention vs. distraction) during the first period of the experiment, the second factor was the attentional set (attention vs. distraction) during the second period, and the repeated measure factor was judgment trials.

As can be seen in Figure 5, distress showed a sharp initial increase for the first minute of immersion, peaked at approximately 1½ min into the procedure, and then declined, though differentially, in all groups. Figure 5 also shows that the distress levels declined considerably more in the attention–attention and attention–distraction conditions than in the distraction–distraction and distraction–attention

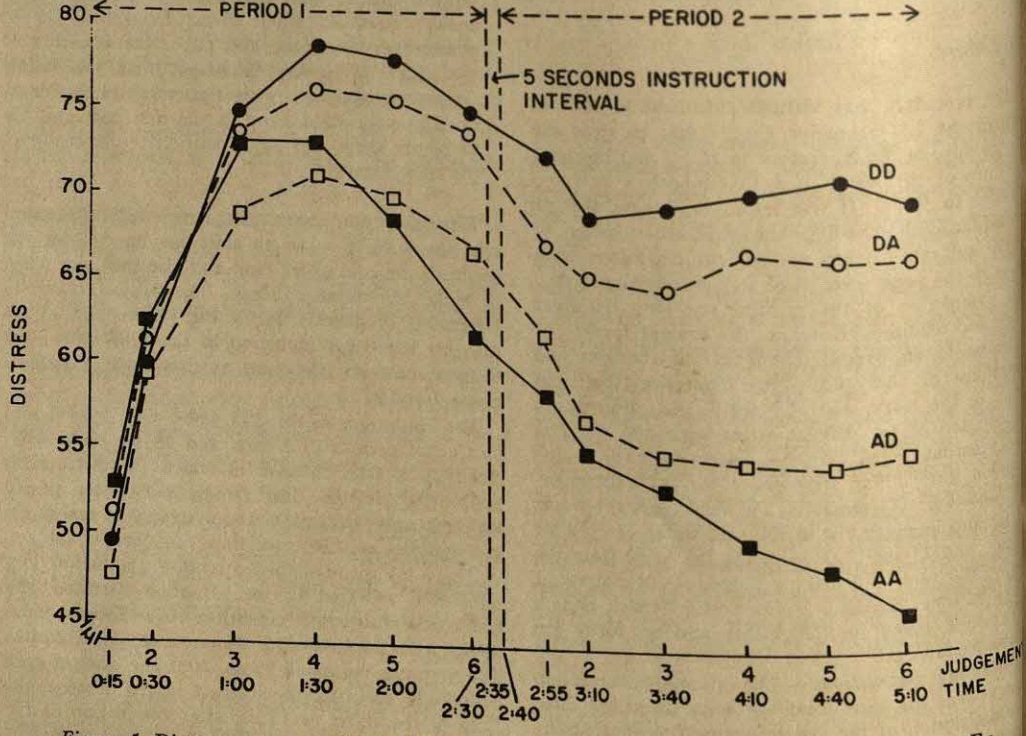


Figure 5. Distress ratings during cold pressor immersion for the four experimental groups in Experiment 3. (DD = distraction–distraction; DA = distraction–attention; AA = attention–attention; and AD = attention–distraction. $n = 16$ in each experimental group.)

conditions. The difference in rate of decline in reported distress resulted in a significant main effect, with the two groups paying attention to the hand in the first period reporting an average level of distress ($M = 59.04$) substantially lower than the average level of distress reported for the two groups that were distracted by slides during the first immersion period ($M = 68.18$), $F(1, 60) = 4.61$, $p < .05$. More important, however, there was an interaction between trials and the attentional set adopted during the first 2-min. and 35-sec period, $F(11, 660) = 5.15$, $p < .001$. The level of distress reported during the second period of cold pressor was substantially lower for the two groups attending to the hand during the first period than was the level of distress for the two groups watching slides during the first period. There were no significant effects as a function of attention or distraction during the second period of the cold pressor.

A pair of subsidiary analyses were conducted, one on the six ratings before the change in instructions from attention to distraction and another on the six ratings made after the change. The analysis of the first period showed an interaction between attentional set and trials, $F(5, 300) = 2.87$, $p < .05$, because the groups, although virtually identical for judgments 1–3, began to separate by the fourth judgment. The peak for the two early distraction groups ($M = 77.0$) was only slightly greater, however, than the peak for the two early attention groups ($M = 71.8$).

In the second set of six judgments, the drop in distress was significantly greater for the groups that were instructed to pay attention to the hand during the first period (the attention–attention and attention–distraction groups) than for the groups that were instructed to watch slides during the first period (the distraction–distraction and distraction–attention conditions), interaction $F(5, 300) = 2.47$, $p < .05$. The difference between the two pairs was sufficiently great to produce a significant main effect as well, $F(1, 60) = 8.46$, $p < .01$. The analyses support the pattern visible in Figure 5 and suggest that early attention to the hand began to affect distress reports $1\frac{1}{2}$ min. after immersion in the cold water and that these effects lasted and became still stronger in the second period of exposure. (Once again, trial effects were highly significant.)

Finger Temperature

Finger temperature dropped during the early part of immersion and rose during the postimmersion period for all conditions, but there were no treatment effects.

Postexperimental Questionnaire Reactions

Focus of attention. Subjects instructed to attend to their hand indicated that they spent a substantially higher proportion of their time monitoring their hand than did subjects not so instructed, in both the first (M attend = 76.5%, M distract = 50.3%), $F(1, 59) = 50.75$, $p < .001$, and the second period (M attend = 79.6%, M distract = 42.9%), $F(1, 59) = 76.05$, $p < .001$, of cold pressor. There were no significant differences for attention to body reactions or to other events in the setting.

The sensation and features checklist showed that *more* attention was paid to the hand (sum of eight items) during the first ($M = 73.7$) than during the second ($M = 58.7$) period of the cold pressor, $F(1, 59) = 17.12$, $p < .001$, whereas the sum of the 11 body sensation items showed less awareness of body responses in the first ($M = 51.7$) than in the second ($M = 66.2$) cold pressor period, $F(1, 59) = 21.97$, $p < .001$. These shifts in awareness from hand sensations in the first half of the cold pressor to body sensations in the second half held for all groups.

Feeling of fatigue. The same three-part question used in Experiment 2, "Toward the end of the time your hand was in the water, did you feel (a) tired, (b) tense and jumpy, (c) detached and daydreamy?" showed a significant effect for the tiredness rating (made on a 21-point scale); subjects who were instructed to attend to their hand in the first half of the cold pressor exposure reported being less tired ($M = 4.30$) than did subjects who were distracted by slides during the early phase ($M = 6.14$), $F(1, 58) = 4.77$, $p < .05$. There were no other significant effects.

Discussion

The results of Experiment 3 confirm and extend those of Experiment 2. First, attention to the site of cold pressor impact produced substantial reductions in reported distress. Second, it was clear that attention during the first portion of exposure was critical for later distress reduction; this finding held for reported distress and for reported fatigue on the postexperimental questionnaire. The introduction of the distractor slides facilitated the deliberate direction of attention and the ease with which subjects could report on its deployment, making clear that subjects gave a higher proportion of attention to the impact site during this early period. And the attentional effects seem to have begun very early, taking hold by the end of the fourth distress judgment or about 1½ min. into the cold pressor test.

Third, there were no signs of decline in reported distress or reported fatigue for subjects who monitored their hand during the second half of exposure; the reports of these subjects did not differ from those of subjects who were continually distracted with slides. It seems that attention during the latter part of exposure is relatively ineffective for distress reduction. While there are many reasons why this may be so, the present data and past observations point to one conclusion in particular. First, it is clear that distress is already at relatively high levels when attention is first deployed to the hand in the distraction–attention group. Second, it is clear that the experience of the cold pressor changes over time; the sensation checklist showed declining awareness of hand sensations and growing awareness of body sensations from the first to the second half of cold pressor immersion. Wolf and Hardy (1943) have detailed the many complex changes in the experience of cold pressor from the coldness and bright, pricking, pain sensations early in exposure to the deeper aching and numbness experienced later on, changes which have been attributed to activity in different fiber systems (Mountcastle, 1968). The earlier experiences are far more localized and informational in nature than the diffuse, unchanging aching sensations that begin to dominate the later pain experience. These latter sensations seem to be of the type that are more easily summated with other sources of input than are the clear, discrete, and localized sensations of early pain experience (Melzack, 1973). In conclusion, it would appear that subjects attending to the impact site in the latter half of exposure would have relatively less information available for the construction of a schema of the stimulus input and relatively more information (diffuse aching) for the formation of a generalized state of arousal and distress. The pattern of findings tends to confirm the hypothesis that monitoring during the onset of stimulation, when features are rich and informative and diffuse, aching distress is relatively low, is optimal for the formation of a stimulus schema and the habituation (or cessation of recruitment) of distress emotion (Sokolov, 1963).

General Discussion

All three of the experiments show that monitoring the impact of noxious stimulation can substantially reduce the emotional distress experience component of the experience on stimulus impact. The studies suggest that the mechanism underlying this effect is the formation of a schema of the objective, or informational, features of the stimulus, which captures new stimulus inputs and reduces the degree to which these new inputs can stimulate distress reactions. This habituation process is empirically similar to that reported in habituation studies in animals (Groves & Thompson, 1970), though the data from the present study point to a mechanism similar to that suggested by Sokolov (1963). It should be understood that the findings do not rule out alternative modes of distress reduction. Analgesic sets, under hypnotic (Hilgard, 1969, 1971) or nonhypnotic conditions (Barber & Hahn, 1962; Melzack, Weisz, & Sprague, 1963; Spanos, Horton, & Chaves, 1975), may also reduce distress to noxious inputs, but the empirical findings in these areas (immediate distress reduction, successful reduction for relatively short periods of time) suggest a different mechanism, namely, the blockage of sensations and distress from conscious awareness, rather than the habituation of distress

per se (Hilgard, 1973; Hilgard, Morgan, & Macdonald, 1975). And it should also be clear that monitoring stimulus inputs does not necessarily lead to distress reduction. If the subject is threatened and regards the stimulus sensations as warnings of further harm and danger, monitoring may enhance emotional distress rather than facilitate emotional habituation. Monitoring may be essential for the occurrence of habituation, but it is not sufficient.

Our interpretation of the cold pressor effects assumes, therefore, two quite different ways of coding or schematizing noxious stimulus inputs—an objective informational coding in which cues are coded for their concrete, immediate meaning and an emotional memory schematization in which cues are coded for their anticipatory, threat value (Engel, 1959; Leventhal, in press; Leventhal & Everhart, in press). The relative strength of these two types of coding will obviously vary as a function of set and subject disposition. The theoretical interpretation is the same as that offered several years ago for the "new look" studies of perception (Bruner, 1957) and in more recent theorizing respecting the role of categorization in awareness (Broadbent, 1977).

While the approach is not novel, it does allow the synthesis of a wide range of data. For example, Morgan and Pollock (1977) studied world-class runners and found that elite runners carefully monitor leg and muscular sensations while performing. Outstanding, but nonelite runners distract from these sensations; they regard them as threat cues or signs of an anticipated wall or limit to their endurance. Thus, monitoring per se is not the critical factor in the runner's control of distress; it is the schematization of the cues that is central. Similarly, cases of phantom pain appear to involve a pain schema of body parts and a vivid central neural representation of the somesthetic and distress reactions associated with preoperative injuries to amputated body parts (Melzack, 1971; Morgenstern, 1970; Simmel, 1962). Case examples illustrate that these painful experiences may vanish if the patient is exposed, by means of epinephrine injection to nerve roots, to a rich panoply of concrete sensations from the missing limb without pain inputs; monitoring the fresh, nonpainful set of stimulus features allows the reassertion of the neutral, nonpain schema of the missing body part.

It also seems that under some conditions, emotional schemata (pain memories) are more available than so-called objective codes. Indeed, the availability of these schemata in real-life situations is very likely responsible for the differences in experimental findings between laboratory and clinical studies of pain (Beecher, 1959). We have been impressed with the reliability with which one can find distress reduction by using monitoring strategies in hospital settings. The patient is in a threatening situation, pain and emotional schematization are highly likely, and distress reduction is readily achieved by combinations of monitoring strategies (sensation information) and reassurance that the procedures will help (Johnson et al., 1975; Johnson & Leventhal, 1974; Johnson et al., 1973; Johnson et al., Note 1; Fuller, Endress, & Johnson, Note 2; Wilson, Note 3).

The field studies also make clear that coping or control is an important factor in the distress-reduction process. The availability of a coping reaction may not only substitute instrumental responding for emotional responding but will also help sustain attention to the objective features of repeated, noxious stimulus inputs, facilitate placing benign interpretations on these inputs, and permit habituation (see Johnson & Leventhal, 1974; Johnson et al., Note 1). For example, a postsurgical cardiac by-pass patient must engage in stretching and breathing exercises that will generate intense sensations and distress from the chest. Knowing what the sensations will be, carefully monitoring them as he or she breathes, and interpreting them as signs of healing and return to normal function will likely encourage the patient's objective schematization of the cues and the habituation of distress reactions. By contrast, an unprepared patient who does not expect the signs and has no way of interpreting them as benign might inhibit coping reactions in order to minimize and distract him- or herself from pain and distress and fears of potential injury.

A complete model of the stress control process will involve, therefore, a series of steps going from the representation of stimulus features to their coding in objective or emotional memory schemata, the execution of coping reactions guided by these cues, and the interpretation or evaluation of feedback from the coping behaviors (Lazarus, 1966). The three experiments reported here addressed only part of the problem. It is also clear that we have merely touched on the process of distress habituation. Habituation may be a passive dropping out of responding (e.g., Groves & Thompson, 1970) or a product of one or more active, inhibitory processes related to gating of noxious stimulus impulses by informational factors (Melzack & Wall, 1965, 1970) or by inhibition generated by endorphins in the medial forebrain system (Snyder, 1977). Interesting questions remain respecting the mechanisms underlying the behavioral effects reported here, and these mechanisms may suggest new psychological manipulations for pain control.

Reference Notes

1. Johnson, J. E., Rice, V. H., Fuller, S. S., & Endress, M. P. *Sensory information, behavioral instruction, and recovery from surgery.* Paper presented at the 85th annual convention of the American Psychological Association, San Francisco, August 1977.
2. Fuller, S. S., Endress, M. P., & Johnson, J. E. *Control and coping with an aversive health examination.* Paper presented at the 85th annual convention of the American Psychological Association, San Francisco, August 1977.
3. Wilson, J. F. *Coping styles influencing the effectiveness of preoperative intervention procedure.* Paper presented at the 85th annual convention of the American Psychological Association, San Francisco, August 1977.

References

Barber, J. X., & Hahn, K. W. Physiological and subjective responses to pain producing stimulation under hypnotically-suggested and waking-imagined "analgesia." *Journal of Abnormal and Social Psychology,* 1962, *65,* 411–418.

Beecher, H. K. *Measurement of subjective responses.* New York: Oxford University Press, 1959.

Blitz, B., & Dinnerstein, A. J. Role of attentional focus in pain perception: Manipulation of response to noxious stimulation by instructions. *Journal of Abnormal Psychology,* 1971, *77,* 42–45.

Brady, J. V., Porter, R. W., Conrad, D. G., & Mason, J. W. Avoidance behavior and the development of gastroduodenal ulcers. *Journal of the Experimental Analysis of Behavior,* 1958, *1,* 69–72.

Broadbent, D. E. The hidden preattentive processes. *American Psychologist,* 1977, *32,* 109–118.

Bruner, J. S. On perceptual readiness. *Psychological Review,* 1957, *64,* 123–152.

Calvert-Boyanowsky, J., & Leventhal, H. The role of information in attenuating behavioral responses to stress: A reinterpretation of the misattribution phenomenon. *Journal of Personality and Social Psychology,* 1975, *32,* 214–221.

Cohen, F., & Lazarus, R. S. Active coping processes, coping dispositions, and recovery from surgery. *Psychosomatic Medicine,* 1973, *5,* 375–389.

Cupchik, G., & Leventhal, H. Individual differences in the objective and subjective information about cartoon stimuli. *Journal of Personality and Social Psychology,* 1974, *30,* 429–442.

Egbert, L. D., Battit, G. E., Welch, C. E., & Bartlett, M. K. Reduction of postoperative pain by encouragement and instruction of patients. *New England Journal of Medicine,* 1964, *270,* 825–827.

Engel, G. L. Psychogenic pain and the pain-prone patient. *American Journal of Medicine,* 1959, *26,* 899–918.

Epstein, S. Expectancy and magnitude of reaction to a noxious UCS. *Psychophysiology,* 1973, *10,* 100–107.

Epstein, S. & Clarke, S. Heart rate and skin conductance during experimentally induced anxiety: Effects of anticipated intensity of noxious stimulation and experience. *Journal of Experimental Psychology,* 1970, *84,* 105–112.

Geer, J. H., Davison, G. C., & Gatchel, R. I. Reduction of stress in humans through nonveridical perceived control of aversive stimulation. *Journal of Personality and Social Psychology,* 1970, *16,* 731–738.

Glass, D. C., & Singer, J. E. Behavioral aftereffects of unpredictable and uncontrollable aversive events. *American Scientist,* 1972, *60,* 457–465.

Groves, P. M., & Thompson, R. F. Habituation: A dual-process theory. *Psychological Review,* 1970, *77,* 419–450.

Hall, K. R. L., & Stride, E. The varying response to pain in psychiatric disorders: A study in abnormal psychology. *British Journal of Medical Psychology,* 1954, *27,* 48–60.

Hilgard, E. R. Pain as a puzzle for psychology and physiology. *American Psychologist,* 1969, *24,* 103–113.

Hilgard, E. R. Hypnotic phenomena: The struggle for scientific acceptance. *American Scientist,* 1971, *59,* 567–577.

Hilgard, E. R. A neodissociation interpretation of pain reduction in hypnosis. *Psychological Review,* 1973, *80,* 396–411.

Hilgard, E. R., Morgan, A. H., & Macdonald, H. Pain and dissociation in the cold pressor test: A study of hypnotic analgesia with "hidden reports"

through automatic key pressing and automatic talking. *Journal of Abnormal Psychology,* 1975, *84,* 280–289.

Holmes, D. S., & Houston, B. K. Effectiveness of situation redefinition and affective isolation in coping with stress. *Journal of Personality and Social Psychology,* 1974, *29,* 212–218.

Janis, I. L. *Psychological stress.* New York: Wiley, 1958.

Janis, I. L. Effects of fear arousal on attitude change: Recent developments in theory and experimental research. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 3). New York: Academic Press, 1967.

Johnson, J. E. Effects of accurate expectations about sensations on the sensory and distress components of pain. *Journal of Personality and Social Psychology,* 1973, *27,* 261–275.

Johnson, J. E. Stress reduction through sensation information. In I. G. Sarason & C. D. Spielberger (Eds.), *Stress and anxiety* (Vol. 2). Washington, D.C.: Hemisphere, 1975.

Johnson, J. E., Kirchoff, K. T., & Endress, M. P. Deterring children's distress behavior during orthopedic cast removal. *Nursing Research,* 1975, *75,* 404–410.

Johnson, J. E., & Leventhal, H. Effects of accurate expectations and behavioral instructions on reactions during a noxious medical examination. *Journal of Personality and Social Psychology,* 1974, *29,* 710–718.

Johnson, J. E., Leventhal, H., & Dabbs, J. M., Jr. Contribution of emotional and instrumental response processes in adaptation to surgery. *Journal of Personality and Social Psychology,* 1971, *20,* 55–64.

Johnson, J. E., Morrissey, J. F., & Leventhal, H. Psychological preparation for an endoscopic examination. *Gastrointestinal Endoscopy,* 1973, *19,* 180–182.

Kanfer, F. H., & Goldfoot, D. A. Self-control and tolerance of noxious stimulation. *Psychological Reports,* 1966, *18,* 79–85.

Krueger, F. E. The essence of feeling. In M. Arnold (Ed.), *The nature of emotion: Selected readings.* Baltimore Md.: Penguin Books, 1928.

Lazarus, R. *Psychological stress and the coping process.* New York: McGraw-Hill, 1966.

Leventhal, H. Findings and theory in the study of fear communications. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 5). New York: Academic Press, 1970.

Leventhal, H. Emotions: A basic problem for social psychology. In C. Nemeth (Ed.), *Social psychology: Classic and contemporary integrations.* Chicago: Rand McNally, 1974.

Leventhal, H. The consequences of depersonalization during illness and treatment. In J. Howard & A. Strauss (Eds.), *Humanizing health care.* New York: Wiley, 1975.

Leventhal, H. A perceptual–motor theory of emotion. In P. Pliner, K. R. Blankstein, & I. M. Spigel (Eds.), *Advances in the study of communication and affect: Vol. 5. Perception of emotion in self and others.* New York: Plenum Press, in press.

Leventhal, H., & Cupchik, G. C. A process model of humor judgment. *Journal of Communication,* 1976, *26,* 190–204.

Leventhal, H., & Everhart, D. Emotion, pain, and physical illness. In C. E. Izard (Ed.), *Emotion and psychopathology.* New York: Pergamon Press, in press.

Melzack, R. Phantom limb pain: Implications for treatment of pathologic pain. *Anesthesiology,* 1971, *35,* 409–419.

Melzack, R. *The puzzle of pain.* New York: Basic Books, 1973.

Melzack, R., & Wall, P. D. Pain mechanism: A new theory. *Science,* 1965, *150,* 971–979.

Melzack, R., & Wall, P. D. Psychophysiology of pain. *International Anesthesiology Clinics,* 1970, *8,* 3–34.

Melzack, R., Weisz, A. Z., & Sprague, L. T. Strategies for controlling pain: Contribution of auditory stimulation and suggestion. *Experimental Neurology,* 1963, *8,* 293–347.

Morgan, W. P., & Pollock, M. L. Psychologic characterization of the elite distance runner. In P. Milvy (Ed.), *The marathon: Physiological, medical, epidemiological, and psychological studies.* New York: New York Academy of Sciences, 1977.

Morgenstern, F. S. Chronic pain. In O. U. Hill (Ed.), *Modern trends in psychosomatic medicine* (Vol. 2). New York: Appleton-Century-Crofts, 1970.

Mountcastle, V. B. Pain and temperature sensibilities. In V. B. Mountcastle (Ed.), *Medical physiology* (Vol. 2). Saint Louis, Mo.: C. V. Mosby, 1968.

Neufeld, R. W. J., & Davidson, P. O. The effects of vicarious and cognitive rehearsal on pain tolerance. *Journal of Psychosomatic Research,* 1971, *15,* 329–335.

Nisbett, R. E., & Schachter, S. Cognitive manipulation of pain. *Journal of Experimental Social Psychology,* 1966, *2,* 227–236.

Pervin, L. A. The need to predict and control under conditions of threat. *Journal of Personality,* 1963, *31,* 570–587.

Ross, L., Rodin, J., & Zimbardo, P. G. Toward an attribution therapy: The reduction of fear through induced cognitive–emotional misattribution. *Journal of Personality and Social Psychology,* 1969, *12,* 279–288.

Schachter, S. The interaction of cognitive and physiological determinants of emotional state. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 1). New York: Academic Press, 1964.

Schachter, S., & Singer, J. E. Cognitive, social, and physiological determinants of emotional state. *Psychological Review,* 1962, *69,* 379–399.

Sime, A. M. Relationship of preoperative fear, type of coping, and information received about surgery to recovery from surgery. *Journal of Personality and Social Psychology,* 1976, *34,* 716–724.

Simmel, M. L. The reality of phantom sensations. *Social Research,* 1962, *29,* 337–356.

Snyder, S. H. Opiate receptors and internal opiates. *Scientific American,* 1977, *236,* 44–56.

Sokolov, E. N. *Perception and the conditioned reflex.* New York: Macmillan, 1963.

Spanos, N. P., Horton, C., & Chaves, J. F. The effects of two cognitive strategies on pain threshold. *Journal of Abnormal Psychology,* 1975, *84,* 677–681.

Staub, E., & Kellet, O. Increasing pain tolerance by information about aversive stimulation. *Journal of Personality and Social Psychology,* 1972, *21,* 198–203.

Staub, E., Tursky, B., & Schwartz, G. E. Self-control and predictability: Their effects on reactions to aversive stimulation. *Journal of Personality and Social Psychology,* 1971, *18,* 157–162.

Teichner, W. H. Delayed cold-induced vasodilation and behavior. *Journal of Experimental Psychology,* 1965, *19,* 426–432.

Weiss, J. M. Effects of coping behavior with and without a feedback signal on stress pathology in rats. *Journal of Comparative and Physiological Psychology,* 1971, *77,* 22–30.

Wolf, S. G., & Hardy, J. D. Studies on pain: Observations on pain due to local cooling and on factors involved in the "cold pressor" effect. *Research Publications of the Association for Research on Nervous and Mental Disease,* 1943, *23,* 123–142.

Wolfer, J. A., & Davis, C. C. Assessment of surgical patients' preoperative emotional conditions and postoperative welfare. *Nursing Research,* 1970, *19,* 402–414.

Received December 22, 1977 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 5, 715-727

Effects of Pitch and Speech Rate on Personal Attributions

William Apple
Columbia University

Lynn A. Streeter
Bell Laboratories
Murray Hill, New Jersey

Robert M. Krauss
Columbia University

In three experiments, subjects listened to recordings of male speakers answering two interview questions and rated the speakers on a variety of scales. The recordings had been altered so that the pitch of the speakers' voices was raised or lowered by 20% or left at its normal level, and speech rate was expanded or compressed by 30% or left at its normal rate. The results provided clear evidence that listeners use these acoustic properties in making personal attributions to speakers. Speakers with high-pitched voices were judged less truthful, less emphatic, less "potent" (smaller, thinner, faster), and more nervous. Slow-talking speakers were judged less truthful, less fluent, and less persuasive and were seen as more "passive" (slower, colder, passive, weaker) but more "potent." However, the effects of the acoustic manipulations on personal attributions also depended on the particular question that elicited the response.

Human speech provides a listener with at least two sources of information: a *verbal* channel, encoding the message's linguistic content, and a *vocal* channel, conveying paralinguistic information by variations in pitch, speech rate, loudness, and the like.

One important type of information communicated via the vocal channel concerns a speaker's affective state. The vocal characteristics associated with the expression of emotion are beginning to be understood (see, for example, Fairbanks, 1940; Hecker, Stevens, von Bismarck, & Williams, 1968; Williams & Stevens, 1972). It is now reasonably well established that stressful situations raise the voice's fundamental frequency (the number of glottal pulses per second) and that "active" emotions such as anger and fear tend to be reflected in increased mean pitch and pitch variance, whereas "low energy" states such as sorrow and indifference are associated with a lower mean pitch and a slower speech rate.

Given that emotional states do differ reliably in their paralinguistic expression, to what extent do listeners use these vocal cues in judging the immediate affective state or more enduring personality traits of a speaker? An early investigation of the noncontent aspects of speech by Allport and Cantril (1934) demonstrated that listeners could judge, at better than chance levels, a speaker's age and at least some personality characteristics from voice alone: In four of six experiments, speakers' scores on a test of ascendance–submission (Allport's A–S reaction study) were judged with significant accuracy. However, reviewing much of the voice-attribution work, Kramer (1963) concluded that more common than accuracy in such judgment studies was the finding of "vocal stereotypes." That is, certain voices were reliably, though sometimes incorrectly, judged as belonging to certain personality types.

Unfortunately, as Kramer noted, the description of the vocal parameters that under-

We wish to thank John Lin and Nina Macdonald for their valuable assistance. We are also grateful to Myron Wish, Nancy Morency, and four anonymous reviewers for their helpful comments.

Requests for reprints should be sent to Robert M. Krauss, Department of Psychology, 400 Schermerhorn Hall, Columbia University, New York, New York 10027.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3705-0715$00.75

lie such stereotypes leaves much to be desired. While not all the earlier research on voice attributions can be faulted on these grounds, studies attempting to specify critical stimulus dimensions often have neglected or confounded important paralinguistic cues. For example, in reporting two recent field experiments, Miller, Maruyama, Beaber, and Valone (1976) concluded that the persuasiveness of a communication can be directly related to the rate at which it is delivered. More rapid speech was found to be more persuasive, presumably because a fast talker is viewed as more credible. Although Miller et al. have ruled out certain alternative explanations (such as the effect of having limited opportunity to counterargue against a rapid presentation), the stimulus materials used in their experiments warrant closer examination. In both experiments the persuasive messages were recorded by the same speaker at either a slow or a fast speech rate. This was accomplished "by simply instructing the speaker to practice delivering the same speech as rapidly and slowly as possible while controlling his level of enthusiasm and involvement" (Miller et al, 1976, p. 618). The stimulus recordings in the two conditions do, of course, differ in speech rate, but it is quite likely that they differ in other respects as well. In natural speech, such vocal parameters as amplitude, pitch, and rate tend to covary. For example, rapid speech is likely to be louder and higher pitched than normal speech (Black, 1961). Consequently, it is quite possible that subjects in the Miller et al. study were responding to pitch and/or loudness cues as well as to rate.

Because it is so difficult to assess, in a controlled way, the contribution of various vocal parameters to the attribution process using natural speech, a number of workers have attempted to deal with the problem by using nonspeech stimuli. For example, Scherer (1974) presented listeners with simple tone sequences generated on a Moog synthesizer, in which a minimal set of acoustic cues (pitch level and variation, amplitude level and variation, and tempo) were varied factorially; listeners rated the emotional quality of these tone sequences on a set of semantic differential scales. Scherer found that judgments were most influenced by tempo and pitch variations: Fast tempo led to attributions of highly active and potent emotions (e.g., interest, anger, and happiness) and slow tempo to attributions of sadness, disgust, and boredom; extreme pitch variation with rising contours produced ratings of highly pleasant, active, and potent emotions (happiness, surprise, interest). Other parameters influenced the perception of specific emotional qualities. While Scherer's method effectively deals with the confounding of vocal qualities in natural speech and suggests the effect that independently varied acoustic cues can exert, in view of the artificial nature of the stimuli used, it is not clear how directly the results can be generalized to the processing of speech.

Other investigators, using natural speech, have achieved a degree of control over stimulus materials by means of editing techniques. For example, investigators have added or removed such speech disfluencies as filled pauses and repetitions from stimulus audiotapes and asked listeners to judge speakers' credibility and other personal attributes (Lay & Burron, 1968; Miller & Hewgill, 1964). Typically, highly hesitant or disfluent speakers are assigned relatively undesirable personality traits, and their communications are judged to be low in credibility.

While the methods thus far described have all afforded some degree of stimulus control, recent developments in speech synthesis technology permit investigators to vary independently one or more parameters of natural speech. This approach has been exploited by Brown and his colleagues (Brown, Strong, & Rencher, 1973, 1974; Smith, Brown, Strong, & Rencher, 1975), who have manipulated speech rate, mean fundamental frequency, and fundamental frequency variance, using a computer-based analysis–synthesis system. Two major personality dimensions (termed "competence" and "benevolence") have emerged from factor analyses of judgments of such manipulated speech. Generally speaking, higher pitch seems to result in a speaker being judged less competent and less benevolent (Brown et al., 1974), whereas faster speech rates produce judgments of higher

competence but yielded an inverted-U relationship on the benevolence dimension (Smith et al., 1975).

The studies of Brown and his associates make an important contribution to our understanding of the significance that listeners ascribe to vocal qualities. Nevertheless, they are not free of methodological problems. They typically have relied on a small number of stimulus voices, all of whose acoustic permutations are presented to the same raters. For example, the Brown et al. (1974) study used only two adult male speakers uttering the same sentence ("We were away a year ago"), which was then manipulated to fill the 27 cells of a $3 \times 3 \times 3$ factorial design. It is not clear that judgments based on hearing 54 repetitions of the same content bear much similarity to the kinds of everyday personality ascriptions listeners make under more natural conditions. Additionally, the particular parametric values used in these studies are problematic. Brown et al. (1974) reported that in the high-pitch condition, the original fundamental frequency was multiplied by a factor of 1.8. Although they did not report average fundamental frequency values for their two speakers, assuming that they fell into the normal range for males, Brown et al.'s high-pitch manipulation would have raised a male voice into the female range. Similar criticisms apply to their choice of speech-rate scale factors.

The present article reports an exploratory series of experiments designed to demonstrate the effects of two acoustic parameters—average fundamental frequency and speech rate—on judgments of several state and trait variables. It was hoped that by using a large number of speakers and naturally produced utterances, together with more conservative values for acoustic alteration, and by independently manipulating the two parameters factorially, we could overcome the methodological problems noted above.

Experiment 1

Experiment 1 derives from a finding reported by Streeter, Krauss, Geller, Olson, and Apple (1977), who demonstrated that the mean fundamental frequency of speakers' voices increased when they lied, relative to a truth-telling baseline, and that this effect was stronger when the speaker had been motivated to lie effectively. In addition, Streeter et al. found that although judges' ratings of truthfulness were essentially uncorrelated with pitch level, there *was* a significant negative correlation between judged truthfulness and average pitch for listeners who heard the speech after it had been passed through a content filter—a device that destroys intelligibility without affecting vocal features of the utterances. Apparently, the natural pitch increments during lying (on the average about 3 Hz) were too small to affect the judgments of listeners who had content available but were taken into account by listeners who could not understand the responses' verbal content. However, since pitch increments and speech rate decrements were correlated in their data, Streeter et al. could not rule out the possibility that listeners were attending to rate differences rather than pitch differences between true and false utterances. From the finding of Miller et al. (1976), one would expect slower speech to result in lower perceived speaker credibility. In Experiment 1 we assessed the effects of these two variables on truthfulness judgments.

Method

Stimulus materials. Forty male Columbia College undergraduates, all native speakers of English, answered questions for use as stimulus materials. They were individually recorded in a sound-isolated booth on high-quality audio equipment, and they received course credit for their participation. Each speaker answered six standard questions dealing with his opinion on a range of topics. The questions were administered in a fixed order, and speakers were instructed to answer all questions honestly and frankly. They were told to give brief answers, but to give more than a yes-or-no response. All speakers were debriefed concerning how their interviews might be used and signed informed-consent releases permitting the later use of the recordings.

Two of the six interview questions were selected for use in the experiment. One asked the subject's opinion of college admissions quotas designed to favor minority groups (Question 3); the other asked the subject what he would do if he suddenly won or inherited a large sum of money (Question 6). These two questions were chosen because both elicited a diversity of responses and because the two differed

on a dimension of personal salience for the respondents. (The question of quotas is a matter of genuine concern and frequent discussion among Columbia undergraduates.) Twenty-seven speakers, who gave responses that were not excessively long and represented the spectrum of opinions on the two questions, were selected.

Each of the 27 speakers was then randomly assigned to one of nine cells of a 3 (rate: slow, unmanipulated, fast) × 3 (pitch: low, unmanipulated, high) completely crossed factorial design, with three speakers in each cell. All speech material was digitized and analyzed by the linear predictive coding (LPC) method of Atal and Hanauer (1971). (The LPC analysis calculates 14 parameters every 10 msec. Twelve of these parameters represent pseudoarea functions of the vocal tract; the other two are values for amplitude and fundamental frequency. The advantage of LPC analysis is that such variables as rate, pitch, and amplitude can be manipulated without changing other voice parameters. In addition, since the parameters are derived from the original speech, the quality of the resulting synthesis is high.) The two responses of each speaker were manipulated on a DDP-224 computer using an interactive program (Nakatani, Note 1) that displays the parameters as functions of time and allows the user to manipulate any or all of the 14 parameters as well as to linearly expand or compress the time base. The altered utterances can then be synthesized and recorded on audiotape.

The scale factors chosen for the low- and high-pitch manipulations were 80% and 120%, respectively, of the speaker's unmanipulated fundamental frequency. The values chosen for the slow and fast speech rate manipulations were 70% compression and 130% expansion, respectively, of the utterance's time base, resulting in speech rates that were 77% and 143% of the unmanipulated rates.[1] These scale values were chosen because the resulting speech still sounded natural and the acoustic properties remained more or less within the normal range of values. (See Hanley, 1951; Mysak, 1959; Peterson & Barney, 1952; and Terango, 1966, for data on pitch and Goldman Eisler, 1968, for normative speech rate data.) The mean premanipulation fundamental frequency and speech rate for our speakers were 109.79 Hz ($SD = 14.82$) and 3.27 syllables/sec ($SD = .66$). Following manipulation, all utterances were resynthesized to produce a stimulus audiotape of the answers to the quota question (in one random order) and a tape of the answers to the money question (in a different order). The tapes were formatted to allow 15 sec of silence between each response.

Procedure. Twenty undergraduates, 14 males and 6 females, were paid for their participation.[2] Raters were run in groups of three to seven. They were told that the purpose of the study was to determine how well people can tell, from the sound of the speaker's voice, whether someone is lying or telling the truth. They were informed that approximately half the speakers had been instructed to tell the truth, whereas the remaining half had been told to lie—that is, to give answers that did not correspond to their actual beliefs or feelings.

The two stimulus tapes were played over high-quality audio equipment with the order of tape presentation counterbalanced. A 500-Hz warning tone preceded and followed each answer by 1 sec. Raters were given 15 sec in which to rate the truthfulness of the answer on a 7-point scale ranging from "not at all truthful" (1) to "entirely truthful" (7).

Because of variations in the quality of the recordings, raters were also told that recordings had been made using a variety of equipment in several different environments.[3] The experimental session lasted about an hour, and there was a 5-min. rest period between tapes. Participants in the three rating studies were sent a report describing the purpose and detailing some of the findings of the study several weeks after its conclusion.

Results

Some of the premanipulation characteristics of the stimulus materials are summarized in Table 1. Note that the more involving topic, college admissions quotas, resulted in significantly longer, slower, and higher pitched responses. Since the quota topic always preceded the money topic in the original order, we cannot rule out the possibility that these differences are due to serial position effects.

---

[1] The rate manipulation algorithm linearly expanded or compressed consonant and vowel information alike. In natural speech, rate variation tends to be reflected more in changes of vowel duration than in consonant duration. However, the rate manipulation used resulted in reasonably natural-sounding speech.

It may not be immediately apparent why 70% compression of an utterance's time base results in a speech rate that is 143% of the original. Consider a response consisting of n syllables spoken over s seconds of time; that utterance's rate would be n/s syllables/sec. Seventy-percent compression changes the rate to $n/(.7s)$, or $1.43\ n/s$—a faster speech rate that is 143% of the original. Likewise, a 130% expansion of an utterance's time base results in a slower speech rate that is 77% of the original rate.

[2] One additional rater was run, but his data were discarded after he expressed some suspicion regarding possible splicing of the tapes. No other raters voiced any suspicion that the stimuli had been altered.

[3] There is some quality variation across speakers in the LPC synthesis; in particular, speakers who tend to mumble and have a high degree of nasalization seem to suffer the greatest quality degradation.

Table 1
Premanipulation Characteristics of Stimulus Tapes

	Quota tape		Money tape		
	M	*SD*	*M*	*SD*	*t*
Response length (syllables)	111.1	40.7	84.6	42.7	3.89**
Response time (sec)	36.8	13.8	25.6	13.2	4.68**
Speech rate (syllables/sec)	3.0	.50	3.4	.74	2.12*
Fundamental frequency (Hz)	112.2	14.9	107.4	14.6	5.60***

Note. $N = 27$ segments per tape.
* $p < .05$, two-tailed. ** $p < .01$, two-tailed. *** $p < .001$, two-tailed.

However, it seems more plausible to regard these differences as reflections of differences in our subjects' personal involvement with the two questions and/or the cognitive complexity of the answers they called for (see Goldman Eisler, 1968, and Williams & Stevens, 1969).

To check the adequacy of the random assignment, premanipulation speech rate (syllables/sec) and average pitch were subjected to analyses of variance (3 rate levels × 3 pitch levels × 2 questions) with speakers nested within the rate and pitch factors. In addition to the between-questions differences already noted, the analysis of the premanipulation pitch failed to reveal other significant effects. However, there was a marginally significant difference in premanipulation speech rate across the three assigned rate conditions, $F(2, 18) = 3.00$, $p < .08$, primarily because of some slower speakers' random assignment to the slow rate condition. Therefore, prior to all further analyses, we adjusted the raw data for covariation on speakers' premanipulation speech rate and fundamental frequency: The appropriate beta weights were derived from linear regression of the data collapsed across raters. This adjustment has the virtue of controlling for spurious rate and question effects in the analyses of variance.[4] Having thus transformed each dependent variable, we computed min-F ratios (and their approximate degrees of freedom) for the analyses of variance. The F'' statistic we report (Winer, 1971, pp. 375–378) treats both speakers and raters as random effects, permitting simultaneous generalization over both groups.[5]

A 3 (pitch levels) × 3 (rate levels) × 2 (questions) analysis of variance was performed on the mean truthfulness ratings with repeated measures on the last factor. Premanipulation speech rate and premanipulation fundamental frequency were used as covariates, and all means reported below have been adjusted for these covariates. A significant main effect was found for the pitch manipulation, $F''(3, 34) = 3.37$, $p < .05$, and a marginally significant effect for rate, $F''(3, 51) = 2.26$, $p < .10$. The mean truthfulness ratings for the three pitch conditions (going from low to high pitch) were 4.62, 4.40, and 4.09, respectively, indicating that lower pitch enhanced credibility. The corresponding means for the rate manipulation were, going from slow to fast rate, 4.10, 4.63, and 4.37; the unmanipulated rate was judged most credible and the slow rate least credible.

No significant effect was found for the

[4] To ensure that our subjects' ratings were not affected by variations in acoustic quality across our nine experimental conditions, we had six undergraduates rate the 54 recorded segments for intelligibility and correlated these ratings with our subjects' ratings of truthfulness. The two sets of ratings were not significantly correlated ($r = .15$). Similarly, to ensure that response content was well distributed across conditions, we had 12 undergraduates rate from a transcript how pro- or antiquota (for the quota question) or generous or selfish (for the money question) each response was. Their ratings were then subjected to a 3 (pitch levels) × 3 (rate levels) analysis of variance. For neither question were significant main effects or interactions found.

[5] In cases where the min-F value (F'') is marginal, we will also report the conventional F ratios for raters (considering speakers as a fixed effect) and speakers (considering raters as a fixed effect). These values represent less conservative tests.

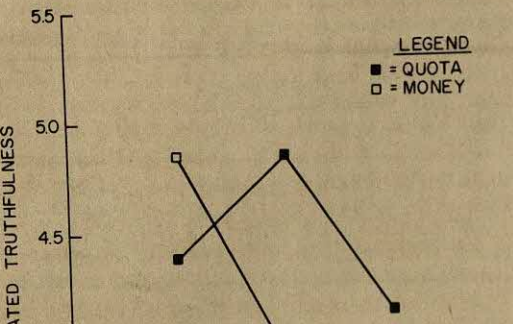


Figure 1. Average rated truthfulness plotted as a function of pitch condition (low, normal, high) for each of the two question topics.

Pitch × Rate interaction. However, there was a nearly significant Pitch × Question effect, $F''(2, 32) = 3.02$, $p < .10$; raters' $F(2, 38) = 10.25$, $p < .001$; speakers' $F(2, 18) = 3.84$, $p < .05$. For the quota question, truthfulness ratings and pitch were curvilinearly related; for the money question, low-pitched speakers were judged as most truthful. The two sets of means are shown in Figure 1.[6]

Discussion

The results demonstrate that the acoustic manipulations performed on the speech stimuli affected judgments of truthfulness. Consistent with the findings of Streeter et al. (1977), judges rated high-pitched voices as less truthful than lower pitched voices. Perhaps listeners perceived high pitch to be an indication of stress and attributed such stress to attempted deception. The pitch manipulations used, although not extreme enough to place voices outside the normal male pitch range, were evidently large enough to produce attributions of lying from naive listeners. It will be recalled that the smaller, naturally occurring pitch increments accompanying deception in the Streeter et al. experiment did not evoke such attributions, except in the filtered listening condition, in which verbal content was unintelligible.

While the rate effect was only marginally significant, it appears that the effect of rate on truthfulness judgments is not linear; rather, the pattern is an inverted-U function, with slower speech perceived as least credible. These results are consistent with the findings of Miller et al. (1976), who demonstrated that a faster speaker was perceived as more intelligent, knowledgeable, and objective than a slower speaker. Since Miller et al. used only two levels of speech rate, it is, of course, not possible to establish the effect of intermediate rates on credibility judgments in their study. It is relevant, however, that Brown and coworkers have reported a similar inverted-U relationship between manipulated speech rate and their "benevolence" dimension, on which the adjective pair sincere–insincere loads significantly (see Smith, Brown, Strong & Rencher, 1975). To the extent that ratings of sincerity correspond to this study's truthfulness measure, the two results are consistent.

The most plausible explanation of the Question × Pitch interaction (Figure 1) is that listeners took question content into consideration in making their truthfulness judgments. When listening to a potentially "loaded" topic (quota question), raters were willing to call both low- and normal-pitched voices more truthful than high-pitched voices. For the less involving question (money), raters were willing to call only low voices more truthful. This interaction argues for a pitch threshold, above which deception is signaled to raters. Such a threshold would interact with response content, so that for an emotionally involving topic, it would be set higher than for a less involving topic. With an emotionally involving topic, some of the vocally reflected stress can be attributed to the topic; given a non-involving topic, the high-pitch responses are likely to be attributed to attempted deception. However, since only two questions were used, further research is needed to test this content-attribution hypothesis.

The truthfulness ratings reflect judgmen

[6] The corresponding analysis performed on the intelligibility ratings revealed a significant effect only for the rate manipulation, with fastest speech suffering greatest loss in intelligibility.

processes that give rise to attributions of a speaker's transient state. If listeners had not been told that certain speakers were lying, differences in the acoustically manipulated variables might have been seen as enduring vocal properties reflecting stable personal predispositions. Such properties have been referred to by the linguist Trager (1958) as the "voice set" and involve "the physiological and physical peculiarities resulting in the patterned identification of individuals as . . . persons of a certain sex, age, state of health, body build, rhythm state" (p. 4). The voice set, therefore, acts as a relatively permanent background against which transient vocal changes are superimposed. In the absence of situational factors (e.g., the possibility that the speaker was lying) that could explain the voice qualities produced by our acoustic manipulations, listeners would be likely to ascribe such qualities to the voice set. How such variables affect person perception and contribute to vocal stereotypes was explored in Experiment 2.

Experiment 2

Method

Stimulus materials. The quota and money tapes from Experiment 1 were used.

Procedure. Eleven college students, nine males and two females, were paid for their participation as raters. The procedure was essentially the same as that used in Experiment 1, with the following differences: Raters were instructed that the study's purpose was to investigate how listeners form impressions of speakers from the things they say as well as from the way they say them. Accordingly, raters were told to focus both on content and delivery when making their judgments.

The speaker of each recorded segment was rated on nine bipolar adjective pairs taken from the semantic differential (Osgood, Suci, & Tannenbaum, 1957); scales were chosen that had high loadings on one of Osgood et al.'s semantic space factors and relatively low loadings on the other two. The scales for the evaluation factor were sour–sweet, awful–nice, and bad–good. Scales for the potency factor were thin–thick, small–large, and weak–strong. Those for the activity factor were slow–fast, cold–hot, and passive–active.

A warning tone followed each recorded segment, signaling judges to begin making the nine ratings on 7-point scales. (The second adjective of each pair was scored as 7.) Instructions stressed that ratings should reflect the listener's impression of the speaker and not agreement or disagreement with the content of the answer. Subjects were told that all speakers had answered the questions truthfully.

Results

Means of the nine scales were computed for each recorded segment, and the intercorrelation matrix was factor analyzed using principal factoring followed by a varimax rotation. A three-factor solution accounted for 84.5% of the total variance and roughly corresponded to the three dimensions of Osgood et al.'s (1957) semantic space. We chose scales loading greater than .60 (in absolute value) as representative of the respective factors. Using that cutoff, Factor 1 consisted of all three activity scales (slow–fast, cold–hot, and passive–active) as well as the strong–weak scale; it accounted for 54.4% of the variance. Factor 2 was a pure evaluation dimension (with only the three evaluative scales loading appreciably: sour–sweet, awful–nice, bad–good); it accounted for 20.2% of the variance. The third factor consisted of two potency scales (thin–thick, small–large) and an activity scale (slow–fast); it accounted for 9.9% of the variance.

Each rater's data were reduced to three factor scores weighting the original scales by the factor loadings; the factor scores were adjusted for covariates (as in Experiment 1) and entered into univariate analyses of variance of the same design as that used in Experiment 1.

For Factor 1, a significant rate effect, $F''(2, 32) = 10.49$, $p < .001$, and a nearly significant Rate × Question interaction, $F''(2, 31) = 3.02$, $p < .10$, were obtained; raters' $F(2, 20) = 9.72$, $p < .01$; speakers' $F(2, 18) = 3.87$, $p < .05$. The configuration of means is shown in Figure 2. In both cases, slow speakers were perceived as less active. No significant main effects of interactions were found for Factor 2. For Factor 3, significant main effects were found for rate, $F''(2,36) = 12.87$, $p < .001$, and pitch, $F''(2, 33) = 7.94$, $p < .01$. In both cases the relationship was monotonic, with increasing pitch and rate resulting in judgments of decreasing potency. No other significant effects were found.

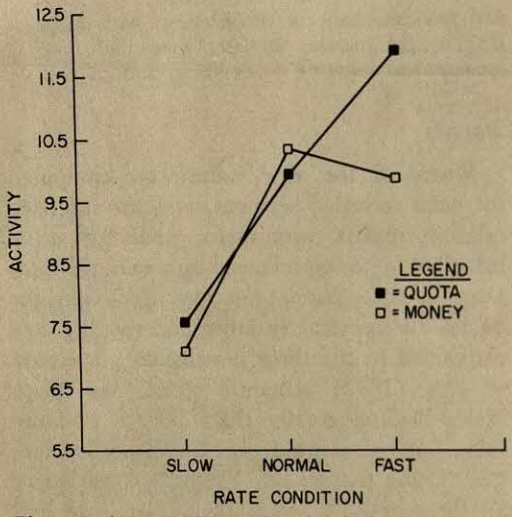


Figure 2. Averages for the activity factor plotted as a function of rate condition (slow, normal, fast) and question topic.

Discussion

These results extend the findings of Experiment 1 to judgments of more stable speaker dispositions. Men speaking in higher pitched voices were perceived as less potent (smaller, thinner, slower) and slow-speaking men were perceived as more passive (slower, colder, more passive, weaker) and more potent.

These findings are to some extent consistent with correlational evidence provided by Scherer, Koivumaki, and Rosenthal (1972). In their experiment, listeners rated taped segments, taken from a recorded play, on semantic differential scales similar to the ones used here, as well as on scales reflecting the segments' acoustic properties (e.g., bass–treble, soft–loud). Unlike the present study, raters judged the emotion portrayed and not their impression of the speaker, and a variety of listening conditions were used to degrade semantic content and prosodic features. Nevertheless, Scherer et al. found a marginally significant relationship between pitch rating (bass–treble) and potency ratings (strong–weak) paralleling the main effect reported above: Lower pitched speech was placed toward the "stronger" pole in the potency dimension of the emotional-meaning space. In contrast to our results, they also found a marginal correlation between articulation rate (slow–fast) and potency ratings: Segments with slower speech were rated as "weaker." Not surprisingly, Scherer et al. also found that segments with faster speech were heavily loaded on the activity dimension.

Our findings are likewise consistent with the results of Brown et al. (1974). Although the methodological differences previously noted preclude direct comparison, Brown found that high fundamental frequency decreased competence ratings—a scale probably related to our potency dimension.

The Rate × Question interaction (Figure 2) again suggests the influence that content exerts on raters' judgments: When speakers talked about quotas, ratings on the passive–active dimension were linear with manipulated rate; for the money question, they were not.

Experiment 3

Experiment 3 returned this work to the area of judgments of the speaker's affective state. Impressions of nervousness, emphaticness, seriousness, fluency, and persuasiveness illustrate how these acoustic variables serve to convey a speaker's self-presentation under conditions in which raters believe that answers are being given honestly. These state ratings (with the exception of persuasiveness) were chosen because Krauss, Geller, and Olson (Note 2) found significant correlations between them and truthfulness ratings in a previous study of deception interactions.

Method

Stimulus materials. The quota and money tapes from Experiment 1 were used.

Procedure. Ten college-student subjects, two males and eight females, were paid to rate all segments on five state variables: fluency, emphaticness, persuasiveness, nervousness, and seriousness. The rating scales spanned 7 points, with 1 indicating that the smallest amount of a variable was judged and 7 indicating that the largest amount was judged. For example, anchors of the nervousness scale were "Not at all nervous" and "Very nervous." Subjects were encouraged to adopt their own criteria for all ratings; no external standards were given. Instructions were virtually identical to those used in Experiment 2. Again, listeners were asked to take into consideration both the content of an answer and the manner in which it was delivered.

Table 2
Min-F Values for the Five State Variables

State variable	Rate (R) effect	Pitch (P) effect	Question (Q) × R	Q × P
Persuasiveness	3.93**	2.94*	3.54**	
Fluency	6.03***		4.07**	2.73*
Emphaticness	6.66***		3.12**	3.33**
Nervousness	6.16***	5.90***		
Seriousness				4.60**

* $p < .10$. ** $p < .05$. *** $p < .01$.

All other procedural details were identical to those of Experiment 2, except that this study was run at Bell Laboratories, using students from a number of colleges who were home on summer vacation.

Results

Segment means were computed for all state measures and analyzed using covariance analyses as described above. Table 2 summarizes the findings. Note that all variables with the exception of seriousness show main effects for the rate manipulation. These effects are all of the inverted-U type with the normal (unmanipulated) speakers judged most fluent, persuasive, and so forth, and slow speakers judged lowest on these scales. (Ratings of nervousness go in the direction opposite to the other three scales.)

Only nervousness yielded a significant main effect for pitch; the pitch manipulation was marginally significant for persuasiveness. Rated nervousness increased with higher pitch, whereas rated persuasiveness decreased.

In addition, ratings of emphaticness and seriousness showed significant Question × Pitch interactions. The shape of these interactions, shown in Figure 3, is quite similar to the corresponding effect on judged truthfulness (Figure 1). For the quota question, only the highest pitched group was "underrated" on emphaticness and seriousness, whereas for the money question, both the normal- and high-pitched groups were "underrated." The fluency measure showed a comparable interaction.

There were Question × Rate interactions for fluency, persuasiveness, and emphaticness. All interactions had a similar shape; for the quota question only the slowest group suffered low ratings, while for the money question both the slow and fast groups received low ratings. The effect for seriousness judgments was marginally significant, but of the same shape.

Discussion

These results again demonstrate the effect acoustic variables have on person perception processes. Decreasing speech rate has a particularly deleterious effect on a speaker's perceived persuasiveness, fluency, and emphaticness. Similarly, increased pitch lowers ratings of persuasiveness and increases greatly the impression of nervousness.

Our findings also suggest that context plays a role in the attribution process, as evidenced by the question interactions. When a speaker

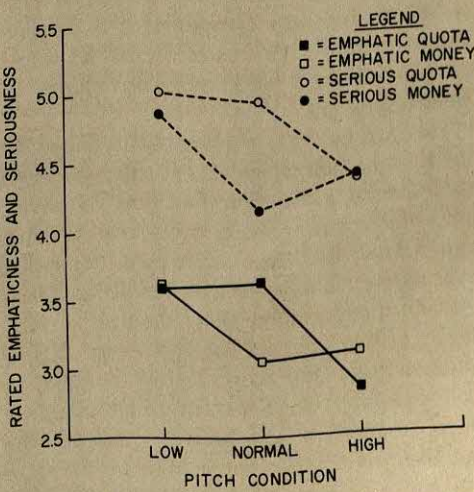


Figure 3. Average emphaticness and seriousness ratings plotted as a function of pitch condition (low, normal, high) and question topic.

answered the quota question, higher pitch levels could be discounted by attributing them to the stressfulness of the topic. Similarly, it may have been inappropriate on the money topic to talk too slowly, (perhaps because of the topic's relative simplicity) or too rapidly (perhaps because rapid speech is perceived as an attempt to be inappropriately serious or persuasive). On the other hand, rapid speech on the quota question might have been attributed to the speaker's conviction regarding his argument.

General Discussion

The three experiments taken as a whole provide clear evidence that acoustic properties of a message have considerable impact on judgments of a variety of state and trait variables. The impressions of high-pitched or slow-talking speakers seem particularly negative. For example, men with high-pitched voices are judged less truthful, less persuasive, weaker, and more nervous. Similarly, slow-talking men are judged to be less truthful, fluent, emphatic, serious, and persuasive, and more passive, although they are also seen as more potent.

By using a large number of speakers and factorially varying speech rate and pitch, we have a reasonable measure of confidence in the validity of our findings. However, since no female voices were used, we cannot generalize these results to women; it is conceivable that these same acoustic variables would produce different effects on perceptions of women speakers.

Message context, presumably mediated by the two question topics, was also demonstrated to influence attribution processes. However, since there was only one question of each type, what follows must remain somewhat speculative. The question interactions suggest an interpretation along the lines of Kelley's (1971) discounting principle. Kelley suggests that the more factors a situation contains—any one of which might plausibly have resulted in an observed outcome—the less likely is any one factor to be perceived as the cause of that outcome. With fewer possible causes present, the cause-to-effect attribution is more compelling. We hoped that our acoustic manipulations would be potent enough to affect speaker state or trait attributions. However, it is evident that listeners may have, at least partially, attributed our acoustic alterations to the question topic. The quota question called for an answer that was both complex and emotionally involving for our college-student speakers and raters. As Table 1 shows, before any manipulations were done, the quota answers were longer and slower (suggesting greater cognitive complexity; see Goldman Eisler, 1968) and higher pitched (suggesting greater stressfulness; see Hecker et al., 1968). Because of this, slower and higher pitched answers might have been perceived as more appropriate for quota responses than for money responses.[7] But even when premanipulation speech rate and pitch were covaried, the question interactions remained, supporting the discounting principle. For example, higher pitch levels seemed to be discounted on ratings of truthfulness, fluency, emphaticness, and seriousness for speakers answering the quota, but not the money, question.

It should be noted that manipulating fundamental frequency by multiplying by a scale factor as we did has the effect of multiplying the *variance* of the fundamental frequency by the square of the scale factor. Thus, high-pitched segments were both high-pitched and high pitch-variance segments, and vice versa. Thus, we cannot rule out the interpretation that the pitch effects observed could be pitch-variance effects. However, this interpretation seems unlikely considering the findings of Brown et al. (1974), who did the appropriate factorial experiment and found that increased mean fundamental frequency lowered judgments of speakers' competence and benevolence, while decreased variance also lowered these ratings. Thus, it seems that average pitch and pitch variance affect judgments on

[7] Subsequently, in connection with another study, we had 12 undergraduates rate a long list of potential interview topics (including the two used in the present study) as to how stressful and how complex they would be for a typical Columbia undergraduate to discuss. Subjects judged the quota question to be significantly more stressful and more complex than the money question.

a wide range of scales in opposite ways. In the present study, in which both pitch and pitch variance were positively correlated, the pitch-correlated variance could only have attenuated the effects of average pitch.

Given that the experimental manipulations used here affected the perception of speakers' personality and state, it remains to be explained why these particular data patterns were observed. Miller et al. (1976), for example, conclude that the effect of speech rate on message persuasiveness is mediated by the effect that variable has on the perception of a speaker's credibility. This, they assert, is a "less rationalistic view" of attitude change than other interpretations (e.g., change mediated by comprehension effects or counterargument disruption—two hypotheses their experiments ruled out).

Certainly such a conclusion would be justified if it were the case that variations in voice quality bore no relation to the actual internal state or predisposition of the speaker. However, there is considerable evidence that stressful situations do produce discernible changes in voice quality (Fairbanks, 1940; Hecker et al., 1968; Williams & Stevens, 1969, 1972), and it seems reasonable to assume that listeners in the present study used such variations appropriately to infer a speaker's state from the quality of his speech. For example, Hecker et al. demonstrated that task-induced stress raised the fundamental frequency of those speakers who did not talk more softly under stress. In our experiment, attributions of increased nervousness to higher-pitched speakers are quite "rationalistic," given the similarity of our pitch manipulations to the effect of real-life stress. Similarly, Streeter et al. (1977) demonstrated that pitch increments accompany deceptive responses; listeners' truthfulness judgments in Experiment 1 appropriately reflect this relationship.

The effects of speech rate can be similarly interpreted in light of Goldman Eisler's (1968) finding that rate and the cognitive complexity of the topic were negatively related. Listeners may have assumed that lying increases speakers' cognitive load, resulting in slower rates. Unpublished speech rate data from the Streeter et al. (1977) study lend plausibility to this argument. A marginally significant interaction ($p < .08$) indicated that subjects did, in fact, speak more slowly when lying than when telling the truth, provided they had been given instructions engaging their motivation to lie effectively. However, speakers not receiving such instructions spoke more rapidly when lying.

Judgments of what is or is not "rationalistic" are probably to a large extent matters of personal preference, but it does seem to us that a listener would be ill-advised to ignore reliable information concerning a speaker's internal state in evaluating the speaker's message, especially when the internal state seems incongruent with the situation or with the message's content.

Furthermore, the significant Question × Manipulation interactions for ratings of state variables suggest some qualifications on the findings of Miller et al. (1976). Fast speakers are not always more persuasive; talking too quickly in response to the money question produced lower persuasiveness ratings than did responses at a normal rate. Apparently, listeners take more into account than meets the ear—at least, more than simply the acoustic data.

Evidence of veridicality for vocally based attributions of enduring personality traits is less firm. We have found no reliable data to indicate that fast talkers actually are more active people or that higher pitched men are weaker than their lower pitched counterparts. Apart from studies of psychiatric patients, much of the work in this area deals with the traits of introversion–extraversion and dominance. Mallory and Miller (1958) found small but significant negative correlations between *judged* pitch and rate and the dominance scale of the Bernreuter Personality Inventory. While these findings support our results on judged potency along the pitch dimension (Experiment 2), it is possible that Mallory and Miller's acoustic judgments are either inaccurate or subject to biasing effects from other sources, since raters judged speakers in a live situation. Furthermore, our results for judged potency on the speech-rate variable appear opposite to Mallory and

Miller's. A more recent study (Ramsay, 1966) found that the speech of subjects classified as extraverts on the Eysenck Personality Inventory had longer unbroken phonation times and shorter silences than those of introverts across a variety of speaking tasks; no data on speech rates were presented.

The acoustic stimulus, of course, contains more information than average pitch and rate. In addition, there is sequential information (provided by intonation contours and duration pattern), loudness, and variability of both pitch and loudness over time. There is also voice quality information (e.g. "breathy" or "raspy" voices) that may not be so readily specified in terms of physical parameters. All of these factors can be expected to enter into the person perception process via stereotypes with larger or smaller kernels of truth.

On none of the measures we examined was the Rate × Pitch interaction statistically significant. The median F interaction was 1.12—close to its expected value under the null hypothesis. This absence of interaction argues for an additive model, in which pitch and rate exert independent effects on listeners' judgments. It remains to be seen whether support for such a model will continue as the role of additional vocal factors is explored.

Reference Notes

1. Nakatani, L. H. *SYNLOG: An interactive system for manipulating speech.* Unpublished manuscript, Bell Laboratories, Murray Hill, N.J., 1976.
2. Krauss, R. M., Geller, V., & Olson, C. T. *Modalities and cues in the detection of deception.* Paper presented at the meeting of the American Psychological Association. Washington, D.C., September 1976.

References

Allport, G. W., & Cantril, H. Judging personality from voice. *Journal of Social Psychology,* 1934, *5,* 37–55.

Atal, B. S., & Hanauer, S. L. Speech analysis and synthesis by linear prediction of the speech wave. *Journal of the Acoustical Society of America,* 1971, *50,* 637–655.

Black, J. W. Relationships among fundamental frequency, vocal sound pressure, and rate of speaking. *Language and Speech,* 1961, *4,* 196–199.

Brown, B. L., Strong, W. J., & Rencher, A. C. Perceptions of personality from speech: Effects of manipulations of acoustical parameters. *Journal of the Acoustical Society of America,* 1973, *54,* 29–35.

Brown, B. L., Strong, W. J., & Rencher, A. C. Fifty-four voices from two: The effects of simultaneous manipulations of rate, mean fundamental frequency, and variance of fundamental frequency on ratings of personality from speech. *Journal of the Acoustical Society of America,* 1974, *55,* 313–318.

Fairbanks, G. Recent experimental investigations of vocal pitch in speech. *Journal of the Acoustical Society of America,* 1940, *11,* 457–466.

Goldman Eisler, F. *Psycholinguistics: Experiments in spontaneous speech.* London: Academic Press, 1968.

Hanley, T. D. An analysis of vocal frequency and duration characteristics of selected samples of speech from three American dialect regions. *Speech Monographs,* 1951, *18,* 78–93.

Hecker, M. H. L., Stevens, K. N., von Bismarck, G., & Williams, C. E. Manifestations of task-induced stress in the acoustic speech signal. *Journal of the Acoustical Society of America,* 1968, *44,* 993–1001.

Kelley, H. H. *Attribution in social interaction.* Morristown, N.J.: General Learning Press, 1971.

Kramer, E. Judgment of personal characteristics and emotions from nonverbal properties of speech. *Psychological Bulletin,* 1963, *60,* 408–420.

Lay, C. H., & Burron, B. F. Perception of the personality of the hesitant speaker. *Perceptual and Motor Skills,* 1968, *26,* 951–956.

Mallory, E. B., & Miller, V. R. A possible basis for the association of voice characteristics and personality traits. *Speech Monographs,* 1958, *25,* 255–260.

Miller, G. R., & Hewgill, M. A. The effect of variations in nonfluency on audience ratings of source credibility. *Quarterly Journal of Speech,* 1964, *50,* 36–44.

Miller, N., Maruyama, G., Beaber, R. J., & Valone, K. Speed of speech and persuasion. *Journal of Personality and Social Psychology,* 1976, *34,* 615–624.

Mysak, E. Pitch and duration characteristics of older males. *Journal of Speech and Hearing Research,* 1959, *2,* 46–54.

Nakatani, L. H. Computer-aided signal handling for speech research. *Journal of the Acoustical Society of America,* 1977, *61,* 1057–1062.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. *The measurement of meaning.* Urbana: University of Illinois Press, 1957.

Peterson, G. E., & Barney, H. L. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America,* 1952, *24,* 175–184.

Ramsay, R. W. Personality and speech. *Journal of Personality and Social Psychology,* 1966, *4,* 116–118.

Scherer, K. R. Acoustic concomitants of emotional dimensions: Judging affect from synthesized tone sequences. In S. Weitz (Ed.), *Nonverbal communication.* New York: Oxford University Press, 1974.

Scherer, K. R., Koivumaki, J., & Rosenthal, R Minimal cues in the vocal communication of affect: Judging emotions from content-masked speech. *Journal of Psycholinguistic Research,* 1972, *1,* 269–285.

Smith, B. L., Brown, B. L., Strong, W. J., & Rencher, A. C. Effects of speech rate on personality perception. *Language and Speech,* 1975, *18,* 145–152.

Streeter, L. A., Krauss, R. M., Geller, V., Olson, C., & Apple, W. Pitch changes during attempted deception. *Journal of Personality and Social Psychology,* 1977, *35,* 345–350.

Terango, L. Pitch and duration characteristics of the oral reading of males on a masculinity–femininity dimension. *Journal of Speech and Hearing Research,* 1966, *9,* 590–595.

Trager, G. L. Paralanguage: A first approximation. *Studies in Linguistics,* 1958, *13,* 1–12.

Williams, C. E., & Stevens, K. N. On determining the emotional state of pilots during flight. An exploratory study. *Aerospace Medicine,* 1969, *40,* 1369–1372.

Williams, C. E., & Stevens, K. N. Emotions and speech: Some acoustical correlates. *Journal of the Acoustical Society of America,* 1972, *52,* 1238–1250.

Winer, B. J., *Statistical principles in experimental design* (2nd ed.), New York: McGraw-Hill, 1971.

Received January 18, 1978 ■

Corrections to Koretzky, Kohn, and Jeger

In the article "Cross-Situational Consistency Among Problem Adolescents: An Application of the Two-Factor Model" by Martin B. Koretzky, Martin Kohn, and Abraham M. Jeger (*Journal of Personality and Social Psychology,* 1978, Vol. 36, No. 9, pp. 1054–1059), there are errors in two correlations reported in the first paragraph on page 1058. The Factor I and Factor II scores for the Koretzky (1976) study cited there should be reversed. The correct correlation for Factor I is .40, and for Factor II it is .60. Thus, the sentence should read, "Consistency correlations between classroom and residence settings were even stronger in this experiment for Factor II ($r = .60$) and were also respectable for Factor I ($r = .40$)."

Martin B. Koretzky's affiliation was erroneously given as the Veterans Administration Hospital, Bronx, New York. His affiliation at the time of the original research was the State University of New York at Stony Brook.

Journal of Personality and Social Psychology
1979, Vol. 37, No. 5, 728–734

Smiles Can Be Back Channels

Lawrence J. Brunner
Department of Behavioral Sciences
University of Chicago

The placement of auditor smile beginnings in the stream of dyadic interaction was investigated, using detailed transcriptions of the language, paralanguage, and body motion of the participants in four two-person conversations recorded on videotape. Auditor smile beginnings showed a strong tendency to occur at the same kinds of location as "back channel" responses (such as "yeah," "uh-huh," and head nods). This finding indicates that the smile can function as a type of back channel. It is argued that smiles, like other forms of back channel, make communication more efficient by providing the speaker with feedback on a number of levels simultaneously.

Most quantitative investigations of smiling in adults (see reviews by Berlyne, 1969; Chapman & Foot, 1976; and Goldstein & McGhee, 1972) have treated it as expressive of underlying states such as mirth or pleasurable arousal. Possibly because researchers in social psychology and personality tend to focus their attention more on hypothetical constructs than on the behavior that those constructs are intended to explain (a thesis that Fiske, 1978, argues cogently and in detail), the frequent failure to find significant correlations between ratings of amusement and actions such as smiling and laughing (Chapman, 1976, pp. 156–157) has been attributed (by Berlyne, 1969, and Chapman, 1976, for example) to the greater "sensitivity" of self-report measures.

The present study is deliberately opposite in approach. Without denying that people often smile when they are happy or when something strikes them as funny, it is concerned with smiles as actions rather than expressions of unobservable inner states. It is an attempt to partly specify the "meaning" of smiles—specifically, the smiles of the person who is not speaking in a two-person conversation—in what Goffman (1971, p. 235) calls the "structural" sense of the term. That is, it is an attempt to describe their placement in the stream of behavior.

The study is part of a program of investigation (Duncan, 1972, 1974; Duncan & Fiske, 1977; Duncan & Niederehe, 1974) whose purpose is to discover the organization of face-to-face interaction. The term *organization* reflects the assumption that just as language operates according to rules that give it a definite structure, so does the process of face-to-face interaction in which language ordinarily occurs.

The research of Duncan and his associates on the structure of interaction is based on detailed transcription of speech and body motion in two-person conversations. Exploration of these materials has led to the formulation of a set of hypothesized signals and rules that people apparently use to exchange speaking turns and otherwise regulate the flow of interaction in conversations. These rules depend on a distinction between the "speaker" (the person who holds the speaking turn)

This study is based on a dissertation submitted in partial fulfillment of the requirements for a doctoral degree in psychology at the University of Chicago. The author would like to acknowledge the support and guidance of his committee—Starkey Duncan, Jr. (Chairman), Donald W. Fiske, and Norman Bradburn.

The research was supported by National Science Foundation Grant SOC 74-24084 to Starkey Duncan, Jr., and Donald W. Fiske.

Requests for reprints should be sent to Lawrence J. Brunner, 1605 East 55th Street, Chicago, Illinois 60615.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3705-0728$00.75

and the "auditor" (a person who is in the conversation without holding the turn). To hold the speaking turn is not the same as to emit speech sounds; in the conversations we have studied, brief pauses seldom if ever constitute relinquishment of the speaking turn, and it is possible for a person to talk without claiming the speaking turn (for example, to say things like "yeah" and "uh-huh").

Speaking Turns

Duncan (1972) hypothesized a turn signal composed of six separate cues in language, paralanguage, and body motion. If any one or any combination of these cues is displayed, the auditor may take the turn if he or she wishes to. The auditor is not required to take the turn when the turn signal is displayed; he or she merely has the option of doing so. There is a marked tendency for the turn signal to precede orderly (smooth) exchanges of the speaking turn, but it seldom precedes instances of "simultaneous turns" in which both participants attempt to hold the floor at the same time.

Complementary to the turn signal is the "gesticulation" signal. While the speaker is gesticulating, the auditor should not attempt to take the turn, regardless of whether the turn signal is displayed. The evidence for this formulation is that the auditor is much less likely to attempt the turn when the speaker is gesturing and that if the auditor does, the result is usually simultaneous turns, even when one or more turn cues are also present.

Back Channels

The turn signal and the gesticulation signal are both conceived to facilitate the smooth exchange of turns. In further development of the turn system, attention has been directed toward speaker–auditor interchanges that occur *during* speaking turns. Duncan (1974) reported a study of auditor "back channels" (head nods and statements such as "yeah" and "uh-huh") that explored some of the differences between back channels and turn attempts and related auditor back channels to the actions of the speaker.

Unlike turn attempts, auditor back channels are not suppressed by the speaker's gesticulation. In addition, back channels and turn attempts are distributed differently with respect to the boundaries of segments into which the conversations were divided for purposes of statistical analysis. It is likely that these segments reflect structural properties of speech and communicative body motion (Duncan & Fiske, 1977, discuss them in detail). Turn attempts usually occur exactly between segments, whereas back channels are more loosely distributed.

Another important distinction between auditor back channels and turn attempts is that they seem to be related to different sets of cues given by the speaker. Duncan (1974) hypothesized a speaker "within-turn" signal originally consisting of two cues: grammatical completion and turning of the head toward the auditor. Brunner (1977) has reanalyzed the data and found evidence for changing the head turn cue to simple head direction toward the auditor. Responses to the within-turn signal, like responses to the turn signal, are conceived to be optional; the speaker is not obliged to give a back channel, but may do so if a signal is present.

The thesis of the present study is that the placement of smile beginnings in conversations is closely parallel to that of back channels and that, in fact, smiling can be a type of back channel. This idea was suggested by the entirely serendipitous finding that most auditor smile beginnings were preceded by the within-turn signal. In order to claim, however, that the auditor's smile beginnings occupy a place that is similar to back channels in the structure of interaction, three other requirements must be met.

1. There must be an acceptable statistical relation between occurrence of a smile beginning and previous display of the within-turn signal.

2. Auditor smile beginnings must be more loosely distributed with respect to segment boundaries than auditor turn attempts are.

3. The auditor's smile beginning must not be suppressed by the speaker's gesticulation.

Method

Conversations and Transcription

The four conversations on which this report is based are a subset of the ones described by Duncan and Fiske (1977). Conversations 1, 2, and 3 are 7 minutes long. All three were between female law students and male social work students at the University of Chicago; the participants were paid volunteers who were previously unacquainted. Conversation 4 is 19 minutes long and, unlike the other three conversations, would have occurred whether or not it was videotaped. It was between two 40-year-old male psychotherapists from the Counseling Center at the University of Chicago. They were good friends and had known each other for about 10 years. The subject of their conversation was a client whom both of them had been seen recently. Duncan and Fiske give more detailed background information about these conversations. They refer to Conversations 1 through 4 as numbers 5, 6, 7, and 2, respectively.

In all four conversations, the participants were seated in adjacent chairs turned slightly toward each other, facing a video camera. The participants in the first three conversations were requested to get acquainted and carry on a short conversation, and in the fourth they were asked to continue a conversation that they had been unable to complete on a previous occasion.

The conversations have been transcribed in considerable detail, but for the present study, only the following actions were important: spoken syllables, completions of subject–predicate clauses, intonation (according to Trager & Smith's, 1957, system), gestures, head movements, and smiles. In this study, the term *smile* refers to apparent bilateral or unilateral contraction of the risorius and buccinator muscles. Each event was carefully located with respect to the other events in the conversation so as to produce a transcript in which the sequence of actions is represented, accurate to one syllable.

In transcribing body motion for this study, the author and a research assistant worked together with the goal of making a transcription on which they completely agreed; in cases of doubt, other people familiar with the transcription system were consulted. Typically, agreement between observers is very high without discussion, and previous studies of the reliability of the transcription system used in the present investigation have yielded a median kappa (Cohen, 1960) of .81 (roughly 95% agreement) between separate observers (Duncan & Fiske, 1977, p. 342).

Data Analysis

Unit of analysis. In order to assess behavioral regularities within an interaction, it must be divided into units to which scores are assigned. The reason is that in order to present evidence that some action, A, precedes B more often than one would expect by chance, it is necessary to know not only how many times A preceded B and how many times each action occurred separately but also how often each action did not occur. In the analyses to be reported here, the unit of analysis is a stretch of speech that can be as short as a single phonemic clause (Trager & Smith, 1957) and as long as an entire speaking turn.

The boundaries of the units of analysis (which will be called "segments" for convenience) occur after the boundaries of phonemic clauses that contain one or more of the following actions: unfilled pause, audible inhalation or exhalation, turning of the speaker's head toward the auditor, a drop in paralinguistic pitch or loudness, any pitch level – terminal juncture combination other than 2 2 | at the end of a phonemic clause, paralinguistic drawl on the final or the stressed syllable of a phonemic clause, termination of a gesture or relaxation of a tensed hand position by the speaker, use of a stereotyped expression such as "but uh" or "you know," or the completion of a grammatical clause consisting of a subject and predicate. Duncan and Fiske (1977) explain segments in much more detail.

In addition to serving as the basic units of analysis, segments are used to describe the distribution of events within an interaction. The locations of the auditor's actions with respect to segment boundaries can be represented in a fourfold typology. The four categories of location are the following:

1. Speech overlap—from roughly the middle to slightly before the end of the segment.

2. Pause—between the last syllable of the unit in question and the first syllable of the next unit.

3. Sociocentric sequence—before or during a sociocentric sequence such as "but uh" or "you know."

4. Postboundary—from the very beginning to roughly the middle of the segment after the one in question.

Statistics. Evidence for the relation between the speaker's within-turn signal and the auditor's smile beginning will be presented in a fourfold table where the columns indicate presence versus absence of the signal, and the rows indicate presence versus absence of a subsequent smile beginning by the auditor. If the results are parallel to those obtained for back channels, and the speaker's within-turn signal permits an auditor smile beginning but does not require one, almost every auditor smile beginning should be preceded by the signal. If the signal operated perfectly, the cell representing auditor smile beginning in the absence of the within-turn signal would be zero. There is no requirement, however, that the other cell in that diagonal (failures to respond to the signal) be zero or even have a particularly low frequency; permissive signals allow the option of not responding.

Consequently, the most useful index of association between signal and response for the table just described is one that reaches its maximum when only one cell is empty. The statistic Q (described by Yule, 1900, and further developed by Goodman & Kruskal, 1954, 1959, 1963, 1972) has this property: It equals

1.0 or −1.0 when one cell or both cells on one diagonal of a 2 × 2 table equals zero. The *Q* also has the advantage of being a true index of association, directly interpretable as percentage of improvement in predicting the dependent variable when the independent variable is known. The standard error of *Q* can be calculated, and the hypothesis that *Q* is significantly different from zero may be tested by dividing *Q* by its standard error and referring the result to the normal curve. The *Q* and its two-tailed probability value will be reported for all the 2 × 2 tables in this article.

Replication. Since the results to be reported here are the products of very thorough exploration rather than hypothesis testing, they must be replicated in order to ensure that they do not merely reflect random contingencies in the data. I have used Conversations 1 and 2 for exploration and Conversations 3 and 4 for replication. Although pooled data are reported for both the exploratory and replication samples, the results for all eight participants were examined separately to make sure that the same trends held for all individuals.

The analyses to be reported do not include segments with auditor turn attempts or smiles continued from the previous segment. Segments with turn attempts were excluded because when a person smiled as he or she claimed the speaking turn, it was often impossible to decide whether to call the action a speaker smile or an auditor smile. Segments with auditor smile continuations were excluded because the auditor could not begin to smile when he or she was already smiling.

Results

The Within-Turn Signal

Previous research has shown that the speaker's within-turn signal regularly precedes auditor back channels whose location with respect to segment boundaries is late or between units, but not those which are early or occur before or during a sociocentric sequence. This is also the case for auditor smile beginnings.

Table 1 shows that in both the exploratory and replication conversations, the speaker's within-turn signal (composed of grammatical completion or head direction toward the auditor or both) preceded a high percentage of the auditor's late and between-units smile beginnings and was significantly related to their occurrence. In Conversations 1 and 2, 23 of 24, or 95.8%, of the auditor's smile beginnings occurred while the within-turn signal was present. For the 2 × 2 table relating the

Table 1
Relationship Between Display of Speaker's Within-Turn Signal and Subsequent Display of Auditor Smile Beginning

Auditor's smile beginning	Speaker's within-turn signal: Proportion of segments with signal absent	Speaker's within-turn signal: Proportion of segments with signal present	Row total of segments
Pause and postboundary smiles: Exploratory sample[a] (Conversations 1 and 2)			
Absent	.27	.72	153
Present	.04	.96	24
Pause and postboundary smiles: Replication sample[b] (Conversations 3 and 4)			
Absent	.26	.74	487
Present	.00	1.00	15
Speech overlap and sociocentric sequence smiles[c] (Conversations 1, 2, and 4)			
Absent	.23	.77	697
Present	.33	.67	12

[a] $Q = .79$, $p = .00006$.
[b] $Q = 1.00$, $p \cong 0$.
[c] $Q = -.26$, $p = .36$.

auditor's smile beginning to prior display of the speaker's within-turn signal, $Q = .79$, $p = .00006$. In Conversations 3 and 4, all 15 between-units and late auditor smile beginnings happened in the presence of the within-turn signal. $Q = 1.0$, and the associated probability cannot be calculated because the standard error is zero.

The bottom third of Table 1 shows that there is an insignificant negative relation between presence of the speaker's within-turn signal and auditor smile beginnings located at speech overlaps or sociocentric sequences. Similar results are reported by Duncan and Fiske (1977) for vocal back channels and head nods that occur in these locations.

Effect of Speaker's Gesticulation

Auditor turn attempts are suppressed by the speaker's gesticulation, but back channels are not. To test whether smile beginnings are also unaffected by the speaker's gesticulation, data were arranged in 2 × 2 tables where the rows indicated presence versus absence of the speaker's gesticulation and the columns indicated presence versus absence of the auditor's smile beginning. In the exploratory sample (Conversations 1 and 2), the auditor's smile and the speaker's gesticulation were not significantly related ($Q = -.36$, $p = .14$).

In the replication sample (Conversations 3 and 4), it seemed at first that the auditor was somewhat less likely to smile while the speaker was gesticulating ($Q = -.45$, $p = .04$). These are not strong results, but it is strange that smile beginnings should show a structural similarity to turn attempts rather than back channels, particularly when this did not happen in the exploratory study.

Examination of the data from Conversations 3 and 4 separately indicates that there is virtually no relationship between the speaker's gesticulation and the auditor's smile in either conversation. Conversation 4 has a low rate of auditor smiling and a high rate of speaker gesticulating; Conversation 3 has a relatively high rate of auditor smiling and an extremely low rate of speaker gesticulating. Consequently, although there is no relation between the speaker's gesticulation and the auditor's smile beginning in either conversation, an artificial one appears when data from the two are pooled. This finding shows how important it is to examine subsets of data separately before combining them.

Table 2
Locations of Auditor Smile Beginnings and Turn Attempts With Respect to Segment Boundaries

Location	Auditor turn attempts: Frequency	Auditor turn attempts: %	Auditor smile beginnings: Frequency	Auditor smile beginnings: %
Exploratory sample (Conversations 1 and 2)				
Speech overlap	16	15.4	8	24.2
Pause	84	80.8	16	48.5
Postboundary	2	1.9	8	24.2
Sociocentric sequence	2	1.9	1	3.1
Total	104		33	
Replication sample (Conversations 3 and 4)				
Speech overlap	13	22.0	2	11.8
Pause	40	67.8	9	52.9
Postboundary	2	3.4	6	35.3
Sociocentric sequence	4	6.8	0	0
Total	59		17	

Table 2 compares the distributions of auditor smiles and turn attempts with respect to segment boundaries. Auditor smile beginnings are less concentrated in the "pause" location (the same finding that has been obtained for back channels), although this tendency is not as pronounced in the replication study.

Discussion

The organization of auditor smile beginnings in these conversations is parallel to that of back channels. They are both consistently preceded by the speaker's within-turn signal, they are both unaffected by the speaker's gesticulation, and they are both more loosely distributed than turn attempts with respect to segment boundaries. Taken together, these results are grounds for placing auditor smile beginnings in the category "back channel," along with nods and statements such as "yeah" and "uh-huh."

Nods and vocal back channels seem intuitively to belong together, because they both seem to say, in effect, "Yes, you are understood. Proceed." Although it is not immediately obvious that smiles belong in this category, I believe that auditor smile beginnings, like auditor back channels, make communication more efficient by providing the speaker with feedback.

The idea of communicative efficiency lies behind much of the turn system, though it has not been given a great deal of emphasis. For example, the turn signal was constructed so that at least one cue would be present before each smooth exchange of turns, but not before instances of simultaneous turns. The reasoning was that a system of rules for taking speaking turns ought to result in smooth exchanges when it is working properly, because human beings have only a limited capacity to send and receive information simultaneously.

The people who have written about the actions that we call back channels generally agree that the auditor uses them to provide the speaker with information. There is less unanimity, however, concerning the exact sort of information that back channels convey. Fries (1952, p. 49) suggests that they are primarily signs of attention and involvement in the conversation. Yngve (Note 1) agrees, but proposes that they are also "very important in providing for monitoring of the quality of the communication" (p. 568). Dittmann and Llewellyn (1967) imply that in addition, they are sometimes used for directly requesting further information or explanation (this is consistent, of course with Duncan's treatment of short questions as back channels).

Kendon (1967) distinguishes between "point granting" and "assenting" signals, both of which we would describe as back channels. This distinction merely reflects the fact that usually the auditor's feedback indicates that he or she is following what the speaker says, and sometimes it also indicates agreement. It is important to mention in this connection that head shakes and verbal statements of disagreement were treated as back channels in the present study and in Duncan's earlier investigations. This was done mostly on intuitive grounds (they "felt" like back channels), but the implicit rationale was that feedback need not always be positive in tone, although it usually is.

There is a consistency among the various sorts of information that back channels are supposed to provide. I would like to suggest that back channels (including smiles) give the speaker feedback on three levels. On the first level, back channels signal the auditor's involvement and participation in the interaction. They indicate that the auditor is attending to what the speaker says and that a conversation, not a monologue, is occurring.

On the second level, back channels provide information about the auditor's level of understanding, allowing the speaker to adjust his or her communicative endeavor so as to get the ideas across efficiently. "Mhm's" and head nods tell the speaker that the message is being decoded successfully; when the location of these signals is "speech overlap" (before the end of a segment of speech is reached), they may indicate that the auditor is "ahead" of the speaker and a little less elaboration would be in order; a puzzled or blank look may indicate that more information is needed (no work has been done on the actions that signify the auditor's lack of understanding, except for clarifying questions).

On the third level, back channels can signal the auditor's personal response to what the speaker has just said. This might mean agreement or disagreement, shock, amusement, scorn, or any number of other reactions. In general, if an action functions on a higher level it also functions on the ones below it; any indication of a personal reaction implies understanding (or the lack of it), and any indication of a degree of understanding implies some involvement in the conversation. Smiles are higher level back channels, providing feedback on all three levels. They indicate a positive personal reaction (often merely politeness), as well as understanding of the preceding statement and participation in the conversation.

We do not know how far the results of the present investigation can be generalized. The findings of the exploratory study were fully

replicated, which indicates that at least they are not unique to a particular conversation. Definitive information about the limits within which they hold can be provided only by further studies of smiling in which many people are observed in many different settings.

Reference Note

1. Yngve, V. H. On getting a word in edgewise. In *Papers from the sixth regional meeting of the Chicago Linguistic Society*. Chicago: Chicago Linguistic Society, 1970.

References

Berlyne, D. E. Laughter, humor and play. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (2nd ed.). New York: Academic Press, 1969.

Brunner, L. J. *Smiling and laughter in four two-person conversations*. Unpublished doctoral dissertation, University of Chicago, 1977.

Chapman, A. J. Social aspects of humorous laughter. In A. J. Chapman & H. C Foot (Eds.), *Humor and laughter: Theory, research and applications*. New York: Wiley, 1976.

Chapman, A. J., & Foot, H. C. (Eds.). *Humor and laughter: Theory, research, and applications*. New York: Wiley, 1976.

Cohen, J. A. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 1960, *20*, 37–46.

Dittmann, A. T., & Llewellyn, L. G. The phonemic clause as a unit of speech decoding. *Journal of Personality and Social Psychology*, 1967, *6*, 341–349.

Duncan, S., Jr. Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology*, 1972, *23*, 283–292.

Duncan, S., Jr. On the structure of speaker–auditor interaction during speaking turns. *Language in Society*, 1974, *2*, 161–180.

Duncan, S., Jr., & Fiske, D. *Face-to-face interaction*. Hillsdale, N.J.: Erlbaum, 1977.

Duncan, S., Jr., & Niederehe, G. On signaling that it's your turn to speak. *Journal of Experimental Social Psychology*, 1974, *10*, 324–247.

Fiske, D. W. *Strategies for personality research*. San Francisco: Jossey-Bass, 1978.

Fries, C. C. *The structure of English*. New York: Harcourt, Brace, 1952.

Goffman, E. *Relations in public*. New York: Harper & Row, 1971.

Goldstein, J. H., & McGhee, P. E. (Eds.). *The psychology of humor: Theoretical perspectives and empirical issues*. New York: Academic Press, 1972.

Goodman, L. A., & Kruskal, W. H. Measures of association for cross-classifications. *Journal of the American Statistical Association*, 1954, *49*, 732–764.

Goodman, L. A., & Kruskal, W. H. Measures of association for cross-classifications, II: Further discussion and references. *Journal of the American Statistical Association*, 1959, *54*, 123–163.

Goodman, L. A., & Kruskal, W. H. Measures of association for cross-classifications, III: Approximate sampling theory. *Journal of the American Statistical Association*, 1963, *58*, 310–364.

Goodman, L. A., & Kruskal, W. H. Measures of association for cross-classifications, IV: Simplification and asymptotic variances. *Journal of the American Statistical Association*, 1972, *67*, 415–421.

Kendon, A. Some functions of gaze direction in social interaction. *Acta Psychologica*, 1967, *26*, 22–63.

Trager, G. L., & Smith, H. L., Jr. *An outline of English structure*. Washington, D.C.: American Council of Learned Societies, 1957.

Yule, G. U. On the association of attributes in statistics. *Philosophical Transactions of the Royal Society of London*, 1900, *A 194*, 257–319.

Received February 14, 1978

Journal of Personality and Social Psychology
1979, Vol. 37, No. 5, 735–746

Motoric and Symbolic Mediation in Observational Learning

Seymour M. Berger, Linda L. Carli, Kathy S. Hammersla,
Judith F. Karshmer, and M. Estela Sanchez
University of Massachusetts—Amherst

The observers' motoric and symbolic representations of a model's behavior are important mediators in observational learning. The observers' *spontaneous* use of these mediators may be influenced by their familiarity with responses performed by a model and by their intention to learn these responses. Unfamiliar observers do not have symbolic codes available for the model's responses, so they may rely on motor mimicry. Familiar observers have symbolic codes available, so they may employ these codes and, possibly, motor mimicry as mediators. Spontaneous mediation may also depend on whether observers intend to learn these responses. Experiments revealed that familiar observers use motoric and symbolic mediators, whereas unfamiliar observers primarily use motor mimicry. Symbolic coding facilitates familiar observers' but not unfamiliar observers' learning; unfamiliar observers' learning is related to motor mimicry. Intention to learn increases motor mimicry but not symbolic coding. An interpretation is offered for the observers' pervasive use of motor mimicry.

Bandura (1974) has emphasized the role of imaginal and verbal coding, and rehearsal of these symbolic codes, in observational learning. Berger (1966, 1968), on the other hand, has focused on the role of motor mimicry, motor rehearsal, and conditioning in observational learning. Studies supporting these analyses, however, either have not examined the observers' *spontaneous* use of different forms of mediation or else have *instructed* observers to use different mediators (e.g., Bandura, Grusec, & Menlove, 1966; Bandura & Jeffery, 1973; Berger, Irwin, & Frommer, 1970; Bernal & Berger, 1976; Gerst, 1971; Jeffery, 1976). The following experiments examine some conditions that affect the observers' spontaneous motoric and symbolic mediation, and the relationship between these mediators, in an observational learning situation.

The authors are listed in alphabetical order. The authors involved in each experiment are identified separately in the report of each experiment.

Requests for reprints should be sent to Seymour M. Berger, Department of Psychology, University of Massachusetts, Amherst, Massachusetts 01003.

Experiment 1: Letter Code Availability and Learning Instructions [1]

Bandura (1974) argues that observers represent a model's behavior through symbolic codes and that these codes guide subsequent motoric reproduction of the model's behavior. This analysis assumes that observers are able to code and that the response components already exist in the observers' behavioral repertoire. If observers cannot symbolically code because there is insufficient time for them to employ appropriate symbolic codes, then observers may attempt some other form of mediation, such as mimicry, in order to represent

[1] The authors of Experiment 1 were Seymour M. Berger and Kathy S. Hammersla. A portion of this experiment was presented under the general title of this article at the annual convention of the American Psychological Association, Toronto, Canada, in August 1978. This project was formulated, with the assistance of Norbert Vanbeselaere, while the first author was a Fulbright–Hays Senior Research Scholar at the Laboratory for Experimental Social Psychology, Catholic University of Leuven, Belgium. The authors wish to thank Margo McMahon for her assistance in selecting the hand signals and in checking the experimenter's reliability.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3705-0735$00.75

the model's behavior. Furthermore, if the response components of the model's behavior are not part of the observers' repertoire, these components may be acquired during the exposure period through the observers' mimetic practice. Thus, mimicry may facilitate learning for observers who are unable to symbolically code and/or unable to perform the appropriate response components.

When observers have symbolic codes and appropriate response components available, there appears to be no need for them to actually mimic the model; in this case, symbolic codes would *compensate* or substitute for mimicry, and mimicry would be reduced or not occur at all. It also is conceivable, however, that motoric and symbolic mediators *complement* rather than compensate for one another. In this case, observers would engage in both types of mediation. Motoric and symbolic mediators may be elicited directly by the model's behavior, or observers may use both types of mediation because they have found, through past experience, that using multiple mediators is more effective for learning than relying on a single form of mediation.

Half of the observers in this experiment were familiar with the manual alphabet for the deaf and half were not. It has been shown that observers tend to overtly mimic these gestures and that such mimicry is positively correlated with observer learning (e.g., Berger, 1966). If the compensatory theory of mediation is correct, then observers who know the manual alphabet can use the letter associated with each gesture as a mediator; therefore, they should engage in less overt mimicry than observers who are unfamiliar with the alphabet. If the complementary theory of mediation is correct, then familiarity with the manual alphabet should not be associated with a reduction in overt mimicry.

In addition to varying response familiarity, observers were instructed either to learn from the model's performance or merely to watch. This was done for two reasons: (a) The observers' spontaneous use of motor and symbolic mediators may depend on whether they intend to learn; and (b) we wanted to provide a comparison with previous research.

Method

Subjects. Thirty-two undergraduates at the University of Massachusetts—Amherst were included in this study.[2] Twenty-five of these subjects received course credit for their participation; seven subjects were volunteers from a course in sign language.

Procedure. Subjects were run individually and were assigned alternately to either the learning or watch-only conditions. Since observers tend not to overtly mimic the model if they are being watched (Berger, 1966), observers were led to believe that they were alone.

Briefly, observers were told that the experimenter had inadvertently signed up two subjects for the same time period, but that the observer could watch the experiment over closed-circuit television from an adjacent "recording" room. After being seated in the recording room, half of the observers were told to watch the television monitor and to follow the instructions given to the subject in the other room. Since the experiment was described as a study of nonverbal learning, the experimenter explained that she would return to test the observer after all the learning materials were presented. The other half of the observers were told that they would receive participation credit if they merely watched the experiment on their monitor. (Seven observers could not receive credit. This presented no problem with three subjects who were assigned to the learning instruction condition, since they had volunteered anyway. The remaining four subjects were in the watch-only condition, and they were told they could help the experimenter by evaluating her performance.)

The experimenter explained that she would turn on the camera when she returned to the subject in the adjacent room. Upon returning to that room, she started the videotape player and seated herself at the one-way mirror in order to record the observer's mimicry. The observer was led to believe that the mirror was covered, but the experimenter could see the observer without being noticed.

The videotape showed the experimenter reentering the adjacent room and giving the instructions to the other subject (model). These instructions stated that the subject had to learn pairs of gestures and that the subject would be tested after the pairs were presented. The pairs were presented by means of a slide projector. Each gesture appeared on a separate slide for 5 sec, with a blank 5-sec period inserted between each pair of gestures. The pairs were gestures associated with *X* and *Q*, *R* and *A*, *P* and *Q* and *K* and *B* in the manual alphabet. The observer

[2] Forty-nine subjects were recruited for this study. Eight of these subjects were eliminated because of procedural errors or because procedural requirements could not be met. The remaining 41 subjects formed the pool from which the 32 observers were selected as explained in the procedure.

saw a same-sex model perform each gesture once while it was projected on the screen.

After all the pairs were shown, but before the subject was tested, the experimenter explained that she had to go into the next room for a moment. The videotape was stopped, as if the camera had been turned off. She returned to the recording room and asked the observer to show her all the pairs that the observer could recall. After noting the observer's response, she asked the observer to complete a questionnaire.

Observers were asked to indicate how familiar they were with the manual alphabet and (with the aid of printed drawings of the gestures used) to indicate in separate items which gestures they actually performed, which they formed images of, which they identified by letter, and which they represented in some "other" way (e.g., words, numbers, etc.).[3] The final questions dealt with the observers' understanding of the instructions and whether they were "suspicious about the instructions . . . received at the time they were given or while . . . watching the TV set"; if observers responded affirmatively, they were to answer the following question: "What were you suspicious about?"

According to their questionnaire responses, 16 of the 41 observers knew some or all of the gestures from the manual alphabet. Seven of these observers were volunteers from the sign language course; the remaining nine familiar observers may have been attracted to the experiment because it was described as a study of nonverbal learning involving gestures. Eight of these familiar observers (six females and two males) were instructed to learn, and the remaining eight observers (all females) were instructed to merely watch. In order to obtain equal numbers of unfamiliar observers, eight observers (six females and two males) under instructions to learn and eight observers (six females and two males) under instructions to merely watch were randomly selected from the remaining pool of 25 unfamiliar observers.

Results

The number of different gestures performed was used as a dependent measure in this experiment (and the experiments that follow), rather than total amount of performance, because the former measure was used previously (Berger, 1966) and the two measures are highly correlated (.85). The experimenter's records of different gestures performed correlated .98 ($p < .001$) with an independent assessment by another person during a pretest reliability check and .73 ($p < .001$) with the observers' self-reported mimicry of different gestures in this experiment.

In general, the experimenter recorded more mimicry of different gestures ($M = 5.00$) than observers reported ($M = 3.03$). For observers who were seen to mimic or reported mimicking, this difference is highly significant, $t(24) = 4.40$, $p < .001$. Observers apparently underestimated the number of different signals they performed.

The data did not meet the criteria for analysis of variance (ANOVA). The most serious violation was the lack of variability in certain cells involving unfamiliar observers. Therefore, nonparametric tests were used to verify each significant effect found in the ANOVAS, and only those effects which were supported by the nonparametric tests are reported below. The results of the F tests are reported, and the significance levels obtained from the nonparametric tests are reported in parentheses next to F-test levels. All significance tests are two-tailed.

Self-reported mediation. The mixed design ANOVA revealed that familiar observers engaged in more total mediation than unfamiliar observers, $F(1, 28) = 34.16$, $p < .001 (< .002)$, and that there were differences, $F(3, 84) = 7.56$, $p < .001 (< .01)$, among the observers' uses of the various mediators (i.e., mimicry, images, letters, or "other"). These main effects, however, have to be interpreted in terms of the Mediators × Familiarity, $F(3, 84) = 3.30$, $p < .05$, and the Mediator × Learning Instruction, $F(3, 84) = 4.04$, $p < .01$, interactions.

The means for the Mediation × Familiarity interaction are reported in Table 1, along with the percentages of observers who engaged in each type of mediation. A Newman-Keuls test of these means indicated that familiar and unfamiliar observers did not differ in their mimicry or in their use of "other" mediators. Familiar observers used more im-

[3] There were two series of questions regarding the observers' mediational activities. The first series asked about mediation during the video presentation, whereas the second series dealt with mediation following the presentation. Since the observers' mimicry following the presentation may have been influenced by the observers' expectation that the experimenter would return and see them performing, the questionnaire responses to the second series were excluded from analysis in Experiments 1 and 2 because they could have been influenced by this situational artifact.

Table 1
Mediation × Familiarity Interaction: Means and Percentages

	Mediator									
	Mimicry		Images		Letters		Other		Total	
Observer	*M*	%	*M*	%	*M*	%	*M*	%	*M*	%
Unfamiliar	2.81	69	.44	13	0	0	.31	13	.89	8
Familiar	3.25	69	3.38	75	2.75	63	1.25	44	2.66	10

Note. There were eight gestures, so each mean could range between 0 and 8.

ages, $\alpha = .01(<.02)$, and more letters, $\alpha = .01(<.002)$, than unfamiliar observers.

Familiar observers engaged in more mimetic and imaginal mediation than "other" forms of mediation, $\alpha = .05(<.05)$, for both comparisons. None of the other differences between mediators for familiar observers was significant.

Unfamiliar observers used mimicry more than images, letters, or "other" mediators, $\alpha = .01(<.01)$, for each comparison. There were no significant differences between images, letters, and "other" mediators for unfamiliar observers. These observers tended to employ mimicry rather than symbolic mediators.

This analysis supports the complementary theory of mediation, because symbolic coding did not reduce motor mimicry. Familiar and unfamiliar observers engaged in about the same amount of mimicry even though familiar observers employed significantly more imaginal and letter coding.

The Newman-Keuls test showed that the Mediator × Learning Instruction interaction was produced by observers who were instructed to learn engaging in more mimicry than any other form of mediation by these observers or by the watch-only observers, $\alpha = .01(<.01)$, for all comparisons. Table 2 presents the means for this interaction along with the percentages of observers who engaged in each form of mediation. There were no differences in symbolic mediation as a function of learning instructions. Since the instruction to learn increased motor mimicry but not symbolic mediation, it appears that observers believed that mimicry would facilitate their learning more than symbolic mediation.

Observed mimicry. The ANOVA based on the experimenter's records revealed only a significant main effect of learning instructions, $F(1, 28) = 20.14$, $p < .001(<.002)$, and no interaction with familiarity. Observers who were instructed to learn performed more different gestures ($M = 7.06$) than the watch-only observers ($M = 2.94$). This finding supports the results of the self-report analysis regarding the effect of instructions to learn.

Observer learning. The ANOVA of the number of pairs learned yielded only a main effect for familiarity, $F(1, 28) = 10.51$, $p < .005(<.02)$, and no interaction with learning instructions. Familiar observers learned more

Table 2
Mediation × Learning Instruction Interaction: Means and Percentages

	Mediator									
	Mimicry		Images		Letters		Other		Total	
Condition	*M*	%	*M*	%	*M*	%	*M*	%	*M*	%
Watch only	1.62	44	1.88	50	1.06	25	.94	31	1.38	81
Learning instructions	4.44	94	1.94	38	1.69	38	.62	19	2.17	100

Note. There were eight gestures, so each mean could range between 0 and 8.

pairs ($M = 1.69$) than unfamiliar observers ($M = .44$).

Correlations between mediators and between mediators and number of pairs learned. None of the correlations between the self-reported mediators was significant. Unfamiliar observers' self-reported mimicry correlated .62 ($p < .05$) with pair learning, and these observers' observed mimicry correlated .53 ($p < .05$) with pair learning. For familiar observers, letter coding correlated .55 ($p < .05$) with pair learning, and their imaginal coding correlated $-.51$ ($p < .05$) with pair learning. Because of a lack of variability, it was not possible to correlate unfamiliar observers' symbolic coding (i.e., images, letters, and "others") with pair learning. Familiar observers' "other" coding, self-reported mimicry, and observed mimicry were not significantly correlated with pair learning.

These correlations suggest that in the absence of other forms of coding and well-practiced responses, motor mimicry prevails as an effective mediator in learning; but given a readily available code that is associated with well-practiced responses, symbolic coding is more effective than mimicry.

Manipulation checks and suspiciousness. Most observers correctly recalled the instructions given them. Approximately one third of the observers instructed to learn and one third of the observers instructed to merely watch reported being suspicious. The major sources of suspicion were the cover story, particularly with regard to whether the other subject was real and whether observers were being watched. None of the observers guessed the actual purpose of the study.

Discussion

The finding that unfamiliar observers relied almost exclusively on motoric mediation contrasts sharply with the familiar observers' use of a variety of mediators. Unfamiliar observers could have used imaginal and other forms of coding, but they did not do so. A possible explanation for the familiar observers' mediational behavior, as well as for similar findings in the studies that follow, will be presented in the final section of this article.

There was a significant negative correlation between imaginal coding and learning, and a tendency (which was not significant) for imaginal coding to be negatively correlated ($-.44$) with letter coding for familiar observers. Apparently, letter coding was a more effective form of mediation than imaginal coding for these observers. Perhaps familiar observers who chose to code imaginally did not have time to letter code as well, and their learning suffered as a consequence. Other studies report that imaginal coding facilitates observational learning (e.g., Gerst, 1971; Ito, 1975; Jeffery, 1976). These studies may have provided more time for mediation, but they also differed from the present study in several other ways. Therefore, it is not possible to identify the basis for these inconsistent findings.

Experiment 2: Symbolic Mediation With Relatively Available Verbal Codes [4]

In Experiment 1, familiar observers' learning was positively correlated with letter coding. In order to determine whether this correlation reflected a causal relationship, an attempt was made in Experiment 2 to bring the observers' use of verbal coding under experimental control by either exposing observers to a verbal coding model (verbal coding condition) or instructing them to use verbal codes to help them learn the model's behavior (instruction condition).

Exposure to a model who engages in overt verbal coding may have one or more consequences for observer mediation and learning. The model's coding behavior may suggest a learning strategy to observers who might not otherwise use verbal coding. Furthermore, the model's coding behavior may provide useful codes to observers who intend to use verbal codes but cannot think of any. Finally, it also is possible that the model's coding behaviors will elicit unintentional imitation and, thereby, facilitate observational learning. No matter which of these processes is involved, there is

[4] The authors of Experiment 2 were Seymour M. Berger, Linda L. Carli, Judith F. Karshmer, M. Estela Sanchez, and Kathy S. Hammersla.

ample reason to believe that a verbal coding model will increase the observers' verbal coding.

Instructing observers to code verbally is another way to increase the observers' coding behavior, provided that they have an available verbal code. Since several studies of observational learning have used such instructions (e.g., Bandura et al., 1966; Bandura & Jeffery, 1973; Gerst, 1971; Jeffery, 1976; van Hekken, 1969), it seemed desirable to compare this treatment with treatments involving a verbal coding model and a nonverbal coding model.

A secondary purpose of this investigation was to demonstrate that the mimicry phenomenon has some degree of generality. Instead of gestures from the manual alphabet, gestures commonly used in United States culture were selected, because all observers probably had performed these gestures previously and would have a readily available word or phrase that they could associate with each gesture (e.g., "come here," "good-bye," "stop"). Thus, observers in this experiment were comparable to familiar observers in Experiment 1 in that they had a highly salient symbolic code available for mediation.

It was hypothesized that observers exposed to a model who performed the gestures but did not engage in overt verbal coding (nonverbal coding condition) would use less verbal coding and would recall fewer pairs and fewer responses than observers in either the verbal coding or instruction conditions. It could be argued, of course, that almost all (if not all) observers would respond to the gestures with the verbal codes regardless of the treatments they received, because the verbal codes are readily elicited when the gestures are performed. This ceiling effect was not expected to occur, however, because the model performed the responses relatively rapidly, thus reducing the observers' time to react with an appropriate verbal code. No differences were expected between conditions in the amount of mimicry, since the model's performance was the same in all conditions and Experiment 1 demonstrated that observer mimicry was independent of the observers' use of symbolic coding.

Method

Subjects. Sixty undergraduates were recruited from a variety of undergraduate courses in psychology at the University of Massachusetts—Amherst.[5] Subjects received extra course credit for their participation.

Procedure. Observers were run individually by one of four different female experimenters. Each experimenter ran five subjects in each of the three conditions.

The procedure for this experiment was similar to that used in Experiment 1. Observers were asked to learn pairs of gestures and were told that two subjects were being run at the same time for purposes of efficiency. The gestures would be presented by means of videotape to the subject in the adjacent room, but the observer would be able to see the presentation over closed-circuit television, because a camera had been installed in the adjacent room. The observer was seated in front of a small television monitor located directly in front of a one-way mirror that appeared to be completely covered.

After giving the instructions, the experimenter left the room, explaining that she had to turn on the camera and the videotape recorder from a third room (observers saw this room and a videotape deck when they arrived for the experiment). She stated that she would return to test the observer after the presentation. The experimenter actually went to the adjacent room, turned on the videotaped model presentation, and positioned herself behind the one-way mirror, which afforded her a clear view of the observer through a small uncovered space at the bottom of the mirror.

After recording any gestures performed, she turned off the videotape, returned to the observer's room, asked the observer to perform the pairs of gestures, and recorded all gestures correctly performed. Observers then completed a questionnaire similar to the one described in Experiment 1, but with drawings of the new gestures. On this questionnaire, observers also were asked to specify which gestures they associated with words and the words they used.

The videotaped gesture presentation began with a medium close-up picture of a large television set on top of a desk. This videotaped television set was used to present additional instructions and gestures, which were actually performed. The instructions, given by a male voice, indicated that the experiment was about to begin. The subject would see three pairs of gestures. Each gesture would be presented separately, and the first member of each pair would be followed by the second member of that pair, with a blank

[5] Because incomplete records were kept regarding the sex of the subjects, the distribution of males and females cannot be given with complete confidence. It appears that there were an equal number of males and females in each condition. Nine additional subjects were run, but these were excluded from the analysis either because their data were uninterpretable or because they did not meet procedural requirements

space appearing between the pairs of gestures. At the end of these instructions, the male voice asked the subject to get ready. At this point, the female model (presumably, the other subject) moved into the picture that the observer saw, as if she were leaning toward the television set in front of her to get a better view. In all conditions, this model performed each of the gestures immediately after it was presented.

The six gestures used were culturally common signs for "come here" (a slightly closed hand with a beckoning index finger), "OK" (index finger and thumb forming a circle, with the remaining fingers vertical and slightly separated), "good-bye" (thumb stationary while the remaining fingers bent down and up), "you" (index finger pointing directly toward the viewer, with remaining fingers forming a fist), "stop" (palm facing viewer, with all fingers vertical, thrust toward the viewer), and "two" (index and middle finger forming a vertical *V*, with thumb overlapping remaining bent fingers).

The gestures were presented in this sequence so that each gesture appeared on the screen for about 2 sec. There was a 2-sec interval between pair members and a 3½-sec interval between the pairs.

Observers were assigned to one of three conditions in the order in which they appeared for the experiment. Observers assigned to the verbal coding condition (VC) heard the female model say the word associated with the gesture as she performed each gesture. Observers assigned to either the nonverbal coding condition (NC) or the instruction condition (IC) saw a tape in which the model performed each gesture but did not say anything; however, IC observers received the following instruction from the experimenter prior to the videotaped presentation: "Each gesture can be associated with a common English word. Use the word you associate with each gesture to help you learn the pairs."

Results

The reliability of the four experimenters' records of the number of different gestures performed was assessed by having two pairs of these experimenters watch a different set of observers. The reliability correlation for each pair of experimenters was .99 ($p < .001$), based on samples of 10 and 15 observers. The experimenters' records correlated .77 ($p < .001$) with the observers' self-reported mimicry of different gestures.

As in Experiment 1, observed mimicry of different gestures ($M = 4.13$) was greater than self-reported mimicry ($M = 3.37$). This difference, based on observers who mimicked (observed or reported), was significant, $t(39) = 2.63$, $p < .02$.

The observers' self-reported use of "other" mediators (e.g., numbers, letters, etc.) was not included in the data analysis because only four observers reported using "other" mediators (one in VC, one in IC, and two in NC). This also eliminated the ANOVA problem encountered in Experiment 1, because of cells having little or no variability. All significance tests are two-tailed.

Self-reported and observed mediation. The mixed design ANOVA yielded a main effect for mediators, $F(2, 114) = 5.89$, $p < .01$, and a significant Conditions × Mediators interaction, $F(4, 114) = 3.97$, $p < .01$. The means for these significant effects are reported in Table 3, along with the percentages of observers using the mediators.

The main effect for mediators must be interpreted in terms of the Conditions × Mediators interaction. In order to carry out 18 preplanned comparisons of the means involved in this interaction, the Bonferroni *t* (see Myers, 1978) familywise error rate was set at

Table 3
Means and Percentages for Self-Reported Mediation: Experiment 2

Condition	Mediator: Mimicry *M*	Mimicry %	Images *M*	Images %	Words *M*	Words %	Total *M*	Total %
Verbal coding	2.75	65	2.65	60	4.50	95	3.30	100
Nonverbal coding	3.85	90	1.80	65	2.25	75	2.63	100
Instruction	3.50	90	3.00	75	4.20	90	3.57	100
Total	3.37	82	2.48	67	3.65	87	3.17	100

Note. There were six gestures, so each mean could range between 0 and 6.

.10; thus the alpha level for any specific comparison was .006. Looking first at the pattern of mediation *within* each condition, one can see that VC observers used significantly more words than either images or mimicry; the difference between images and mimicry was not significant. The NC observers used significantly more mimicry than images; their use of mimicry and images did not differ significantly from their use of words. There were no significant differences in the IC observers' use of these mediators. Thus, each condition manifests a somewhat different pattern of self-reported mediation; the observers' use of images tended to be relatively low in all conditions, which accounts for the mediator main effect.

The comparisons *across* conditions for particular mediators are most relevant to the main purpose of the study. As predicted, NC observers used significantly fewer words than either VC or IC observers; the difference between VC and IC observers' use of words was not significant. There were no significant differences in the comparisons between conditions for images or for mimicry. (This is a somewhat conservative test, because comparisons within the rows are not independent of the comparisons within the columns of means reported in Table 3.)

An ANOVA using the experimenters' reports of observed mimicry also showed no significant difference between conditions, $F(2, 57) = 1.26$. Therefore, the hypothesis regarding mediation is supported, because exposure to a verbal coding model or being given instructions to use a verbal code facilitated the observers' use of verbal coding as compared to observers who were exposed to a nonverbal coding model. Also, it is clear that observers spontaneously mimic these cultural gestures, so mimicry is not limited to gestures from the manual alphabet.

A separate analysis was conducted to determine whether observers who used verbal codes tended to use *exactly* the same words for gestures as the VC model. The mean percentage of words that were exactly the same as the VC model's words was 90% for VC observers, 66% for NC observers, and 67% for IC observers.

Observer learning. The ANOVA compar the number of pairs learned between cond tions yielded $F(2, 57) = 2.79$, which is no significant. Nevertheless, comparisons betwee pairs of conditions were carried out, becaus specific predictions were made regarding th pattern of differences. The trend in the mea was as predicted. The mean number of pai learned by NC observers (1.05) was less tha the mean number of pairs learned by VC ob servers (1.55) and by IC observers (1.75 However, the Bonferroni t with a familywis error rate set at .10 for the comparison o three means revealed that only the differenc between the number of pairs learned by N and IC observers was significant ($\alpha = .03$ f each specific comparison).

A second measure of learning involve merely counting the number of different ge tures recalled by each observer, regardless o whether they were properly paired. Th ANOVA for this measure of learning yielded highly significant conditions effect, $F(2, 5$ $= 8.55$, $p < .001$. The NC observers recalle significantly fewer gestures (3.95) than eithe the VC observers (4.85) or the IC observer (5.00) (the VC and IC observers' learnin was not significantly different). These wer highly significant differences: The error ra for each comparison was less than .003, usin the Bonferroni t. The learning hypothesis i strongly supported by this analysis of num ber of gestures learned and is partially sup ported by the analysis of pair learning.

Manipulation checks and suspiciousnes The manipulation checks indicated that mos observers correctly recalled the instruction they received and the model's performance gestures and words (in the VC condition).

As in Experiment 1, approximately on third of the observers indicated that the were suspicious. These observers were aske "What were you suspicious about and whe did you become suspicious?" Most of thes observers questioned whether the other su ject was real. Some observers reported th they were suspicious about all psychologic experiments. A few observers thought some one might be watching them; two of these thought that someone might be watching see how they learned.

scussion

The lack of a clear conditions main effect for number of pairs learned is somewhat surprising given the highly significant conditions main effect for number of gestures learned. Perhaps a stronger conditions effect for pair learning might have been achieved if the learning task had included more pairs, longer exposure to the gestures, or a longer interval between the pairs of gestures. Several observers complained that the exposure time was too short for them to think of words and to make the necessary associations; other observers complained that the distinction between pairs was not clear in the videotaped presentation.

Experiment 3: Symbolic Mediation With Relatively Unavailable Verbal Codes [6]

In Experiments 1 and 2, the use of symbolic mediators (i.e., letters and words, respectively) was related to increased learning. This relationship existed only when observers had symbolic codes available for the model's behavior. Experiment 3 investigated the influence of verbal mediators when observers had no prior symbolic (verbal) code for the model's behavior.

In Experiment 1, the unfamiliar observers did not engage in symbolic coding; their learning was related to their mimicry of the model's responses. Without prior practice in performing the model's responses and without salient verbal codes for these responses, it seemed unlikely that unfamiliar observers would profit from the use of symbolic codes.

In order to test this hypothesis, observers who were unfamiliar with the manual alphabet were instructed to learn certain pairs of gestures from the alphabet. Half of the observers also were instructed to associate words with gestures to help them learn (instructed condition); the remaining observers did not receive these instructions (noninstructed condition). Observers in the instructed condition were expected to engage in more verbal mediation than observers in the noninstructed condition, but no differences were expected between these conditions in learning. Furthermore, since mimicry was found to be unaffected by verbal mediation in Experiments 1 and 2, no differences in mimicry were expected. Observers in the instructed condition, however, were expected to engage in more imaginal and "other" types of coding than those in the noninstructed condition, because the results of Experiments 1 and 2 seemed to indicate that those observers who engaged in symbolic coding also engaged in imaginal and "other" forms of coding.

Method

Subjects. Forty undergraduates were recruited from a variety of undergraduate summer session courses at the University of Massachusetts—Amherst.[7] All subjects received course credit for their participation.

Procedure. Observers were run individually by one of three different female experimenters. Two of these experimenters each ran 7 observers, and one ran 6 observers in each condition, so that there were 20 observers in each condition. The recruiting folder stated that the study involved gesture learning and that students who were familiar with the manual alphabet for the deaf should not participate.

Observers were assigned to conditions in the order in which they appeared for the experiment. All observers were brought into the subject room and told that this was a study of nonverbal learning. They were asked to learn pairs of gestures that would be presented by means of videotape. It was explained that the videotape showed a subject in a previous experiment. Only one tape was used—the one employed in Experiment 1 showing a female model. The instructions given to all observers were those used in the learning condition in Experiment 1.

The 6 male and 14 female observers assigned to the instructed condition (IC) were given the following additional instruction: "Previous experiments have shown that if you associate words with gestures it will help you learn. So please use this form of verbal coding to help you learn." The remaining 7 male and 13 female noninstructed observers (NC) did not receive this added instruction.

The one-way mirror in the room was completely covered by a fiberboard. The videotape deck was located in the subject's room alongside the small television monitor. The experimenter explained that she would turn on the deck and leave the room so as not to distract the observer during the presentation and

---

[6] The authors of Experiment 3 were Seymour M. Berger, M. Estela Sanchez, Judith F. Krashmer, and Linda L. Carli.

[7] Two additional subjects were excluded from the data analysis because they did not meet the procedural requirements.

Table 4
Means and Percentages for Self-Reported Mediation: Experiment 3

Condition	Mediator									
	Mimicry		Images		Words		Other		Total	
	M	%	*M*	%	*M*	%	*M*	%	*M*	%
Noninstructed	5.50	100	2.70	55	.40	15	1.65	35	2.56	100
Instructed	4.55	80	2.40	55	3.25	65	1.20	30	2.85	100
Total	5.03	90	2.55	55	1.83	40	1.43	33	2.71	100

Note. There were eight gestures, so each mean could range between 0 and 8.

that she would return after the presentation was over to test the observer's learning of the pairs of gestures. If there were no questions, the experimenter started the videotape and left the room, leaving the door slightly ajar. In this way, she could hear when the presentation ended and return to the room to test the observer.

After the test, in which she recorded each gesture correctly performed, the experimenter asked the observer to complete the questionnaire. The questionnaire was the same as that used in Experiment 2, except that there were no questions about mediation after the television presentation; the appropriate gestures from the manual alphabet were substituted for cultural gestures.

Results

Self-reported mediation. The mixed design ANOVA revealed a significant main effect for mediators, $F(3, 114) = 17.76$, $p < .001$, and a significant Conditions × Mediators interaction, $F(3, 114) = 5.10$, $p < .005$. The means and percentages involved in these comparisons are reported in Table 4.

The Conditions × Mediators interaction indicates that both conditions showed a similar pattern of mediation except with regard to the use of words. As expected, the IC observers used significantly more words than the NC observers. Within both conditions, use of mimicry was greater than use of images or "others," and the difference between the use of images and the use of "others" was not significant. For NC observers the level of imaginal coding was significantly greater than the level of verbal coding; for IC observers the level of verbal coding did not differ significantly from either the level of imaginal or "others" coding. These differences are based on a Bonferroni t test with a familywise error rate set at .10, so that the alpha level was .00 for the 16 specific comparisons.

Observers' learning. Separate ANOVAs for the two learning measures (i.e., numbers of pairs learned and number of gestures learned) did not yield significant differences between conditions. In pair learning, the mean number of pairs learned for NC observers was 1.25; for IC observers it was 1.35. The mean for number of gestures learned was 4.90 for NC and 5.40 for IC observers.

Manipulation checks and suspiciousness. The manipulation checks revealed that the vast majority of observers correctly recalled the instructions they received and that the model had performed the gestures. Only five observers were suspicious about the instructions. Three of these observers were suspicious because they were not given sufficient justification for the learning task, and two were suspicious because the distinction between pairs of gestures in the videotaped presentation was not clear.

Discussion

Given the absence of both imaginal and "other" forms of coding found in Experiment 1 for unfamiliar observers, the occurrence of both of these forms of coding in Experiment 3 among NC observers was unexpected. Since observers in Experiment 3 were told that they were watching a videotape of a previous subject, perhaps the experimenter's selection of this particular subject's performance implied a demand to mimic as well as approval for

the use of coding strategies in general. Another possibility is that observers pay more attention to a "live" model (Experiment 1) and, therefore, are somewhat distracted from considering or engaging in symbolic coding when they have no available symbolic code. Finally, the summer school undergraduates recruited for Experiment 3 may have been more motivated or more skilled in learning strategies than regular-session students; several subjects were somewhat older than traditional college undergraduates. Whether mode of presentation or subject selection accounts for these differences in mediation between Experiments 1 and 3 cannot be determined without further investigation.

General Discussion

Although observational learning was related to symbolic coding (i.e., letters in Experiment 1 and words in Experiment 2) among observers with a salient code for the model's behavior, these observers also engaged in a considerable amount of mimicry and imaginal coding that had no *apparent* relationship to their learning. It is clear from the observers' questionnaire responses that these observers mimicked and imaginally coded because they believed that these forms of mediation would help them learn.

Since these observers had learned to perform the gestures previously, it is not surprising that the additional mimetic practice failed to directly enhance their learning. It is possible that mimicry was an unintentional reaction to the model's performance, because the model performed the gestures in all conditions. This interpretation would account for mimicry but not for imaginal coding.

Perhaps a dual or multiple-stage coding process is involved, similar to the one suggested by Bandura and Jeffery (1973) and Jeffery (1976). For observers with a highly salient and well-learned letter or word code, it may be more efficient for them to store and retrieve the model's behavior in terms of these symbolic codes. Use of mimicry (or other nonverbal codes) may help observers retain, temporarily, those model responses which they are unable to letter- or word-code immediately. Thus, using mimicry (or other nonverbal codes) provides more time for observers to employ their well-learned letter or word codes for these behaviors.

It also is possible that the self-report measures of mediation are not valid indexes of the observers' actual mediational behavior. Although substantial correlations were found in Experiments 1 and 2 between observed and self-reported mimicry, self-reported mimicry was significantly lower than observed mimicry. Furthermore, even observed mimicry tends to underestimate the total amount of relevant motor activity in this type of situation, as revealed by electromyographic recordings (Berger et al., 1970). Nevertheless, even though observed and self-reported mimicry measures probably lack absolute accuracy, these measures are assumed to be reasonably sensitive to relative differences in the observers' mimetic activities. They both show comparable relationships with observational learning in Experiments 1 and 2 (where both measures were obtained). If the self-report measure of mimicry is considered to have some degree of validity, then it is reasonable to assume that self-reports of other forms of mediation have some degree of validity as well.

Finally, two additional points should be noted. The correlations reported in these studies probably are attenuated because of the low variability in the measures of mediation and learning. The test of learning was administered within a few minutes after the observers' exposure to the model. The long-term effects of mediation in learning, which are of considerable interest to observational learning theorists, were not examined in these experiments.

References

Bandura, A. *Psychological modeling: Conflicting theories*. New York: Lieber-Atherton, 1974.

Bandura, A., Grusec, J. E., & Menlove, F. L. Observational learning as a function of symbolization and incentive set. *Child Development*, 1966, *37*, 499–506.

Bandura, A., & Jeffery, R. W. Role of symbolic coding and rehearsal processes in observational learning. *Journal of Personality and Social Psychology*, 1973, *26*, 122–130.

Berger, S. M. Observer practice and learning during exposure to a model. *Journal of Personality and Social Psychology*, 1966, *3*, 696–701.

Berger, S. M. Vicarious aspects of matched-dependent behavior. In E. C. Simmel, R. A. Hoppe, & G. A. Milton (Eds.), *Social facilitation and imitative behavior*. Boston: Allyn & Bacon, 1968.

Berger, S. M., Irwin, D. S., & Frommer, G. P. Electromyographic activity during observational learning. *American Journal of Psychology*, 1970, *83*, 86–94.

Bernal, G., & Berger, S. M. Vicarious eyelid conditioning. *Journal of Personality and Social Psychology*, 1976, *34*, 62–68.

Gerst, M. S. Symbolic coding processes in observational learning. *Journal of Personality and Social Psychology*, 1971, *19*, 9–17.

Ito, H. An analysis of mediation processes in ob servational learning: A comparison of imaginal an verbal mediation. *Japanese Psychological Re search*, 1975, *17*, 182–191.

Jeffery, R. W. The influence of symbolic and moto rehearsal in observational learning. *Journal o Research in Personality*, 1976, *10*, 116–127.

Myers, J. L. *Fundamentals of experimental desig* (3rd ed.). Boston: Allyn & Bacon, 1978.

van Hekken, S. M. J. The influence of verbalizatio on observational learning in a group of mediatin and a group of non-mediating children. *Human Development*, 1969, *12*, 204–213.

Received February 20, 1978

Journal of Personality and Social Psychology
1979, Vol. 37, No. 5, 747-755

A Step in the Socialization of Altruism as Hedonism: Effects of Negative Mood on Children's Generosity Under Public and Private Conditions

Douglas T. Kenrick
Montana State University

Donald J. Baumann and Robert B. Cialdini
Arizona State University

Prior research has indicated that although negative mood-induction procedures reliably lead to enhanced helping in adults, such procedures do not produce increased helping in young children. Consistent with the negative state relief model, it was expected that, relative to neutral mood subjects, children in a negative mood would be more generous if the helping opportunity offered the potential for direct reward through social approval. This expectation was supported in a pair of studies wherein first- through third-grade children were asked to imagine either neutral or sad experiences and were then given the opportunity to be charitable either in public or in private. Experiment 2 provided evidence that the enhanced public helping of negative mood subjects is more parsimoniously interpreted as an attempt to remove negative mood than to repair public image. Finally, a three-step account of the development of altruism as a self-reward is proposed.

A large number of studies have shown increased helping as a function of experiences that are likely to have induced negative moods in adult subjects (e.g., Aderman & Berkowitz, 1970; Apsler, 1975; Filter & Gross, 1975; Freedman & Doob, 1968; Freedman, Wallington & Bless, 1967; Greenglass, 1969; Kidd & Berkowitz, 1976; Konečni, 1972; Rawlings, 1968; Steele, 1975). Cialdini and his associates (Cialdini, Darby & Vincent, 1973; Cialdini & Kenrick, 1976) have suggested a negative state relief model of altruism to explain these data. According to this notion, altruistic behavior is self-gratifying and will serve to alleviate temporary negative mood states in the same way that any other mood-enhancing experience will. Such a notion is consistent with research demonstrating that negative mood will increase the likelihood of other forms of self-reward (Rosenhan, Underwood, & Moore, 1974; Underwood, Moore, & Rosenhan, 1973) and receives direct support from the study by Cialdini et al. (1973). These authors found that although a negative mood-inducing experience generally led to increased helping, this relationship did not hold for subjects who received a gratifying event (money or praise) between the negative experience and the helping opportunity. Apparently, the negative affective state was cancelled for these subjects by the application of a reinforcer, and it was no longer necessary to employ the helping response for that purpose.

In contrast to the studies showing a general tendency for mood-lowering procedures to lead to increased helping, there is a body of data indicating that such procedures typically lead to some degree of decreased helping (e.g., Isen, Horn, & Rosenhan, 1973,

The authors are grateful to Nancy Eisenberg-Berg, Susan Somerville, and Antoinette Zeiss for their critical comments on an earlier version of this article. Thanks are also due the administration of Kyrene del Norte and Broadmor Elementary Schools, Tempe, Arizona, especially to their principals, Fredrick Gallow and Herbert Hackett.

Requests for reprints should be addressed to Douglas T. Kenrick, Department of Psychology, Montana State University, Bozeman, Montana 59717.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3705-0747$00.75

Studies 1 and 3; Moore, Underwood, & Rosenhan, 1973; Rosenhan et al., 1974; Underwood, Froming, & Moore, 1977). Cialdini and Kenrick (1976) sought to reconcile these seemingly discrepant findings by pointing out that the data showing a direct relationship between lowered mood and helping have come from adult subjects, whereas the data failing to show such a relationship have generally come from work on children. Cialdini and Kenrick reviewed evidence to support the position that altruism acquires its self-reinforcing qualities as a function of the socialization process, wherein helping behavior is systematically paired with various forms of direct reinforcement (e.g., praise, reciprocated aid, thanks, etc.). On the basis of that review, it was expected that young children who had not completed the socialization experience would not find helping behavior gratifying and, hence, would not use it to dispel a negative mood state. However, it was expected that with age, individuals would increasingly employ prosocial activity as a self-gratifier when in a negative mood. In support of these expectations, Cialdini and Kenrick found that the relationship between negative mood and helping progressively changed from an inverse to a direct one over three groups of increasingly older subjects, only the oldest (high school students) of whom showed significantly enhanced helping under negative mood.

Normative Versus Autonomous Altruism

One implication of the negative state relief formulation is that young children will not show enhanced helping while in a negative affective state unless the helping leads to some kind of direct, extrinsic reward (e.g., social approval). Rosenhan (1970) has suggested that altruism can be divided into two categories: normative altruism, based on sanctions associated with social norms, and autonomous altruism, which occurs in the absence of external reward and which must be internalized. However, in the above-cited studies investigating the influence of negative mood on children's helping, only autonomous altruism has been examined, since the opportunity to behave charitably has always been anonymous. In keeping with the negative state relief formulation, we would expect that, were the helping situation to provide the opportunity for normative social reward, negative mood would increase help giving in young children.

There are data to suggest that primary school children are aware of the normative sanctions for helping. For example, Ugurel-Semin (1952) found in 6- to 9-year-old children an increasing recognition of the pressures exerted by norms and customs on the decision to share. Similarly, Eisenberg-Berg (in press) found a substantial tendency for 7- and 9-year-olds to employ social approval reasons as justification for prosocial action. Further, Bryan and Walbek (1970) found that the majority of their 8- to 9-year-old subjects verbally indicated to an adult experimenter knowledge of the social desirability of helping. These authors found, however, that their subjects' verbalizations about the appropriateness of helping did not predict their actual generosity in the absence of the experimenter. Thus, it appears that while young children are aware that social approval is likely to follow their prosocial activity, they have not yet internalized the norms for helping.

The present investigation was designed to test the hypothesis that primary school children should show *increased* helping as a function of negative mood induction in a situation involving normative (or publicly observable) altruism, since the potential for external reinforcement would be present in the form of social approval. In Experiment 1, first- through third-grade children were given either a private or a public opportunity to behave charitably. In the public condition, an adult was present during an opportunity for the child to contribute some rewards to other children. In the private condition, no adult was present during this time. In accord with the negative state relief model, it was expected that negative mood would produce significant increases in the helping behavior of children, as it does for adults, but only in the public condition. That is, negative mood would serve to increase normative altruism, but not autonomous altruism, since children

at this age would not be expected to have yet internalized the reward properties of prosocial behavior.

Experiment 1

Method

Subjects. Subjects were 72 first- through third-grade children (35 males, 37 females) from an elementary school in a middle-class suburban area.

Procedure. The procedure used in the present investigation was identical to that used by Cialdini and Kenrick (1976) with two exceptions: the inclusion of a brief check on the mood manipulation and the introduction of a public helping condition.

Briefly, subjects were brought individually to an experimental room containing a table and chairs, books, a tape recorder, and a box labeled "Coupons for other students." Subjects were told that they would be participating in two separate tasks, an initial "hearing task" and a subsequent "imagination task." They were told that for their participation they would receive several coupons that could be exchanged for a valuable prize at the end of the week. It was stressed that larger accumulations of coupons would result in better prizes. The experimenter directed the subjects' attention to the sealed box and explained that they could donate some of their winnings to the other students who would not get a chance to participate in the experiments.

Subjects then performed a hearing task and were given 10 coupons for their participation. The first experimenter then left, and a second experimenter, always of the same sex as the subject, entered and began the imagination task, which constituted the mood manipulation. Each subject received 5 coupons before the onset of this task.

Mood manipulation. Half the subjects (negative mood) were asked to reminisce about a sad experience, and half were asked to imagine a book and a chair (neutral mood). This was done in a manner identical to that of Cialdini and Kenrick (1976), except that following the mood manipulation, subjects were asked, "Did what you just thought about make you feel better than you usually feel, about the same as you usually feel, or worse than you usually feel?" The sequence in which these possibilities were presented was counterbalanced across subjects. Responses were coded on a 1–3 scale, with 1 indicating the most positive response.

Helping opportunity. After the mood manipulation, the second experimenter left the room, and the first experimenter (blind to the subject's mood condition) returned. In the private condition, the experimenter explained that the tasks were done and (while ostensibly resetting the tape recorder) thanked the subject and said there would be a short delay before the subject would be accompanied back to the classroom. The experimenter again called attention to the box labeled "Coupons for other students," reminded the subject of the option to share coupons, and left the subject alone for 90 sec.

In the public condition, the experimenter performed the same behaviors except that instead of leaving after the reminder, he or she remained in the room and observed the subject's behavior during the 90-sec helping opportunity.

For ethical reasons, the experimenter had all negative mood subjects imagine happy events at the completion of the experiment. A second manipulation-check procedure indicated that the negative mood was removed as a result. Subjects were then asked not to discuss the tests with any of their classmates until the end of the week.

Independent variables. Three factors were varied: mood (negative vs. neutral), type of helping opportunity (public vs. private), and sex of subject.

Dependent variables. Two dependent variables were employed. The subjects' responses to the mood manipulation check constituted one dependent variable. The major dependent variable was the mean number of coupons contributed per cell.

Predictions. It was expected that for our young subjects, helping would be substantially enhanced only when they were in a negative mood and the possibility of external reward existed. Consequently, the negative mood/public condition was predicted to produce more helping than the other three conditions, which were not expected to differ among themselves.

Results

A 2 × 2 × 2 analysis of variance revealed that sex of subject had no effect on either dependent variable; thus the means for both the number of coupons contributed and the mood manipulation check are collapsed over sex and are presented in Table 1.

Manipulation check. The results of the analysis of variance on the mood-check item

Table 1
Mood and Helping Scores for Experiment 1

Measure	Type of helping opportunity			
	Private		Public	
	Neutral mood	Negative mood	Neutral mood	Negative mood
Mood check	1.72	2.06	1.50	2.55
M coupons contributed	1.44	1.22	1.50	3.50

Note. Range is 1–3 on mood check (1 = positive mood, 3 = negative mood) and 1–15 on coupons contributed. *n* per cell = 18.

showed the predicted main effect for the mood induction. Mean values (displayed in Table 1) suggest that negative mood subjects indicated more negative feelings as a result of their ideations, $F(1, 64) = 17.08$, $p < .001$. In addition, there was an unexpected interaction of mood induction with type of helping opportunity, $F(1, 64) = 4.32$, $p < .05$, suggesting that the mood manipulation, although producing a similar effect for both public-condition and private-condition subjects, was somewhat stronger in magnitude for subjects who were subsequently given a public helping opportunity. No other effects were significant for this variable.

Coupon sharing behavior. Since the prediction was made that the largest number of coupons would be contributed during a public helping opportunity following a negative mood induction, an a priori contrast (Hays, 1963, p. 466) was performed to test this prediction. This contrast pitted the public negative mood group against the mean of all other groups. In addition, the residual between-groups variance was tested to determine whether any remaining effects due to the independent variables existed. It was expected that the between-groups variance for the a priori contrast would account for the majority of the variance.

Results indicated that the prediction was borne out. Subjects in the negative mood/public condition contributed substantially more coupons than did subjects in the other cells (means are displayed in Table 1). The a priori contrast indicated that this effect was significant, $F(1, 68) = 9.55$, $p < .003$. In addition, this contrast accounted for 99% of the between-groups variance, whereas the remaining comparisons testing all other orthogonal effects were nonsignificant ($F < 1$) and accounted for only 1% of the variance.

Discussion

As expected, subjects who were exposed to negative mood-inducing operations and who had the opportunity to help in the presence of an experimenter donated the most coupons. Negative mood subjects whose opportunity to help occurred in private, however, were not especially likely to share their coupons. In fact, they were slightly less generous as a result of the negative induction procedure, a result that is consistent with the prior literature involving anonymous giving in young children. It appears, then, that negative mood did not lead to increased helping in our young subjects unless the helping opportunity offered the likelihood of direct social reward for generosity. It is our argument that helping will be employed by individuals in a lowered mood state in order to relieve their negative affective state. For adults, who have internalized the reward value of benevolence, the help may occur in private and still retain its self-gratifying function. For young children, on the other hand, altruism has not yet become autonomous and, thus, will be gratifying only when it is instrumental to an extrinsic form of reward like public approval.

Negative mood relief or image repair. The finding that our young subjects were more likely to help when in a negative mood and in public is compatible with some earlier results by Isen et al. (1973). They found (Experiments 2 and 3) that a public failure had a positive effect on young children's charitable donations made in the presence of an experimenter who knew of the failure. The authors accounted for the finding with an image repair explanation: The visible failing produced a lowered public image that subjects then sought to repair via prosocial action. Our own interpretation of those data would be somewhat different: The public failure produced a lowered mood that subjects then sought to remove via prosocial action. Setting aside the question of how to best interpret the Isen et al. results, the image repair notion of those authors presents a potential alternative explanation for the data of the present study. Perhaps, in asking negative mood subjects to reminisce aloud about a sad experience, we caused a large percentage of them to divulge image-damaging things about themselves. If so, the enhanced coupon donating that occurred in the negative mood/public condition may have been a result of attempts to raise a lowered public image rather than to relieve a negative mood.

To test the negative state relief versus im-

age repair interpretations of our findings, a second experiment designed to replicate and extend the results of Experiment 1 was conducted. A test of these alternative formulations gains importance because the image repair interpretation could account for many of the studies showing that certain negative mood procedures resulted in greater helping among adult subjects. Although Cialdini and Kenrick (1976) interpreted such results according to the negative state relief model, increased aid following a public transgression (e.g., Konečni, 1972), receipt of an insult (Steele, 1975), receipt of a deviant score on a personality test (Filter & Gross, 1975), or public embarrassment (Apsler, 1975) could conceivably have occurred in an attempt to restore public image.

A second reason for performing a replication of Experiment 1 involved the unexpected interaction that appeared on our mood manipulation check in that study. Even though the means in Table 1 suggest that the mood induction procedure was successful for subjects who were subsequently randomly assigned to both public and private helping opportunity conditions, the induction appears to have been stronger among those in the public cells. While the pattern of means makes it unlikely that this factor mediated our results, it was felt that confidence in the results of Experiment 1 would be enhanced by a replication showing equivalently strong negative mood induction among public and private condition subjects. To these ends, a replication of Experiment 1 was conducted that included an additional manipulated variable—whether the negative mood-induction procedure required subjects to imagine negative experiences that were image related or image unrelated.

Experiment 2

Method

Subjects. Subjects were 77 first- through third-grade children (42 males, 35 females) from an elementary school in a middle-class suburban area. This was not the same school used in Experiment 1.

Procedure. The procedure was identical to that used in Experiment 1 with the exception that subjects in the negative mood cells were divided into an image and a nonimage condition. These conditions were operationalized in the following way: Experiences spontaneously generated by subjects in Experiment 1 were divided into image and nonimage categories by two independent raters. There was 90% agreement between raters. The four most frequently occurring responses in each category had 100% agreement between raters and were selected for experimental conditions.[1]

Negative mood was manipulated by asking one half of the subjects if they had ever experienced the most frequently occurring non-image-related response, whereas the other half were asked the same question about the most frequently occurring image-related response. If subjects responded "yes" to this first probe, the experimenter proceeded to induce a negative mood state in a manner identical to Experiment 1. If subjects responded "no," the experimenter continued down the list of image or nonimage experiences until an affirmative response was generated. None of the subjects responded "no" to all four probes.

Predictions. It was expected that helping would again be greatest in the negative mood/public condition and that this effect would not vary as a function of the image manipulation.

Results

A $3 \times 2 \times 2$ analysis of variance again revealed that sex of subject had no effect on either dependent variable; thus, the means for these measures were collapsed over sex and can be found in Table 2.

Manipulation check. Once again, the manipulation check showed a significant main effect of the mood manipulation, $F(2, 65) = 20.68$, $p < .001$. In this experiment, neither the type of helping opportunity nor the interaction with mood induction was significant for this variable (both $Fs < 1$). No other effects approached significance.

Coupon sharing behavior. The initial $3 \times 2 \times 2$ analysis of variance also revealed that no significant differences occurred as a function of the image manipulation (all $Fs < 1$). Examination of the means in Table 2 indi-

[1] The nonimage condition contained the following experiences: (a) Have you ever had a pet you owned die? (b) Have you ever had a toy you liked broken, lost, or stolen? (c) Have you ever been hurt in an accident? (d) Have you ever had someone you like very much move away? The image condition included (a) Have you ever been called a bad name? (b) Have you ever been caught doing something wrong? (c) Have you ever been punished for doing something wrong? (d) Have you ever been left out of a game you wanted to be in?

Table 2
Mood and Helping Scores for Experiment 2

Measure	Type of helping opportunity: Private — Neutral mood (n = 13)	Private — Negative mood: Nonimage (n = 12)	Private — Negative mood: Image (n = 11)	Public — Neutral mood (n = 14)	Public — Negative mood: Nonimage (n = 13)	Public — Negative mood: Image (n = 14)
Mood check	1.92	2.66	2.72	1.68	2.69	2.78
M coupons contributed	1.39	2.00	1.00	2.07	3.15	2.93

Note. Range is 1–3 on mood check (1 = positive mood, 3 = negative mood) and 1–15 on coupons contributed.

cates that if anything, image-related negative mood led to relatively less helping than nonimage-related negative mood.

Since there was no difference between the image and the nonimage cells, the means for the major dependent measure were collapsed over type of negative mood induction, and the negative mood/public group was pitted against all other groups, as in Experiment 1. Once again, results indicated the most helping by subjects in the negative mood/public cell (see Table 2).[2] The a priori contrast for this effect was significant, $F(1, 73) = 3.05$, $p < .05$, one-tailed test. Further examination of these results indicated that this contrast accounted for 90% of the between-groups variance, whereas the remaining orthogonal comparisons were nonsignificant ($F < 1$) and accounted for 10% of the variance.

Rosenthal (1978) has recently reviewed a number of methods for combining the results of independent studies, suggesting that the most serviceable technique among those available is the method of adding Z scores (Mosteller & Bush, 1954). When looked at in this manner, the combined results of Experiments 1 and 2 strongly support the hypothesis that the greatest coupon sharing would occur in the negative mood/public condition, $Z = 2.96$, $p < .003$.

General Discussion

Taken together, the results of both experiments provided clear support for our major hypothesis. While primary-grade children in a situation requiring autonomous (i.e., internalized) altruism did not show the adult pattern of increased helping in a negative mood, the pattern shown in the public condition indicated that *normative* (i.e., sanction-based) helping was increased by negative mood. This latter outcome fits well with the findings indicating increased helping in adults (e.g., Apsler, 1975; Cialdini et al., 1973; Freedman et al., 1967) as well as increased direct self-reinforcement in children (Rosenhan et al., 1974; Underwood et al., 1973) as a function of negative mood induction. Our results are also consistent with the Isen et al. (1973, Study 3) finding that an image-damaging incident will lead to an increase in helping, although the present results do not seem to be explained via the image repair model.

It should be noted that the negative state relief model and the image repair model are in no way incompatible. In fact, damage to image is doubtless one common and potent form of negative state induction. It is our position, however, that image repair seems subsumed under the more general negative state relief model. Image damage is, thus, sufficient but not necessary to lead to increases in normative helping. Other negative experi-

[2] As the means in Table 2 also indicate, an analysis of variance revealed a trend toward a main effect for type of helping opportunity, $F(1, 65) = 2.42$, $p < .12$, with public opportunity to help producing somewhat more aid. A similar trend was found in Experiment 1, $F(1, 64) = 3.72$, $p < .058$.

ences will have the same result. Further, research by Cialdini et al. (1973) has indicated that relief of a negative mood state abolishes the negative-mood-leads-to-helping relationship even when such relief is in no way related to image improvement.

Qualifications on the Negative State Relief Model

A further point of clarification regarding the negative state relief model should be made here. The model suggests that one helps when one is experiencing a transitory feeling of sadness or depression, in order to feel better. As indicated in previous statements of the model, (Cialdini et al., 1973; Cialdini & Kenrick, 1976) negative experiences likely to arouse feelings of frustration or anger would not be expected to result in altruistic behavior. Additionally, if the helping is effortful or unlikely to yield great benefit, negative mood might be expected to reduce helping even in adults. This latter suggestion is supported in a recent article by Weyant (1978). He found that a negative mood induction increased helping in his subjects, but only when the cost of helping was low and the potential benefits were high. Conversely, when cost of helping was high but potential benefits were low, helping was somewhat decreased by negative mood.

There is also recent evidence that negative mood can sometimes reduce helping by producing an attentional deficit. McMillen, Sanders, and Solomon (1977) found that negative mood subjects were less attentive to their environment than neutral mood subjects and, consistently, were less helpful to a female confederate in distress, unless she drew attention to herself. In the latter condition, negative mood produced greater helping than either neutral or positive mood conditions.[3] The tendency to engage in altruistic behavior as a means of self-reward following negative state induction, then, seems to be mediated by the reward value and the salience of the helping opportunity. In keeping with the hedonistic basis of the model, it appears that if a helping opportunity is not perceived as likely to lead to a better mood, negative mood will not enhance helping. These findings are consistent with the notion that helping will occur in a negative mood in order to alleviate such aversive feelings.[4]

It was interesting that whereas means for helping in the public conditions were invariably higher than those in the relevant private comparison, the public–private effect was not more pronounced in the neutral mood conditions. It would seem that the children generally judged the value of the unknown prize to be potentially more valuable than the immediate social approval of the experimenter, but that negative mood enhanced the importance of such immediate approval.

The Socialization of Altruism as a Self-Reinforcer: A Three-Step Model

It will be recalled that the Cialdini and Kenrick (1976) data indicate that the relationship between helping and negative mood changes from an inverse to a direct one as an individual passes through the socialization process. When those data are taken in combination with the results of the present investigations, a three-step model of the socialization of altruism as a self-reward emerges. Initially, helpful behavior would be rarely performed by an unsocialized individual, since it involves the loss of rewards. There is substantial evidence that preschool-age children help infrequently and that when they do help, it is unrelated to social approval (Eisenberg-Berg & Hand, in press; Fischer, 1963; Ugurel-Semin, 1952). At this step,

---

[3] Another recent study, by Underwood, Berenson, et al. (1977), failed to find increased helping in negative mood subjects as a function of an attention manipulation, and these authors' suggestions may be compatible with Weyant's (1978) more recent findings. They propose that negative mood subjects may simply not have perceived a great need of help. Unfortunately, since mood induction was accomplished by using subjects who were leaving movie theaters and had observed very different films, it is possible that a number of uncontrolled factors may have been operative in this study. Further research is needed to clarify the contradiction between the Underwood, Berenson, et al. (1977) and McMillen et al. (1977) studies.

[4] We are indebted to Robert A. Baron for his contributions regarding the thinking contained in this paragraph.

helpfulness would be unlikely under conditions of negative mood, since it would be experienced as self-punishment.

The second step involves children's acquisition of the norms for prosocial behavior, without concomitant internalization of these norms. Children at this step, usually in the primary grades, are aware that helping can lead to systematic social reinforcement (Bryan & Walbek, 1970; Eisenberg-Berg, in press; Ugurel-Semin, 1952) and employ a public helping opportunity to obtain external social reward when in a negative mood, as suggested by the data of the present studies. However, such children may still tend to be less helpful under negative mood when the helpfulness is anonymous (Cialdini & Kenrick, 1976; Moore et al., 1973; Rosenhan et al., 1974; Underwood, Froming, & Moore, 1977).

Finally, after sufficient experience with external reward for charitable action, that behavior itself takes on the quality of a secondary reinforcer (Weiss, Buchanan, Alstatt, & Lombardo, 1971). Weiss et al. found that for college students, the opportunity to behave altruistically acted as a reinforcer in a learning task and concluded that altruism itself was rewarding for their subjects. A study by Pisarowicz (Note 1) has provided additional support for the notion that altruism is self-rewarding for adults. Subjects in Pisarowicz's study indicated a greater sense of self-satisfaction and a lessened tendency to engage in direct self-gratification (prize tokens) following altruistic behavior. Consistently, charitable behavior in adults occurs with enhanced frequency under conditions of negative affect (see Cialdini & Kenrick, 1976, for a review), presumably to dispel the bad mood. For the individual who has reached this level of socialization, usually by high school age, helpfulness will have become self-gratifying and will frequently occur privately without the expectation of material or social reward.

Reference Note

1. Pisarowicz, J. A. *Self-reinforcement following altruistic behavior.* Paper presented at the meeting of the Rocky Mountain Psychological Association, Phoenix, Arizona, May 1976.

References

Aderman, D., & Berkowitz, L. Observational set, empathy, and helping. *Journal of Personality and Social Psychology*, 1970, *14*, 141–148.

Apsler, R. Effects of embarrassment on behavior toward others. *Journal of Personality and Social Psychology*, 1975, *32*, 145–153.

Bryan, J. H., & Walbek, N. H. Preaching and practicing generosity: Children's actions and reactions. *Child Development*, 1970, *41*, 329–353.

Cialdini, R. B., Darby, B. L., & Vincent, J. E. Transgression and altruism: A case for hedonism. *Journal of Experimental Social Psychology*, 1973, *9*, 502–516.

Cialdini, R. B., & Kenrick, D. T. Altruism as hedonism: A social development perspective on the relationship of negative mood state and helping. *Journal of Personality and Social Psychology*, 1976, *34*, 907–914.

Eisenberg-Berg, N. The development of prosocial moral judgment. *Developmental Psychology*, in press.

Eisenberg-Berg, N., & Hand, M. The relationship of preschoolers' reasoning about prosocial moral conflicts to prosocial behavior. *Child Development*, in press.

Filter, T. A., & Gross, A. E. Effects of public and private deviancy on compliance with a request. *Journal of Experimental Social Psychology*, 1975, *11*, 553–559.

Fischer, W. F. Sharing in preschool children as a function of amount and type of reinforcement. *Genetic Psychology Monographs*, 1963, *68*, 215–245.

Freedman, J. L., & Doob, A. N. *Deviance: The psychology of being different.* New York: Academic Press, 1968.

Freedman, J. L., Wallington, S. A., & Bless, E. Compliance without pressure: The effects of guilt. *Journal of Personality and Social Psychology*, 1967, *7*, 117–124.

Greenglass, E. R. Effects of prior help and hindrance on willingness to help another: Reciprocity or social responsibility. *Journal of Personality and Social Psychology*, 1969, *11*, 224–231.

Hays, W. L. *Statistics for psychologists.* New York: Holt, Rinehart & Winston, 1963.

Isen, A. M., Horn, N., & Rosenhan, D. L. Effects of success and failure on children's generosity. *Journal of Personality and Social Psychology*, 1973, *27*, 239–247.

Kidd, R. F., & Berkowitz, L. Dissonance, self-concept, and helpfulness. *Journal of Personality and Social Psychology*, 1976, *33*, 613–622.

Konečni, V. J. Some effects of guilt on compliance: A field replication. *Journal of Personality and Social Psychology*, 1972, *23*, 30–32.

McMillen, D. L., Sanders, D. Y., & Solomon, G. S. Self-esteem, attentiveness, and helping behavior. *Personality and Social Psychology Bulletin*, 1977, *3*, 257–261.

Moore, B., Underwood, B., & Rosenhan, D. L. Affect and altruism. *Developmental Psychology,* 1973, *8,* 99–104.

Mosteller, F. M., & Bush, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology: Vol. 1. Theory and method.* Cambridge, Mass.: Addison-Wesley, 1954.

Rawlings, E. I. Witnessing harm to others: A reassessment of the role of guilt in altruistic behavior. *Journal of Personaltiy and Social Psychology,* 1968, *10,* 377–380.

Rosenhan, D. Some origins of concern for others. In P. Mussen, J. Langer, & M. Covington (Eds.), *Trends and issues in developmental psychology.* New York: Holt, Rinehart & Winston, 1970.

Rosenhan, D. L., Underwood, B., & Moore, B. Affect moderates self-gratification and altruism. *Journal of Personality and Social Psychology,* 1974, *30,* 546–552.

Rosenthal, R. Combining results of independent studies. *Psychological Bulletin,* 1978, *85,* 185–193.

Steele, C. M. Name-calling and compliance. *Journal of Personality and Social Psychology,* 1975, *31,* 361–369.

Ugurel-Semin, R. Moral behavior and moral judgment of children. *Journal of Abnormal and Social Psychology,* 1952, *47,* 463–474.

Underwood, B., Berenson, J. F., Berenson, R. J., Cheng, K. K., Wilson, D., Kulick, J., Moore, B. S., & Wensel, G. Attention, negative affect, and altruism: An ecological validation. *Personality and Social Psychology Bulletin,* 1977, *3,* 54–58.

Underwood, B., Froming, W. J., & Moore, B. S. Mood, attention, and altruism: A search for mediating variables. *Developmental Psychology,* 1977, *13,* 541–542.

Underwood, B., Moore, B. S., & Rosenhan, D. L. Affect and self-gratification. *Developmental Psychology,* 1973, *8,* 209–214.

Weiss, R. F., Buchanan, W., Alstatt, L., & Lombardo, J. P. Altruism is rewarding. *Science,* 1971, *171,* 1262–1263.

Weyant, J. M. Effects of mood states, costs, and benefits on helping. *Journal of Personality and Social Psychology,* 1978, *36,* 1169–1176.

Received April 14, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 5, 756-768

The Organizational Functions of the Self: An Alternative to the Duval and Wicklund Model of Self-Awareness

Jay G. Hull and Alan S. Levy
Duke University

A model is proposed that defines the essential functions of self-awareness in terms of the cognitive organization or encoding of information according to its self-relevance. This view is contrasted with a definition of self-awareness in terms of self-focused attention, self-evaluation, and motivated affect reduction. Three experiments are presented in support of the proposed encoding model. The first experiment uses an incidental memory paradigm to demonstrate that self-awareness is associated with a particular encoding depth. The remaining two experiments use previous self-awareness paradigms to demonstrate that self-criticism and self-attribution are not characteristic of self-awareness per se, but rather depend on aspects of the immediate situation. On the basis of this evidence, it is suggested that self-awareness phenomena are a function of a particular form of informational encoding rather than the attentional activation of a self-evaluative process.

The importance of an adequate theory of self for any complete account of human social behavior has long been recognized by philosophers and psychologists alike (Freud, 1927; James, 1890/1950; Kant, 1780/1958; Lewin, 1935; Mead, 1934). These theorists have for the most part distinguished two functions of self: that of organizing the individual's understanding of the social environment and that of regulating behavior through a process of self-evaluation. It is this latter regulatory aspect of the self that has been emphasized in modern social psychology (Bem, 1972; Festinger, 1957; Nisbett & Valins, 1971). According to this view, behavior follows from an evaluation of past activities and as such represents a conscious attempt at self-regulation. While not denying that self-evaluation can serve to mediate behavior in this way, we feel that these theories have tended to restrict unnecessarily the role of the self to a process of self-concept regulation while essentially ignoring its broader function of organizing the individual's understanding of the social environment. This is most easily seen in a recent theory of objective self-awareness (Duval & Wicklund, 1972; Wicklund, 1975).

In terms of the Duval and Wicklund model, self-awareness [1] is composed of four discrete stages: self-focused attention → self-evaluation → affective reaction → motivated discrepancy reduction. Very briefly, attention can be focused in one of two directions: inward on the self or outward on the external environment. Self-focused attention is hypothesized to induce an evaluation of the present self in terms of the individual's ideal self-

The authors especially wish to thank Bruce Saltz for his assistance in Experiment 1 and Steven G. West, Charles Bond, and Timothy D. Wilson for their advice and criticism of earlier drafts of this article.

Requests for reprints should be sent to Jay G. Hull, Department of Psychology, Duke University, Durham, North Carolina 27706.

[1] Duval and Wicklund (1972) originally defined self-awareness in terms of the ratio of an objective (self-focused) to a subjective (environment focused) state. In contrasting their theory with others, then, high self-awareness was postulated to correspond to higher levels of objective as opposed to subjective self-awareness. Following the lead of others, we have substituted the term "self-awareness" for the more cumbersome "objective self-awareness" in connection with their position throughout the present article.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3705-0756$00.75

image. Recognition of a real–ideal discrepancy is hypothesized to elicit an affective reaction —positive or negative depending on the direction of the discrepancy. Upon recognizing that one's ideals exceed one's present self and experiencing a negative reaction, the individual is postulated to be motivated to reduce the degree of negative affect by either (a) working to reduce the size of the real–ideal discrepancy or (b) avoiding the self-aware state. The Duval and Wicklund (1972) theory thus explicitly defines the relation between self and behavior in terms of a regulatory process of self-evaluation and in addition specifies a boundary condition on this process in terms of self-focused attention.

An Encoding Theory of Self-Awareness

In contrast to Duval and Wicklund (1972), we would propose that the defining features of self-awareness do not exist in terms of self-focused attention and the operation of a particular kind of self-evaluative process, but rather exist in terms of the individual's organization of the social environment prior to focal attention and comparison processes. Specifically, we propose that self-awareness corresponds to the encoding of information in terms of its relevance for the self and as such directly entails a greater responsivity to the self-relevant aspects of the environment. Situationally, this level of encoding may be a consequence of the nature of the assigned task, the specific instructions, or the presence of self-symbolic cues (e.g. mirrors, videotape and audiotape recordings). Dispositionally it may represent either a general propensity on the part of the individual or a by-product of more elaborate cognitive structures corresponding to relationships between the self and the environment (Markus, 1977; Rogers, Kuiper, & Kirker, 1977).

The present distinction between encoding and focal attention/comparison processes is common to several recent theories of cognition (Neisser, 1967; Shiffrin & Schneider, 1977; Sternberg, 1975). In terms of a model proposed by Shiffrin and Schneider, feature encoding corresponds to the automatic activation of specific cognitive structures in response to a particular input configuration and as such operates through a set of associative connections in long-term memory. On the other hand, controlled processing corresponds to a capacity-limited comparison or search process dependent on focused attention and under the control of the individual. Since feature encoding defines the form of information available to the individual, it need not be associated with different kinds of controlled processes to lead to different behavioral effects. Thus, encoding may be associated with automatic responses that do not require controlled processing, or, by its inherently constructive nature, encoding may determine the outcomes of those controlled processes that do occur.

From our own perspective, then, self-awareness corresponds to a particular form of encoding process that has its effects on behavior independent of focal attention and comparison processes by rendering the individual sensitive to those aspects of the environment that are potentially self-relevant. It follows from such an analysis that before one can define the behavioral effects of self-awareness, one must specify the ways in which particular aspects of the situation are relevant for the self. We propose that information about one's present situational context is principally self-relevant insofar as it specifies contingencies related to the individual's present activities or projects (Mead, 1934). As such, self-awareness will serve to increase the individual's understanding of the immediate situation in terms of contingencies bearing on those activities. To the extent that such self-relevant contingencies are defined by the responses of others, self-awareness prior to behavior will lead to actions consistent with situational definitions of appropriate conduct (e.g., Carver, 1974, 1975; Diener & Wallbom, 1976; Duval, 1976; Liebling & Shaver, 1973; Scheier, Fenigstein, & Buss, 1974; Wicklund & Duval, 1971). While self-awareness will thus be associated with greater responsivity to situational variations in self-relevant contingencies, it will also be associated with behavioral consistency across situations involving activities that define similar self-relevant contingencies (e.g., Pryor, Gibbons, Wicklund, Fazio, & Hood, 1977).

In addition, we propose that information about past performances and present physiological states also constitutes self-relevant information insofar as it specifies particular kinds of relationships between self and environment. Self-awareness as a greater sensitivity to such information will be associated with a greater responsivity to the evaluative and affective connotations of the information (e.g., Gibbons & Wicklund, 1976; Steenbarger & Aderman, in press; Scheier & Carver, 1977; Gibbons, Note 1). The claim, however, that self-awareness involves a naturally self-evaluative and affect-inducing state (Duval & Wicklund, 1972; Wicklund, 1975) is simply unsubstantiated in the literature to date.

In terms of research issues, our alternative raises questions about three separate aspects of self-awareness theory: (a) To what extent is a state of self-awareness based on a state of self-focused attention? (b) To what extent are the behavioral effects of self-awareness mediated by processes of self-evaluation and affect reduction? (c) To what extent are the attributional effects of self-awareness based on attentional differences? Previous research bearing on each of these issues will be briefly considered along with three studies specifically designed to illustrate our own perspective.

Focal Attention or Encoding Processes?

Two of the main tests of the focus of attention hypothesis have used projective measures (Carver & Scheier, 1978; Davis & Brock, 1975). While these studies did indeed find that self-awareness manipulations led to increases in self-referent responses, we would contend that this was not because of an autistic self-focus, but rather reflects a general propensity to construct the environment in terms of its relationship to the self. In contrast to such projective techniques, Geller and Shaver (1976) have recently used the Stroop color-word test (1938) to demonstrate basic processing differences among differentially self-aware subjects. Thus, self-aware subjects were found to be more sensitive to irrelevant aspects of a color-naming task when those aspects involved self-relevant as opposed to self-irrelevant information. Non-self-aware subjects were not responsive to this difference. It should be noted, however, that self-aware subjects did not show the *general* decrement in processing speed that should be expected with an attentional focus away from the task (Gopher & Kahneman, 1971; Shiffrin & Schneider, 1977). We would propose that the lack of such an overall processing decrement suggests that self-awareness does not correspond to an attentional focus on the self and away from the environment, but rather corresponds to a particular form of informational encoding that is cued by the presence of self-symbolic stimuli.

To provide support for the notion of self-awareness as an encoding process, the present authors designed an experiment to examine differences in self-relevant encoding among individuals varying in degrees of dispositional self-awareness. It was predicted that individuals with high scores on a measure of private self-consciousness (see Fenigstein, Scheier, & Buss, 1975) would be more likely to encode self-relevant information at a deeper level than (a) information encoded on other dimensions and (b) all information (including self-relevant) processed by individuals low in self-consciousness. The primary dependent measure of encoding depth involved incidental recall following a word evaluation task (Craik & Tulving, 1975; Rogers et al., 1977). The encoding dimensions of word length and meaningfulness were selected as control baselines for comparisons with the encoding dimension of self-relevance. The present design thus varied three levels of encoding dimensions along with two levels of self-consciousness.

Experiment 1

Method

Subjects

Subjects were 66 undergraduates of Duke University who participated in partial fulfillment of a course requirement. All subjects had participated in a group testing session 2 months earlier, during which the Fenigstein et al. (1975) self-consciousness inventory was administered.

Procedure

Subjects were run in small groups varying in size from 4 to 11 people. The procedure consisted of the

following instructions read verbatim:

> In this experiment you will be asked to evaluate a series of 30 words which will be presented one at a time. Each word is to be evaluated on one specific dimension. These dimensions are listed in order on the rating sheets you have been provided.
>
> I will call out a number and you are then to read the statement associated with that number. I will then call out a single word. You are then to place a check in the appropriate column to designate whether or not the statement is applicable to that word: in the left column if yes; in the right if no. Some of the statements ask for an evaluation of the length of the word. For these statements check the left if the word is long, the right if short.
>
> Each word is to be evaluated only with respect to the single statement associated with the word called out. Are there any questions about this procedure?

As long as there were no questions, the experimenter continued by calling out a series of 30 words. In each instance he would first read aloud the number associated with the word, wait 5 sec, read aloud the word itself, and then wait 18 sec before calling out the next number.

Evaluation statements. Each word was to be evaluated with respect to one of the three statements used by Rogers et al. (1977, Experiment 2). The specific criteria used in the present study were (a) how long or short the word was; (b) whether the word was meaningful to the subject; and (c) whether the word described the subject. The first criterion is hypothesized to require the shallowest level of processing; the last, the deepest. The order in which the statements appeared was counterbalanced such that in every three trials, each statement appeared once. Subjects were randomly assigned one of three rating sheets, each of which contained a different order of evaluation statements. Thus, within a testing session a given stimulus word was evaluated with respect to all three dimensions, the specific dimension depending on which sheet the subject had been randomly assigned.

Dependent measures. After evaluating all 30 words, subjects were instructed to turn over their rating sheets and write down as many of the words as they could remember. Since stimuli had been presented verbally, they were available only in memory. Subjects were informed that they would have 3 min to complete the task. At the end of 3 min, subjects were asked to place their names at the top of the rating sheets and return them to the experimenter. The purpose of the experiment was then explained, and subjects were asked not to discuss the nature of the experiment with others.

Results and Discussion [2]

Of the randomly assigned evaluation orders, there were 24 subjects in Order 1, 21 in Order 2, and 21 in Order 3. Within each order, subjects were divided into high and low self-conscious groups by means of a median split performed on private self-consciousness scores (Fenigstein et al., 1975). Two subjects were randomly discarded in order to equalize the number of high and low self-conscious subjects in each evaluation order.

The data were submitted to a mixed-design analysis of variance with the number of words remembered within each evaluation category used as a repeated measure. The main effect of self-consciousness was marginally significant, $F(1, 62) = 3.03$, $p < .10$. The within-subjects analysis consisted of three planned comparisons. As predicted, high self-conscious subjects recalled more words encoded as self-relevant than did low self-conscious subjects, $F(1, 124) = 5.61$, $p < .025$. In addition, high self-conscious subjects recalled more words encoded according to their self-relevance than words encoded according to the other control dimensions, $F(1, 124) = 3.88$, $p = .05$. Finally, low self-conscious subjects recalled nonsignificantly fewer words encoded according to their self-relevance than words encoded according to other dimensions, $F(1, 124) = .11$, *ns*.[3] A comparison testing the interaction of self-consciousness and self-relevant versus control words did not achieve a conventional level of significance, $F(1, 124) = 2.58$, $p = .13$.

---

[2] This experiment was originally run as two separate experiments, the second being an exact replication of the first. The results of each were consistent with our hypotheses, and there were no independent effects associated with "experimental session" when this was included as a factor in the design. Therefore, for the sake of brevity, the results are reported collapsed across experiments.

[3] Six of the 10 items from the private self-consciousness scale were significantly correlated with incidental memory of self-referenced words ($.21 < r < .31$, $p < .05$, one-tailed). Two of the items ("I sometimes have the feeling that I'm off somewhere watching myself" and "I'm aware of the way my mind works when I work through a problem") failed to predict in either experimental session: Session 1 ($n = 20$), $rs = -.08$ and .02, respectively; Session 2 ($n = 44$), $rs = .09$ and $-.06$, respectively. While we feel that these items are more closely associated with a focal attention interpretation of self-awareness than are the remaining items, this distinction is obviously a personal judgment on our part.

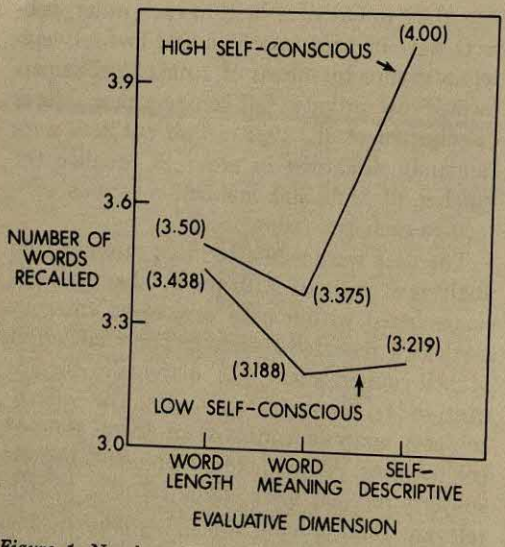


Figure 1. Number of words remembered as a function of self-consciousness and encoding task: Experiment 1.

As can be seen in Figure 1, the pattern of results conformed nicely with predictions. For individuals high in self-consciousness, there is a striking difference between the number of words remembered when encoded as self-relevant as opposed to alternative dimensions. On the other hand, for individuals low in self-consciousness there is no distinction among the different forms of information. This finding occurred in spite of the fact that low self-conscious individuals rated nonsignificantly more of the words as self-descriptive, $t(63) = 1.32$, $p < .20$—a condition that has been found to aid memory on the present task (Rogers et al., 1977).

To determine the effects of the alternative subscales of the self-consciousness inventory, all analyses were repeated using the public self-consciousness and social anxiety indexes as criteria for within-treatment median splits. There were no significant effects for either scale.

The results of the present experiment thus indicate that instructions to self-reference information induce superior incidental recall only for individuals previously identified as dispositionally self-aware. In terms of the analysis provided by Rogers et al., (1977), this would seem to indicate that high self-conscious subjects either (a) possess a more uniform and well-structured cognitive organization corresponding to the self, or (b) this superordinate framework is more easily activated by environmental cues to self-reference. Finally, it should be noted that although self-consciousness was associated with an increased sensitivity to self-relevant information, there was no concomitant decrease in encoding efficiency of non-self-relevant information. We would suggest that these findings, along with those of Geller and Shaver (1976), indicate that self-awareness does not correspond to a bidirectional attentional phenomenon, but rather corresponds to an organizational phenomenon associated with a greater sensitivity to specific forms of environmental information.

Self-Awareness and Self-Evaluation

In addition to denying that self-awareness corresponds to a state of self-focused attention, our analysis is in basic disagreement with the Duval and Wicklund account of the self-evaluative consequences of self-awareness in the absence of performance feedback. The primary evidence for their position exists in the form of an experiment by Ickes, Wicklund, and Ferris (1973) in which self-aware subjects described themselves less favorably relative to their ideals on a "real–ideal self questionnaire." Proponents of the theory claim this study as evidence that under natural conditions, self-aware subjects tend to exaggerate the size of discrepancies between their real and ideal selves. Indirect evidence on subjects whose self-esteem has not been manipulated, however, suggests that self-awareness is associated neither with self-criticism (Carver & Scheier, 1978) nor with the avoidance of generally self-referent responses (Davis & Brock, 1975).

According to our own analysis, a mediational process of self-evaluation and affect reduction is not necessary in order to account for the behavioral consequences of self-awareness. Self-awareness as an encoding process is hypothesized to directly mediate behavioral responses by prefiguring the important and relevant aspects of social situations without engaging a self-evaluative or affect-reduction

process. According to this analysis, subjects' self-critical responses in the experiment by Ickes et al. (1973) were not an inevitable consequence of self-awareness, but rather a consequence of the particular characteristics of the situational context within which the response occurred. It seems likely, for example, that self-deprecation exists as a common self-presentational response to situations similar to that used by Ickes et al. (1973). Lefebvre (Note 2) has shown that female subjects adopt self-deprecation as an ingratiation tactic in the presence of other females and self-enhancement in the presence of males. Self-deprecation in the Ickes et al. study occurred within just such a female (subject) – female (experimenter) interaction. If self-awareness is postulated to heighten the individual's sensitivity to self-relevant aspects of the immediate situation, it might be expected to increase self-deprecation in just this kind of situation.

Although both analyses can explain the results of this particular experiment, they clearly predict different consequences when other variables are introduced. In terms of our analysis, self-deprecation is a response to the immediate circumstances surrounding the self-description. Altering those circumstances by manipulating the characteristics of the audience (e.g., through use of a male rather than a female experimenter) or by rendering the self-descriptions anonymous would be expected to alter the quality of the self-aware response. From the perspective of the Duval and Wicklund (1972) model, such manipulations would not be expected to affect the self-evaluative consequences of self-awareness.

An experiment was designed to test these competing analyses. In accord with our own analysis, it was predicted that self-awareness would lead to self-deprecation when responses were made publicly available to a female experimenter, but the reverse would be true in the presence of a male. In addition, anonymity (by eliminating the self-presentational character of the situation) was predicted to eliminate both of these effects by rendering the sex of the experimenter irrelevant to the subject's behavior. Finally, self-awareness was not predicted to affect subjects' mood. In contrast, Duval and Wicklund (1972) would predict that self-awareness should lead to self-deprecation regardless of other situational variations and that negative affect should be a consequence of this negative self-evaluation.

Experiment 2

Method

Subjects

The subjects were 101 female undergraduates from Duke University who participated in partial fulfillment of a course requirement. Subjects were intentionally restricted to females in order to replicate the conditions of the Ickes et al. (1973) and Lefebvre (Note 2) experiments. Subjects were run individually and randomly assigned to one of eight experimental conditions. Five of these subjects were discarded according to predetermined criteria.[4] The remaining subjects were evenly distributed across conditions.

Procedure

Upon entering the experimental room, subjects were seated before a large (48-inch × 32-inch, 122-cm × 81-cm) mirror that was turned so that its reflective front (high self-aware conditions) or non-reflective back (low self-aware conditions) faced the subject. Subjects' self-descriptions were then obtained under the guise of an experiment described as investigating sex differences in aesthetic reactions. Each subject was shown a stack of art reproductions, along with a series of rating forms for each painting. It was explained that personality and mood differences had to be statistically controlled and that prior to reacting to the pictures it would be necessary for the subjects to fill out two brief self-descriptive questionnaires.

Public and private conditions. Subjects assigned to public conditions were told to write their names and social security numbers on the booklets and to hand them to the experimenter after completing them. In addition, they were told that the booklets would be scored while they rated the pictures, and feedback would be supplied at the end of the experiment.

Subjects randomly assigned to the private conditions were told not to write their names or social

[4] It was determined in advance to discard any subjects in the public conditions who did not put their name on the booklet (two individuals did not) and any subjects in the private conditions who did not place their booklet in the anonymous pile (three individuals did not). The results of all tests of significance are unchanged when these subjects are included in the analysis.

Table 1
Impact of Self-Awareness on Task-Related Self-Descriptions: Experiment 2

Self-awareness	Female experimenter		Male experimenter	
	Public	Private	Public	Private
High				
M	12.67	14.07	14.22	14.11
SD	1.42	1.74	2.01	1.91
Low				
M	15.04	13.11	13.11	14.26
SD	2.03	2.38	1.85	1.18

Note. Higher numbers refer to more positive self-descriptions on a 20-point scale; all statistics are based on 12 subjects per cell.

security numbers on the booklets and after completing them to place them somewhere in the middle of a pile of booklets behind them. No mention was made of the possibility of feedback.

Target other. The characteristics of the individual who would have access to subjects' self-descriptions was varied by using both a female and a male experimenter. For one half of the subjects, the study was conducted by a male experimenter; for the remaining half the experimenter was female.

All subjects were instructed to summon the experimenter after completing the questionnaire. The subject was finally asked if she had any questions about the procedure. As long as there were no questions, the experimenter concluded by commenting that he/she had to share the room with another experimenter and that the bulk of the equipment in the room belonged to the other experimenter's study. The subject was asked not to disturb any of these things, as the experimenter casually indicated the mirror, a tape recorder, and various other experimental stimuli. The subject was then left to work on her own.

Dependent measures. The first questionnaire consisted of a 15-item trait listing. Subjects were instructed to circle numbers from 1 to 10 to indicate the degree of each trait they perceived in themselves. These traits were divided on an a priori basis into three dimensions: task-relevant evaluative (sensitive, creative, unskilled, analytic, and impressionable); general evaluative (careful, intelligent, honest, disagreeable, and crude); and sex-related (masculine, ambitious, forgiving, cooperative, and jealous).[5]

The second questionnaire consisted of a 10-item mood checklist. Once again subjects were instructed to circle numbers from 1 to 10 to indicate the degree of each state they perceived in themselves. These moods were also divided into three dimensions: positive (pleased, positive, pleasant, and lively); negative (depressed, anxious, downhearted, and clutched up); and defiant (defiant, rebellious).

After completion of these questionnaires and the return of the experimenter, subjects were informed that the experiment was over and were probed for suspicion. The actual nature of the experiment was explained to them, and they were thanked for their participation. All subjects were asked not to discuss the experiment with other potential participants.

Results

For the purposes of analysis, the data were combined into the personality and mood categories described above. They were then submitted to a 2 (self-aware vs. non-self-aware) × 2 (public vs. private) × 2 (female vs. male experimenter) analysis of variance.[6]

Personality Self-Descriptions

Task-relevant evaluations. Subjects' responses to these five items were combined by subtracting negative (unskilled) from positive (analytic, sensitive, creative, and impressionable) self-evaluations. The means for each of these eight conditions are reported in Table 1. An analysis of variance on these combined scores revealed a significant triple interaction, $F(1, 88) = 9.27$, $p < .005$. There were no other significant effects. Post hoc analyses indicated that differences in public conditions were primarily responsible for this interaction. Thus, the two-way interaction of self-awareness and sex of experimenter was significant in public (Scheffé procedure, $p <$

[5] The first three task-relevant evaluative and the first four general evaluative traits appeared in the original Ickes et al. (1973) study. An analysis of these seven traits revealed effects identical to the task-relevant and general evaluative dimensions reported below. The two remaining task-relevant traits were regarded as positive skills in the immediate situation of judging art reproductions.

[6] If the distinction is not made between general evaluative, task-relevant, and sex-related traits, and items are combined only on the basis of their positive–negative qualities, the results remain basically the same. Thus, both the triple interaction, $F(1, 88) = 14.92$, $p < .001$, and the sex of experimenter main effect, $F(1, 88) = 3.82$, $p < .06$, are significant. As can be seen in the more extensive analysis, however, the former effect is restricted to general evaluative and task-relevant self-descriptions, whereas the latter only appears on sex-related self-descriptions. Since a combined analysis does not reveal these important differences, separate analyses based on preplanned distinctions were deemed more appropriate.

.001) but not in private conditions ($p > .25$). Self-awareness thus led to more negative self-descriptions in the presence of a female experimenter and more positive self-descriptions in the presence of a male. This was only the case, however, when those self-descriptions were publicly available to those experimenters. Self-descriptions in private conditions were seemingly unaffected by either self-awareness or sex of the experimenter.

General evaluations. Subjects' responses to these five items were combined by subtracting negative (crude, disagreeable) from positive (intelligent, honest, careful) self-evaluations. The means for each of the eight conditions are reported in Table 2. An analysis of variance on these combined scores revealed a significant triple interaction of the same form as that found on the task-relevant items, $F(1, 88) = 9.49$, $p < .005$. Post hoc analyses again indicated that differences in public self-descriptions were primarily responsible for this interaction. Self-awareness interacted strongly with sex of experimenter in public conditions ($p < .001$), but again there were no differences in private ($p > .25$).[7]

Sex-related evaluations. Subjects' responses to these five items were combined by subtracting masculine (masculine, ambitious) from feminine (forgiving, cooperative, jealous) self-evaluations. The means for each of the eight conditions are reported in Table 3. The triple interaction found on the other indexes failed to attain significance on these sex-related

Table 2
Impact of Self-Awareness on General Self-Descriptions: Experiment 2

Self-awareness	Female experimenter		Male experimenter	
	Public	Private	Public	Private
High				
M	13.14	15.48	15.78	14.59
SD	1.81	2.07	1.18	2.51
Low				
M	15.67	15.19	14.63	15.89
SD	1.70	2.34	1.73	2.24

Note. Higher numbers refer to more positive self-descriptions on a 20-point scale; all statistics are based on 12 subjects per cell.

Table 3
Impact of Self-Awareness on Sex-Related Self-Descriptions: Experiment 2

Self-awareness	Female experimenter		Male experimenter	
	Public	Private	Public	Private
High				
M	11.33	11.19	12.59	12.41
SD	1.36	1.30	2.21	1.57
Low				
M	12.37	11.22	12.11	12.89
SD	2.14	2.07	2.03	1.44

Note. Higher numbers refer to more feminine self-descriptions on a 20-point scale; all statistics are based on 12 subjects per cell.

items, $F(1, 88) = 1.78$, $p < 20$. There was, however, a significant main effect for sex of experimenter, $F(1, 88) = 6.98$, $p < .01$. Thus, females described themselves as more feminine in the presence of a male experimenter than in the presence of a female experimenter. This difference did not appear to be contingent on the public quality of the self-description. The most fruitful interpretation of this effect may be in terms of differential trait salience as a function of environmental cues (McGuire, McGuire, Child, & Fujioka, 1978; McGuire & Padawer-Singer, 1976). According to this analysis the present findings of increased feminine trait descriptions in the presence of a male as opposed to a female experimenter would be because of the differential salience of those traits as a function of their situational rarity or distinctiveness.

Mood Self-Descriptions

Subjects' responses to the 10 mood items were combined into three categories for the purpose of analysis: positive, negative, and

[7] Inspection of the means suggests that self-aware subjects in private conditions involving a male experimenter rated themselves slightly less favorably than did non-self-aware subjects. However, insofar as this difference was not significant on this measure —planned comparison $F(1, 88) = 2.55$, $p = .11$—nor even marginal for the task-relevant measure—planned comparison $F(1, 88) = .03$, *ns*—we feel that it is not psychologically meaningful.

defiant. There were no significant effects on any of these indexes.

Discussion

Contrary to the Duval and Wicklund (1972) model, self-awareness did not result in consistently negative self-evaluations and had no apparent effects on reported mood. Instead, the effects of self-awareness appeared contingent on several situational factors. Self-awareness thus interacted with the manipulated anonymity of the response and the sex of the experimenter for both task-relevant and general evaluative self-descriptions.

Consistent with our alternative analysis, the pattern of these results suggests that self-awareness serves to increase sensitivity to self-relevant aspects of the immediate situation. The effects of self-awareness were thus contingent on characteristics of the target for whom the self-descriptions were available.[8] Additional analyses made clear, however, that these effects were restricted to situations in which subjects' self-descriptions were publicly available. Rendering responses anonymous serves the theoretical purpose of eliminating the potential social consequences of the self-descriptions and thus the self-relevance of the sex of the experimenter as a situational cue. Consistent with the present analsis, such a manipulation effectively eliminated the effects of self-awareness.

This is not to say that self-awareness involves "self-presentation" any more than it is to say that self-awareness involves "self-evaluation." The form of the self-aware response is of necessity dependent on the nature of the information that is encoded as self-relevant. Thus, encoding information about past performances in terms of their self-relevance can involve an increased sensitivity to the evaluative connotation of the individual's relationship to the environment. The claim, however, that self-awareness is, in and of itself, a self-evaluative and affect-inducing state is simply unsubstantiated in the literature to date.

Self-Awareness and Causal Attribution

In contrast to an attentional notion of self-awareness, we have proposed that self-awareness involves a particular form of encoding process. Thus, while personally agreeing with the general proposition that attributions follow the individual's focus of attention (Duval & Wicklund, 1972; Jones & Nisbett, 1971; Taylor & Fiske, 1978), we are forced to disagree with the hypothesis that self-awareness affects attribution by increasing *self-focused* attention (Arkin & Duval, 1975; Buss & Scheier, 1976; Duval & Wicklund, 1973). Rather, we would propose that self-awareness affects attribution by increasing the individual's sensitivity to the self-relevant aspects of the immediate situation surrounding the attributional response.

In line with such an analysis, attributional theorists have begun to acknowledge the complex influence of situational factors surrounding attributions of responsibility (Schlenker, 1975; Bradley, 1978; Frey, 1978). Thus, a recent experiment by Frey has demonstrated that when others are aware of a particular outcome, public self-attributions of responsibility are constrained by self-presentational factors of strategic modesty following failure and a complex mix of modesty and self-enhancement following success. On the other hand, private attributions were found to be more responsive to purely defensive concerns.

According to our own analysis, self-awareness should be associated with a heightened sensitivity to just these kinds of self-presentational and defensive concerns. In such circumstances, then, we would predict that the effect of self-awareness on attributions for negative outcomes would depend on the public–private character of the situation.[9]

---

[8] Since only two experimenters were involved, technically it cannot be concluded that the sex of the experimenter was the difference responsible for the experimental effects. However, given the Lefebvre (Note 2) analysis, the brevity of the experimental encounter, and the main effect on sex-related traits, it is reasonable to assume that subjects were responding on the basis of this obvious distinction.

[9] In the present experiment we used hypothetical situations whose outcomes were public knowledge by virtue of having been written by the experimenter and assigned to the subject. The manipulations thus involved public versus private attributions for low-involving situations. Rendering outcomes private or

In contrast, the focus of attention hypothesis would predict a main effect of self-awareness on self-attribution. An experiment was designed to test these opposing predictions by assessing public and private attributions for negative outcomes using procedures based on those developed by Duval and Wicklund (1973) and Buss and Scheier (1976).

Experiment 3

Method

Subjects

The subjects were 52 male undergraduates from Duke University who participated in partial fulfillment of a course requirement. The subject population was restricted to one sex for the sake of convenience. Subjects were run individually and randomly assigned to one of four experimental conditions. Eight of these subjects were eliminated from the analysis on the basis of failure to follow instructions. These subjects were discarded according to predetermined criteria. The remaining subjects were evenly distributed across conditions.[10]

Procedure

Upon entering the experimental room, subjects were seated before a 48-inch × 32-inch (122-cm × 81-cm) mirror that was turned so that its reflective front or nonreflective back faced the subject. All subjects were then introduced to the study in terms similar to those used by Duval and Wicklund (1973) and Buss and Scheier (1976). They were informed that the study involved the development of a social perception questionnaire for use during the following semester and were given a booklet that contained a series of hypothetical situations. They were asked to imagine themselves to be the main character in each of these situations and to respond to each by indicating their degree of responsibility for various hypothetical outcomes. Thus, for example:

> You pull up behind a bus that's stopped at a stop sign, and you want to turn right at this intersection. After waiting 1½ or 2 minutes, the bus hasn't moved. Finally, not knowing what he's going to do, you decide to pull out around him and have to cut back in front to turn right at the corner. Just as you do, he pulls out and runs right into you.

After answering any questions the subject had about this procedure, the experimenter continued with the public-private manipulation.

Public and private conditions. Subjects in the private conditions were asked not to put their names on the booklets and to place the booklets upon completion in the middle of a large pile. Subjects in the public condition were asked to put their names and social security numbers on the booklets, to hand them to the experimenter upon completion, and to anticipate a brief discussion of their responses at the end of the experiment.

Explanation for mirror. For all subjects the experimenter concluded by casually mentioning that he had to share the room with another experimenter and that much of the equipment in the room belonged to the other experimenter's study. He asked the subject not to disturb any of these things, indicating the mirror and various other objects. The subject was then left to work on his own. After completion of the dependent measures, all subjects were carefully probed for suspicion and extensively debriefed.

Dependent measures

The main dependent measures consisted of the percentage of responsibility attributed to the self in each of the five negative-outcome situations used by Duval and Wicklund (1973, Experiment 2). In addition, the amount of time taken to complete the questionnaire was recorded for each subject. This exploratory measure yielded no significant effects and will not be discussed further.

Results

The data were submitted to a mixed-design analysis of variance with two levels each of two between-subjects factors (self-aware vs. non-self-aware; public vs. private) and five levels of the within-subjects factor of hypothetical situation. Contrary to the focus of attention hypothesis, the self-awareness main effect was not significant ($F > 1.00$). However, in accord with our own analysis, the Self-Awareness × Anonymity interaction was significant, $F(1, 40) = 4.38$, $p < .05$.

---

increasing the subject's involvement might well be expected to alter their evaluative connotations and, hence, the character of the self-aware response (cf. Federoff & Harvey, 1976; Frey, 1978).

[10] It was decided in advance to discard any subjects in the public conditions who did not put their name on the booklet (three subjects did not) and any subjects in the private conditions who did not place their booklets in the anonymous pile (two subjects did not).

Table 4
Percentage of Attribution to Self: Experiment 3

Self-awareness	Anonymity	
	Public	Private
High		
M	60.0	51.1
SD	10.9	10.3
Low		
M	53.1	58.2
SD	13.1	9.9

Note. Higher numbers refer to greater attribution to the self on a 100-point scale; all statistics are based on 11 subjects per cell.

Self-aware subjects attributed more responsibility to themselves than non-self-aware subjects in the public condition, but this pattern of results was reversed in private. The only other significant result was a main effect for hypothetical situation, $F(4, 160) = 27.34$, $p < .001$, which simply reflects differences among the hypothetical situations. The means of the conditions collapsed across hypothetical situations appear in Table 4.

Discussion

The present pattern of results suggests that the focus of attention hypothesis by itself is an inadequate conceptualization of the effect of self-awareness on the attribution process: Self-awareness led to greater self-attribution of responsibility than non-self-awareness only under public conditions, the reverse being true in private. Thus, while it may be useful to conceive of causal attributions as affected by the individual's focus of attention (Jones & Nisbett, 1971; Taylor & Fiske, 1978), it is not useful to conceive of the attributional effects of self-awareness in such terms. The present experiment would suggest that self-awareness evokes a much more complex attributional process than merely self-focused attention. Specifically, it is proposed that self-awareness involves an increased sensitivity to the situationally defined meaning of one's actions and, as such, entails a greater responsivity to the presentational constraints and evaluative connotations of self-attributed responsibility. Thus, self-awareness cannot be expected to have a unidirectional impact on attribution (Duval & Wicklund, 1972, p. 103), but rather its effects must be considered in terms of the self-relevant aspects of the attributional context. Anonymity, by rendering different classes of variables self-relevant in the immediate situation, serves to moderate the self-aware response.

Conclusion

In conclusion we would contend that the effects of self-awareness should not be defined by the operation of a particular form of self-regulative process, but rather must be defined in terms of its broader functions of organizing the individual's understanding of the social environment. As such, we would define self-awareness in terms of a heightened sensitivity to particular forms of available information: specifically, the self-relevant contingencies associated with present activity and the self-definitional qualities of information feedback. The three experiments reported in the present article are each attempts to develop the concept of self-awareness along these lines. Thus, Experiment 1 illustrated the particular form of information encoding associated with self-awareness, and Experiments 2 and 3 used conceptual replications of previous research to demonstrate that self-criticism and self-attribution are not characteristic of self-awareness per se, but rather depend on aspects of the immediate situation. Future research, we feel, should be directed toward specifying conditions that determine the self-relevance of various forms of information and thus the effective boundary conditions on the behavioral influences of self-awareness.

Reference Notes

1. Gibbons, F. X. *Self-focused attention and attribution of arousal: Inhibition of the placebo effect.* Unpublished manuscript, University of Texas, 1977.
2. Lefebvre, L. M. *An experimental approach on the use of ingratiation under homogeneous and heterogeneous dyads.* Unpublished manuscript, University of Louvain, Belgium. Reported in Jones, E. E., & Wortman, C. *Ingratiation: An attributional approach.* Morristown, N.J.: General Learning Press, 1973.

References

Arkin, R. M., & Duval, S. Focus of attention and causal attributions of actors and observers. *Journal of Experimental Social Psychology*, 1975, *11*, 427–438.

Bem, D. J. Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6). New York: Academic Press, 1972.

Bradley, G. W. Self-serving biases in the attribution process: A reexamination of the fact or fiction question. *Journal of Personality and Social Psychology*, 1978, *36*, 56–71.

Buss, D. M., & Scheier, M. F. Self-consciousness, self-awareness, and self-attribution. *Journal of Research in Personality*, 1976, *10*, 463–468.

Carver, C. S. Facilitation of physical aggression through objective self-awareness. *Journal of Experimental Social Psychology*, 1974, *10*, 365–370.

Carver, C. S. Physical aggression as a function of objective self-awareness and attitudes toward punishment. *Journal of Experimental Social Psychology*, 1975, *11*, 510–519.

Carver, C. S., & Scheier, M. F. Self-focusing effects of dispositional self-consciousness, mirror presence, and audience presence. *Journal of Personality and Social Psychology*, 1978, *36*, 324–332.

Craik, F. I. M., & Tulving, E. Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 1975, *104*, 268–294.

Davis, D., & Brock, T. C. Use of first person pronouns as a function of increased objective self-awareness and performance feedback. *Journal of Experimental Social Psychology*, 1975, *11*, 381–388.

Diener, E., & Wallbom, M. Effects of self-awareness on antinormative behavior. *Journal of Research in Personality*, 1976, *10*, 107–111.

Duval, S. Conformity on a visual task as a function of personal novelty on attitudinal dimensions and being reminded of the object status of self. *Journal of Experimental Social Psychology*, 1976, *12*, 87–98.

Duval, S., & Wicklund, R. A. *A theory of objective self-awareness*. New York: Academic Press, 1972.

Duval, S., & Wicklund, R. A. Effects of objective self-awareness on attribution of causality. *Journal of Experimental Social Psychology*, 1973, *9*, 17–31.

Federoff, N. A., & Harvey, J. H. Focus of attention, self-esteem, and the attribution of causality. *Journal of Research in Personality*, 1976, *10*, 336–345.

Fenigstein, A., Scheier, M. F., & Buss, A. H. Public and private self-consciousness: Assessment and theory. *Journal of Consulting and Clinical Psychology*, 1975, *43*, 522–527.

Festinger, L. *A theory of cognitive dissonance*. Stanford, Calif.: Stanford University Press, 1957.

Freud, S. *The ego and the id*. London: Hogarth, 1927.

Frey, D. Reactions to success and failure in public and private conditions. *Journal of Experimental Social Psychology*, 1978, *14*, 172–179.

Geller, V., & Shaver, P. Cognitive consequences of self-awareness. *Journal of Experimental Social Psychology*, 1976, *12*, 99–108.

Gibbons, F. X., & Wicklund, R. A. Selective exposure to self. *Journal of Research in Personality*, 1976, *10*, 98–106.

Gopher, D., & Kahneman, D. Individual differences in attention and the prediction of flight criteria. *Perceptual and Motor Skills*, 1971, *33*, 1335–1342.

Ickes, W. J., Wicklund, R. A., & Ferris, C. B. Objective self-awareness and self-esteem. *Journal of Experimental Social Psychology*, 1973, *9*, 202–219.

James, W. *The principles of psychology*. New York: Dover, 1950. (Originally published, 1890).

Jones, E. E., & Nisbett, R. E. *The actor and the observer: Divergent perceptions of the causes of behavior*. Morristown, N.J.: General Learning Press, 1971.

Kant, I. *Critique of pure reason*. New York: Modern Library, 1958. (Originally published, 1780).

Lewin, K. *A dynamic theory of personality*. New York: McGraw-Hill, 1935.

Liebling, B. A., & Shaver, P. Evaluation, self-awareness, and task performance. *Journal of Experimental Social Psychology*, 1973, *9*, 297–306.

Markus, H. Self-schemata and processing information about the self. *Journal of Personality and Social Psychology*, 1977, *35*, 63–78.

McGuire, W. J., McGuire, C. V., Child, P., & Fujioka, T. Salience of ethnicity in the spontaneous self-concept as a function of one's ethnic distinctiveness in the social environment. *Journal of Personality and Social Psychology*, 1978, *36*, 511–520.

McGuire, W. J., & Padawer-Singer, A. Trait salience in the spontaneous self-concept. *Journal of Personality and Social Psychology*, 1976, *33*, 743–754.

Mead, G. H. *Mind, self, and society*. Chicago: University of Chicago Press, 1934.

Neisser, U. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1967.

Nisbett, R. E., & Valins, S. *Perceiving the causes of one's own behavior*. Morristown, N.J.: General Learning Press, 1971.

Pryor, J. B., Gibbons, F. X., Wicklund, R. A., Fazio, R., & Hood, R. Self-focused attention and self-report validity. *Journal of Personality*, 1977, *45*, 513–527.

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology*, 1977, *35*, 677–688.

Scheier, M. F., & Carver, C. S. Self-focused attention and the experience of emotion: Attraction, repulsion, elation, and depression. *Journal of Personality and Social Psychology*, 1977, *35*, 625–636.

Scheier, M. F., Fenigstein, A., & Buss, A. H. Self-awareness and physical aggression. *Journal of Experimental Social Psychology*, 1974, *10*, 264–273.

Schlenker, B. R. Self-presentation: Managing the impression of consistency when reality interferes

with self-enhancement. *Journal of Personality and Social Psychology,* 1975, *32,* 1030–1037.

Shiffrin, R. M., & Schneider, W. Controlled and automatic information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review,* 1977, *84,* 127–190.

Steenbarger, B. N., & Aderman, D. Objective self-awareness as a non-aversive state: Effect of anticipating discrepancy reduction. *Journal of Personality,* in press.

Sternberg, S. Memory scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology,* 1975, *27,* 1–32.

Stroop, J. R. Factors affecting speed in serial verbal reactions. *Psychological Monographs,* 1938, *50*(5, Whole No. 225).

Taylor, S. E., & Fiske, S. T. Salience, attention, and attribution: Top of the head phenomenon. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 11). New York: Academic Press, 1978.

Wicklund, R. A. Objective self-awareness. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 8). New York: Academic Press, 1975.

Wicklund, R. A., & Duval, S. Opinion change and performance facilitation as a result of objective self-awareness. *Journal of Experimental Social Psychology,* 1971, *7,* 319–342.

Received April 18, 1978

Journal of Personality and Social Psychology
1979, Vol. 37, No. 5, 769–777

Disruptive Effects of Disconfirmed Expectancies About Crowding

Kitty Klein and Bruce Harris
North Carolina State University

The experiment utilized a 2 (high vs. low room density) × 2 (forewarning of a crowded room vs. no forewarning) × 2 (simple vs. complex task) design to examine the effects of anticipation of crowding on task performance. More tasks were attempted and efficiency was higher when expectancies about the crowd were confirmed. Subjects not told to anticipate a crowd who actually worked under high density and subjects warned about a crowd that did not materialize performed most poorly. These differences were largest for the complex task. Baum and Greenberg's results were replicated with the performance data. Perceptions of the experimental room also differed as a function of anticipation, but failure to obtain a Crowding × Anticipation interaction did not support their hypothesis that anticipating a crowd induces identical perceptions to those obtained under actual crowding. The results are discussed in terms of disconfirmed expectancies being disruptive of performance, particularly complex task performance.

Although studies of crowding with human populations continue to proliferate, the dynamics of crowding are still largely a mystery. In particular, the search for inhibitory effects of crowding on task performance has not been very successful (Freedman, Klevansky, & Ehrlich, 1971; Sherrod, 1974; Stokols, Rall, Pinner, & Schopler, 1973). Paulus et al. (1976) have suggested three possible reasons to explain the lack of the clear-cut performance deficits predicted by many researchers. First, density levels may have been too low. Second, task variables sensitive to crowding may not have been used; and third, coping mechanisms have not been considered.

Density Levels

Paulus et al. (1976) have suggested that experimental levels of density have not been sufficiently aversive to affect task performance. However, a review of this literature indicates that task and no-task effects occur in relatively similar ratios of persons to space for the high-density subjects. Differences in density that do exist arise from variations in operationalizing low-density (uncrowded) conditions. Investigators reporting no effects of density on task performance, such as Freedman et al. (1971), allowed high-density (crowded) subjects 3.88 square feet (.36 m^2) per person, whereas low-density subjects were afforded 17.70 square feet (1.65 m^2) per person. Similarly, Sherrod's (1974) high-density subjects had 4.62 square feet (.43 m^2) each, and his low-density subjects, 18.75 square feet (1.74 m^2) each. Stokols et al. (1973) utilized 5.64 square feet (.52 m^2) per person for their high-density conditions and 16.73 square feet (1.55 m^2) per person for their low-density condition.

Density levels in experiments in which performance differences occur are very similar to the levels reported above. Heller, Groff, and Solomon (1977), who found a density effect on task performance when subjects interacted closely with each other, used densities of 3–5 square feet (.28–.47 m^2) per person (high density) and 13–15 square feet (1.2–1.4 m^2) per person (low density).

This research was partially supported by a grant to Bruce Harris from the Student Environmental Studies Fund, North Carolina State University. The authors wish to thank Robert L. Wilson for his assistance with the statistical analysis.

Requests for reprints should be sent to Kitty Klein, Department of Psychology, North Carolina State University, Raleigh, North Carolina 27650.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3705-0769$00.75

Paulus et al. (1976), who also report effects of density on task performance, allowed subjects 8.8 square feet (.82 m^2) per person (high-density condition) in the second experiment reported, while their low-density subjects had 26.1 square feet (2.4 m^2) per person. Other researchers reporting some effect of density on task performance include Epstein and Karlin (1975), whose high-density condition permitted 2.66 square feet (.25 m^2) per person and whose low density condition allowed 99 square feet (9.2 m^2) per person, and Aiello, DeRisi, Epstein, and Karlin (1977), who utilized 2.50 square feet (.23 m^2) per person (high density) and 48 square feet (4.4 m^2) per person (low density). Unless subjects are extremely sensitive to these variations in density, density per se is not a very satisfactory explanation of the inconsistent results in the effects of crowding on task performance.

Task Variables

A variety of tasks have been used to assess performance in crowded situations. Density has been reported to have no effects on performance on quiz games (Stokols et al. 1973), on anagrams and number cross-outs (Freedman et al., 1971), or on simple arithmetic and Stroop color-word tests (Sherrod, 1974). Sherrod's crowded subjects, however, did not persist as long as uncrowded subjects at an unsolvable puzzle in a noncrowded room following the experiment.

Other investigators have found some evidence of performance decrements as a function of density. Paulus et al. (1976) found that subjects working in groups of eight solved mazes more poorly than subjects working in groups of four. Epstein and Karlin (1975) investigated delayed effects of crowding on simple and complex task performance to determine if crowding would increase arousal. Increased arousal should inhibit complex task performance while facilitating performance on the simple task. No performance differences were obtained on the complex task (Guilford's creative uses test), but previously crowded subjects did perform better on the simple (counting *a*'s) task. In another study of the delayed effects of crowding, Aiello et al. (1977) employed Torrance's unusual use task and line completion tasks. On some subscores of these tasks, previously crowded subjects were less creative than noncrowded ones.

Heller, Groff, and Solomon (1977) did find performance decrements on a collating task as a function of high density and high physical interaction. This task decrement is not surprising in light of the physical obstruction caused by other subjects in the high-density condition.

Taken together, the results of Paulus et al., Epstein and Karlin, and Aiello et al. do suggest that crowding can facilitate simple task performance and interfere with complex tasks. However, the two latter studies measured performance in noncrowded situations following a period in which subjects were or were not crowded. Therefore, the change in conditions rather than the crowding itself confounds their results.

Coping Mechanisms

In Sundstrom's (1975) model of crowding, coping responses are a possible outcome to stress induced by unwanted social inputs.

One variable that may affect coping responses is the ability to anticipate future conditions. Averill (1973) has applied the term "cognitive control" to information concerning impending stress. Averill's review of laboratory and surgery studies in which information was provided led him to conclude that information is generally preferred but that "there is no consistent relationship between such information and initial reactivity" (p. 295). Applied to the problem of crowding and human behavior, informing subjects of crowded conditions has received little attention.

Langer and Saegert (1977) have proposed a concept of attentional overload, which is caused by a combination of close proximity to others and the necessity to predict future environmental states and coordinate behavior. Langer and Saegert recruited shoppers during crowded and uncrowded hours to shop for a list of 50 items. Half the subjects were

warned that crowded supermarkets induce anxiety. The subjects thus informed showed more positive affect and found more of the items on the list. Shoppers during crowded hours attempted fewer items, correctly completed fewer items, and had more negative attitudes. There were no Density × Information interactions.

Although Baum and Greenberg (1975) and Baum and Koman (1976) led some subjects to anticipate a crowded experimental room, the crowd in fact never materialized. Whether the warning facilitated or impeded coping under conditions of actual crowding was not addressed in their research. Subjects who anticipated a crowded experimental room chose different seat positions and were more negative in their ratings of the experimental room and the confederates than subjects who did not anticipate a crowd.

In the present study, the effects of anticipation and the confirmation or nonconfirmation thereof were tested directly. If anticipating a crowd serves to decrease sensitivity to the stress invoked, subjects who are forewarned and do experience the crowd should report less discomfort and perform better than those who do not have advance warning. If anticipation heightens sensitivity to stress, forewarned individuals whose anticipation is confirmed should feel more stressed when the crowd actually appears.

The design utilized two types of tasks for the purpose of testing differences in the ability to cope with a crowd as a function of forewarning. Applying social facilitation theory (Cottrell, Wack, Sekerak, & Rittle, 1968; Zajonc, 1965), if anticipation of a crowd increases arousal, performance on simple tasks should improve, compared to a no-anticipation condition. Likewise, arousal-increasing anticipation should have deleterious effects on complex task performance. If, however, anticipation allows preparation of coping responses, performance on complex tasks should be better as a consequence. On simple tasks, additional coping would have little effect, although there may be some facilitation. Finally, if Baum and Greenberg (1975) are correct regarding the dynamics of crowding, subjects anticipating a crowd should perceive the experimental room differently than those subjects not anticipating a crowd, and these perceptions should be similar to those of crowded subjects not forewarned.

Method

Subjects

A total of 40 male and 40 female undergraduate students enrolled in an introductory psychology course at North Carolina State University participated in the experiment for partial course credit.

Design

The study was a 2 (anticipate crowding vs. no anticipation) × 2 (performing under crowded conditions vs. not-crowded conditions) × 2 (simple vs. complex tasks) experimental design, to test (a) effects of anticipation on performance of tasks and (b) differential effects on complex and easy tasks. Half of the subjects were led to anticipate a crowded environment. The subjects then performed either simple (vowel cross-out) or complex tasks (logical reasoning puzzles) in a small experimental room (1.75 m^2 and 2.70 m high) under crowded conditions (five subjects in the room) or uncrowded conditions (two subjects in the room). Thus, crowded subjects were afforded .61 m^2 (6.61 square feet) per person, whereas uncrowded subjects had 1.5 m^2 (16.50 square feet) per person. Sexual composition of the groups was not controlled. Upon completion of the tasks, subjects were given a questionnaire to assess their perceptions of the experimental room.

The major dependent variable was the subjects' performance (accuracy and speed) on the simple and complex tasks. Twelve ratings of the experimental room itself were also obtained.

Procedure

The anticipation-no-anticipation manipulation began with the recruiting of subjects for an experiment entitled "Logical Reasoning." For the 20 subjects whose expectancies about a crowd were to be disconfirmed, the sign-up sheets contained five lines with three names (fictitious) already written in. Although there was no indication that the experiment involved crowding or even interaction per se, these anticipation subjects were aware that 3 individuals were to report at the same hour. For the remaining 40 subjects, all of whom would be actually crowded, all five lines on the sign-up sheet were filled by subjects themselves.

The 20 subjects who were not led to anticipate a crowd and who were not crowded signed up on sheets containing only two spaces for each session of the experiment. All sign-up sheets requested that friends not serve together.

Subjects arrived at the experiment in groups of either two or five. They were greeted by the experimenter and told to leave all belongings, except for a pencil, on a table in a large room that adjoined a small experimental room, which was 1.75 m^2 and 2.70 m high.

No-Anticipation Condition

The door to the experimental room was partially closed for those subjects not led to anticipate a crowd. When these subjects (no-anticipation group) were ready with their pencils, the door to the experimental room was opened and subjects were led into the room. In the no-crowd condition, two subjects entered the room, which contained two chairs. In the crowded condition, five subjects entered the room, which contained five chairs. The tasks were placed face down on the chairs before the subjects arrived. Subjects then read the instructions of their tasks, either crossing out and counting vowels from a list of random letters or performing five logical word puzzles taken from Wylie (1957). The instructions urged subjects to work as quickly and accurately as possible and to time their own solutions. The experimenter then closed the door for a specified period of time (13 minutes for the vowel cross-out and 23 minutes for the word puzzles). A pretest with the same tasks had indicated that the majority of subjects completed the puzzle task in 23 minutes, whereas the vowel-counting task required only 13 minutes to complete. All subjects in the room worked on the same type of task. Subjects were asked to record their answers on an answer sheet along with the time of day that each puzzle or vowel was completed. That is, subjects wrote down the time, according to a clock hanging in the experimental room, after completing each of the five puzzles, or after counting each of the five vowels. In addition to the starting time, then, five times were recorded on the answer sheet if the subject completed all five tasks.

After the time period elapsed, the experimenter opened the door, collected the papers, and handed the subjects a questionnaire to assess their feelings about the experimental room. This questionnaire consisted of 12 7-point scales; subjects rated the room in terms of its temperature, how pretty it was, illumination levels, stuffiness, size, adequacy, cheerfulness, suitability, dampness, pleasantness, how crowded it appeared, and how much physical discomfort it caused them. When this questionnaire was completed, subjects left the experimental room and were debriefed and thanked.

Anticipation-of-Crowding Condition

For those subjects who were led to believe that five subjects would participate, the procedure was the same with the following exceptions: The door to the experimental room was open when subjects arrived at the large anteroom, and in all cases, five chairs with five papers (the tasks) on them were clearly visible. Before telling the subjects to lay their belongings on a table, the experimenter pointed to the experimental room and stated that the small room was the location of the experiment. When subjects were ready with their pencils, the experimenter showed them a list of five names, and asking each subject his/her name, conspicuously checked it off. In the condition in which crowding occurred, five subjects entered the experimental room, and the door was closed. In the condition in which the anticipation of crowding was not confirmed, the two subjects entered the experimental room after having their names checked off the list of five names and were told by the experimenter that he would let the others in as soon as they arrived. The experimenter then closed the door, leaving the two subjects to work in a room with three empty chairs. As in the no-anticipation condition, subjects were debriefed and thanked after completing the final questionnaire.

Dependent Measures

The tasks were scored in the following manner: For the vowel-counting task, there were five correct answers (one for each vowel), and a subject was either correct or incorrect in recording the frequency of each vowel. Of the five logical word puzzles, two required only one answer to solve the puzzle correctly. Of the remaining three puzzles, one required two answers for successful completion, one required three answers, and the third required four answers. To be scored correct, all answers of the given puzzle had to be shown correct. In the analyses that follow, standard scores were computed using the overall means of all subjects assigned to the puzzle conditions and of all subjects in the vowel conditions.

Three different measures of performance were analyzed: simple number of correct puzzle solutions and correct vowel counts; number of puzzles and vowel problems attempted; and the number of minutes spent for each solution on both tasks. Because of the nature of the instructions, the minutes/task measure was treated as an index of efficiency similar to Paulus, Gatchel, and Seta's (1978) efficiency (rate). In order to test both the puzzle and the vowel data in an analysis that included type of task as an independent variable, each subject's scores were transformed to standard scores, using the overall means of all subjects assigned to the appropriate task condition.

Results

Preliminary Analyses

All three dependent measures were first analyzed to determine if there were any effects of sex of subject, sexual composition of the group, or the groups themselves. Subject

sex, crowding, and whether the crowding was anticipated were used as factors in a multivariate analysis of variance (MANOVA) on the three measures for each task. Neither main effects of sex nor interactions were found for either the puzzles or the vowels.

The groups, which consisted of either two persons (uncrowded) or five persons (crowded) were further categorized as all male, all female, or mixed in order to determine whether sexual composition of the group affected performance. Again, a multivariate analysis of variance indicated neither a main effect nor interactions for sexual composition for either task.

Finally, a MANOVA using crowding, anticipation, and groups as a nested factor was conducted for each task. The groups effect was not significant in either the puzzle or the vowel conditions. Because of the lack of effects for sex, sexual composition of the group, or the groups themselves, the data below are collapsed across these variables.

Task Performance

The effects of forewarning of a crowd on simple and complex task performance was tested in a three way MANOVA (Anticipation × Crowding × Task) using the number of correct solutions, the number attempted, and the time per correct solution (efficiency). There were significant interactions: Anticipation × Crowding, $F(3, 70) = 3.81$, $p < .01$; Crowding × Task, $F(3, 70) = 2.77$, $p < .05$; and Anticipation × Crowding × Task, $F(3, 70) = 2.58$, $p < .06$; and a significant main effect of tasks, $F(3, 70) = 74.74$, $p < .0001$.

The univariate tests on each dependent variable indicated that on the number correctly completed, only a task effect was obtained. Subjects who counted vowels actually got fewer correct than did subjects who solved puzzles—1.40 correct for the vowel count versus 1.97 for the puzzles, $F(1, 72) = 6.89$, $p < .01$. This difference may arise from the different nature of the feedback involved in each task. The vowel counting offered no check on the correctness of the solution; that is, errors in counting were unlikely to be detected. An incorrect guess concerning the logical puzzle solution could be more easily rejected, because subsequent guesses would prove incorrect.

Table 1
Mean Number of Tasks Attempted as a Function of Forewarning and Crowding

Condition	Crowded	Not crowded
Anticipation	$4.6_{a,b}$	4.15_{b}
No anticipation	$4.4_{a,b}$	4.75_{a}

Note. For each mean, $n = 20$. Means with different subscripts differ significantly using Duncan's multiple-range tests. Number of tasks attempted (both puzzles and vowels) ranged from 3 to 5; maximum possible = 5.

On the number of tasks attempted, the Anticipation × Crowding interaction was significant, $F(1, 72) = 6.03$, $p < .016$. Across both tasks, subjects whose expectancies were confirmed attempted more problems than did subjects whose expectancies were disconfirmed. Subjects who were forewarned about a crowd and were actually crowded tried more tasks than subjects forewarned and not actually crowded. Subjects not forewarned and not crowded tried more tasks than subjects surprised by the crowd (see Table 1).

On the third dependent variable, the amount of time spent per correct solution, the three-way interaction (Anticipation × Crowding × Task) was significant, $F(1, 72) = 5.11$, $p < .02$. As can be seen in Figure 1, it is clearly in the complex task that crowding and anticipation had their greatest effect.

To examine these results in each task separately, a 2 (anticipation – no anticipation) × 2 (crowded – not crowded) analysis of variance was performed on all three dependent variables. In this analysis, no effects of anticipation or crowding appeared for either the number of vowels correctly counted or the number attempted. There was an effect of anticipation on efficiency (number of minutes spent working on each letter) on the vowel-counting task, however, $F(1, 36) = 4.38$, $p < .04$. Subjects who anticipated a crowd spent more time on each vowel count than subjects not anticipating a crowd (3.14 minutes vs. 2.68 minutes per vowel count).

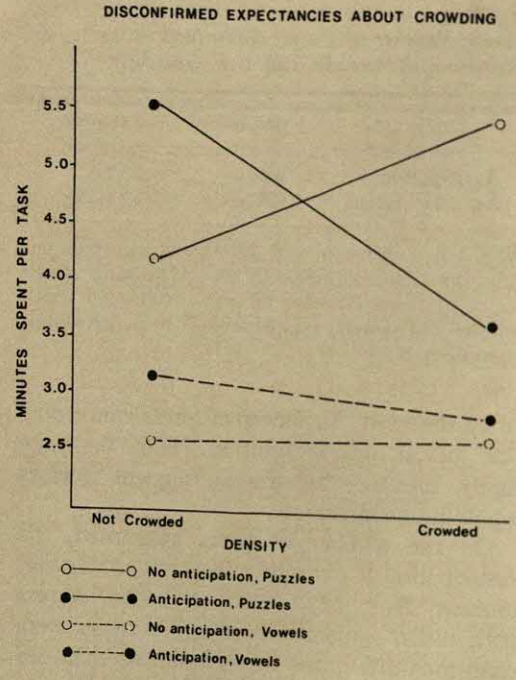


Figure 1. Efficiency as a function of density, anticipation, and type of task.

On the puzzle task, the Crowding × Anticipation interaction was significant on both the number of puzzles attempted, $F(1, 36) = 7.04$, $p < .012$, and on the mean time per solution, $F(1, 36) = 9.10$, $p < .004$. As can be seen in Table 2, fewer attempts to solve were made when crowding was expected but failed to occur and when crowding was not expected but did occur. In both these cells, subjects were caught "unaware." Those who had been forewarned that they would be performing in a high-density environment but actually worked in low-density conditions had their expectancies about the experimental session disconfirmed. Likewise, subjects who were not forewarned suddenly found themselves working in a crowd.

The same pattern of effects appears in the time required for each solution (see Table 2). In this interaction, subjects who expected a crowd and were indeed crowded or who did not expect a crowd and were not crowded performed more efficiently than subjects who were surprised by the experimental conditions.

Perceptions of the Experimental Room

To determine whether the anticipation factor also affects perceptions of the environment, a multivariate analysis of variance was performed on the scales subjects used to rate the experimental room. As in Baum & Greenberg's (1975) study, there was a significant main effect of anticipation, $F(12, 65) = 2.36$, $p < .012$.

Univariate tests of anticipation indicated that anticipation had significant effects on how hot, $F(1, 72) = 7.32$, $p < .01$; how comfortable, $F(1, 72) = 4.65$, $p < .05$; how adequate, $F(1, 72) = 7.99$, $p < .01$; and how suitable, $F(1, 72) = 10.26$, $p < .01$, the room appeared (see Table 3). Subjects who anticipated a crowd rated the room as hotter, less comfortable, less adequate, and less suitable than subjects who were not forewarned. Whether or not the crowd actually materialized also approached significance on the room perception data, $F(12, 65) = 1.82$, $p < .06$. The only rating on which crowding had a significant main effect was the perception of how crowded the room appeared, $F(1, 72) =$

Table 2
Mean Number of Puzzles Attempted and Time (in minutes) per Solution as a Function of Forewarning and Crowding

Condition	Crowded		Not crowded	
	No. attempted	Time	No. attempted	Time
Anticipation	4.80_a	3.77_x	4.20_b	5.63_y
No anticipation	4.20_b	5.21_y	4.80_a	4.25_x

Note. For each mean, $n = 10$. Means with different subscripts differ significantly using Scheffé tests. Time ranged from 1.80 minutes to 7.66 minutes; maximum possible = 23 minutes. Number of puzzles attempted ranged from 3 to 5; maximum possible = 5.

14.50, $p < .001$. Crowded subjects did perceive the room as more crowded than noncrowded subjects ($M = 4.25$ vs. $M = 2.62$, respectively).

The multivariate F for the Crowding × Anticipation interaction was not significant, $F(12, 65) = 1.03$, $p < .4$. Type of task also had no effect on room perceptions, even though puzzle subjects spent a longer time in the experimental room than did subjects counting vowels.

Discussion

Performance

The present data raise a number of issues heretofore unaddressed in either the social facilitation or crowding literature. Subjects in uncrowded conditions who were not misled to anticipate a crowd performed better on both simple and complex tasks compared to subjects warned of a crowd that failed to materialize. Likewise, being crowded when one was not forewarned hurt performance on simple and complex tasks compared to working in crowded conditions with forewarning, although these differences were not significant.

On the simple vowel-counting task, anticipation decreased efficiency for both crowded and uncrowded conditions. This finding fits neither a learned drive (Cottrell, 1972) nor a distraction explanation (Baron, Moore, & Sanders, 1978) of social facilitation. If forewarning about crowding acts to increase arousal, forewarned subjects' performance on the simple vowel-counting task should have been superior to the performance of subjects given no information.

Nor are these results analogous to those reported by investigators concerned with the effects of information about impending stress. In such studies, the amount or kind of information has varied, rather than its accuracy. As Averill (1973) notes, predictability has been found both to enhance and to inhibit stress reactions. However, in the data presented here, subjects with equivalent information about the experimental setting (that is the anticipation–crowded and the no-anticipation–not-crowded conditions) were not equally efficient vowel counters.

Table 3
Mean Ratings of the Experimental Room as a Function of Anticipating a Crowd

Condition	Descriptive dimension			
	Hot	Comfortable	Adequate	Suitable
Anticipation	3.70	4.47	4.50	4.25
No anticipation	4.20	3.67	3.42	3.02
F	7.32**	4.65*	7.99**	10.26**

Note. $n = 40$. Ratings are on a 7-point scale. The lower the mean, the closer to the descriptive dimension.
* $p < .05$. ** $p < .01$.

Further indication that anticipation does not affect the actual crowding experience in any straightforward manner is found in the data from the complex task conditions. Performance on the puzzles was affected by whether the anticipation was confirmed, not by anticipation itself. Subjects who anticipated and did experience the crowd performed better than those who had no advance warning of the crowd. For crowded subjects then, anticipation enhanced performance. Subjects not forewarned and not crowded performed equally well, but those who mistakenly anticipated a crowd were less efficient.

These complex task results fit a social facilitation paradigm only if one assumes that it is the disconfirmation of an expectancy that contributes to arousal or distraction. The disconfirmation explanation also helps explain the data in terms of the information about aversive stimuli research. Lack of predictability decreases complex task efficiency.

As in previous studies (Freedman et al., 1971; Sherrod, 1974; Stokols et al., 1973), no differences in performance were found as a result of crowding alone. Experiments (Aiello et al., 1977; Epstein & Karlin, 1975) in which delayed effects of crowding are reported may inadvertently have confounded expectations about crowding in the test situation with the pretest density manipulation. Subjects exposed to a crowd may anticipate further crowding; their surprise at their subsequent isolation during task performance

may affect their performance as much as the crowd they previously experienced.

The performance data only partially replicate the findings of Langer and Saegert (1977). In that study, forewarning proved beneficial whether or not crowding actually occurred. This discrepancy may arise from the different nature of the tasks used or the type of forewarning manipulation. Langer and Saegert explicitly warned subjects that crowding induced anxiety; in the present study, the forewarning was information only, and subjects were not led to expect any particular psychological consequences.

Room Perceptions

The room perception data do show main effects of anticipation. On several scales, subjects who anticipated the crowd rated the room as more uncomfortable than subjects not forewarned. The room ratings were made at the end of the experimental session. The absence of any Crowding × Anticipation interaction on room perceptions suggests that the density levels used here did not create additional discomfort beyond that generated by the forewarning at the beginning of the experiment.

Baum and Greenberg's (1975) contention that crowding and anticipation are dynamically similar states receives equivocal support in the present results. A strict test of Baum and Greenberg's hypothesis would require similar perceptions from subjects anticipating a crowd that did not materialize and subjects who did not anticipate a crowd but did find one. Complete support for this interpretation of Baum and Greenberg's hypothesis would necessitate an interaction between crowding and anticipation such that anticipating–not-crowded subjects perceived the room in the same way as nonanticipating–crowded subjects. No univariate tests of the Crowding × Anticipation interactions on the room ratings were significant. Thus, although anticipation did cause differences in perceptions of the room, anticipating a crowd did not lead to the same perceptions as actually being in one.

There is more support for Baum and Greenberg's proposal in the performance data. Subjects who anticipated a crowd that never materialized did not perform significantly differently than subjects not anticipating a crowd who worked under crowded conditions.

In the present study, reliable interactions were obtained despite the indirect nature of the forewarning and a crowding experience that was perceived as only moderately uncomfortable. Whether anticipation would have had similar effects if the crowding had truly been stressful has yet to be determined. The degree to which the forewarning is simply an objective description of future conditions, advises on coping mechanisms, or forecasts physical or emotional distress as a consequence of severe crowding also should be examined. Human beings encounter naturally crowded situations both with (e.g., holiday shopping) and without (e.g., delays in air terminals) forewarning. It would seem appropriate to investigate under what conditions different types of foreknowledge increase the ability to cope with a crowd.

References

Aiello, J. R., DeRisi, D. T., Epstein, Y. M., & Karlin, R. A. Crowding and the role of interpersonal distance preference. *Sociometry*, 1977, *40*, 271–282.

Averill, J. R. Personal control over aversive stimuli and its relationship to stress. *Psychological Bulletin*, 1973, *80*, 286–303.

Baron, R. S., Moore, D., & Sanders, G. S. Distraction as a source of drive in social facilitation research. *Journal of Personality and Social Psychology*, 1978, *36*, 816–824.

Baum, A., & Greenberg, C. I. Waiting for a crowd: The behavioral and perceptual effects of anticipated crowding. *Journal of Personality and Social Psychology*, 1975, *32*, 671–679.

Baum, A., & Koman, S. Differential response to anticipated crowding: Psychological effects of social and spatial density. *Journal of Personality and Social Psychology*, 1976, *34*, 526–536.

Cottrell, N. B. Social facilitation. In C. G. McClintock (Ed.), *Experimental social psychology*. New York: Holt, Rinehart & Winston, 1972.

Cottrell, N. B., Wack, D. L., Sekerak, G. J., & Rittle, R. H. Social facilitation of dominant responses by the presence of an audience and the mere presence of others. *Journal of Personality and Social Psychology*, 1968, *9*, 245–250.

Epstein, Y. M., & Karlin, R. A. Effects of acute experimental crowding. *Journal of Applied Social Psychology*, 1975, *5*, 34–53.

Freedman, J. L., Klevansky, S., & Ehrlich, P. I. The effect of crowding on human task performance. *Journal of Applied Social Psychology,* 1971, *1,* 7–26.

Heller, J. F., Groff, B. D., & Solomon, S. H. Toward an understanding of crowding: The role of physical interaction. *Journal of Personality and Social Psychology,* 1977, *35,* 183–190.

Langer, E. J., & Saegert, S. Crowding and cognitive control. *Journal of Personality and Social Psychology,* 1977, *35,* 175–182.

Paulus, P. B., Annis, A. B., Seta, J. J., Schkade, J. K., & Matthews, R. W. Density does affect task performance. *Journal of Personality and Social Psychology,* 1976, *34,* 248–253.

Paulus, P. B., Gatchel, R. J., & Seta, J. J. Enhancement and reduction of task performance by psychological modeling. *Personality and Social Psychology Bulletin,* 1978, *4,* 126–130.

Sherrod, D. R. Crowding, perceived control, and behavioral aftereffects. *Journal of Applied Social Psychology,* 1974, *4,* 171–186.

Stokols, D., Rall, M., Pinner, B., & Schopler, J. Physical, social and personal determinants of the perception of crowding. *Environment and Behavior,* 1973, *5,* 87–117.

Sundstrom, E. An experimental study of crowding: Effects of room size, intrusion, and goal blocking on nonverbal behavior, self-disclosure, and self-reported stress. *Journal of Personality and Social Psychology,* 1975, *32,* 645–654.

Wylie, C. R. *101 puzzles in thought and logic.* New York: Dover, 1957.

Zajonc, R. B. Social facilitation. *Science,* 1965, *149,* 269–274.

Received May 1, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 5, 778-789

Persistence of Opinion Change Induced Under Conditions of Forewarning and Distraction

William A. Watts
Department of Education
University of California, Berkeley

Lewis E. Holt
Stanislaus State College

Ninety-six high school students participated in a study investigating the immediate and delayed effects of forewarning of persuasive intent. It was predicted that subjects would change less immediately after reading persuasive communications because the forewarning would serve as a discounting cue, but that over time, they would tend to forget or dissociate this cue, thus allowing the full impact of the communication to emerge. The results strongly supported this hypothesis.

A second experiment involving 104 high school students was conducted to replicate the first study and to extend the same reasoning to the case of distraction. Distraction was expected to facilitate immediate opinion change, presumably because of interference with counterarguments; but because of its detrimental effect on comprehension and a presumed tendency for subjects to think of opposing arguments after leaving the experimental situation, the change was expected to dissipate more rapidly than in the nondistracted conditions. The data confirmed predictions regarding both forewarning and distraction.

Although forewarning of persuasive intent has been studied extensively (e.g., McGuire, 1966, 1969; McGuire & Papageorgis, 1962; Papageorgis, 1968; Petty & Cacioppo, 1977), there have been no reported investigations of its long-term, or delayed, effects. If forewarning has an immediate detrimental effect on opinion change, which is by no means universally found (e.g., Papageorgis, 1968), there is reason to believe that the inhibiting influence will dissipate over time, thus allowing a greater delayed than immediate impact of the message. That is, if forewarning leads subjects to think of counterarguments either before or while reading a communication (e.g., McGuire & Papageorgis, 1962; Petty & Cacioppo, 1977), arouses psychological reactance (Brehm, 1966; Hass & Grady, 1975), or causes the subjects to perceive the communicator as less fair (e.g., Hass & Grady, 1975), an immediate reduction in persuasive impact should result. However, unless the forewarning significantly interferes with learning the message content, which seems unlikely (e.g., Freedman & Sears, 1965), subjects may tend to forget or spontaneously dissociate these initial reactions over time, thus allowing the full persuasive impact of the material to emerge. For example, Hass and Grady (1975) point out that despite the presence of reactance, one may be persuaded by the informational value of the arguments presented by the communicator. It would logically follow that given compelling arguments, once the initial reactance subsides, the impact of the information should produce increased opinion change over time (e.g., Gruder et al., 1978). Similarly, if forewarning tends to create the perception of a biased communicator (e.g., Hass & Grady, 1975), as this discounting cue becomes dissociated from the message over time, a delayed increase in agreement might be expected. This is analogous to the

Sincere appreciation is extended to Robert S. Wyer, Jr., and Richard E. Petty for their most helpful and astute comments when reading earlier versions of this manuscript.

Requests for reprints should be sent to William A. Watts, Department of Education, University of California, Berkeley, California 94720.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3705-0778$00.75

original interpretation of the sleeper effect in persuasion (Hovland, Lumsdaine, & Sheffield, 1949).

Issue Involvement as a Factor in Forewarning

In attempting to account for the conflicting results in studies of forewarning, Papageorgis (1968) has differentiated between cases involving the mere prior announcement of the topics and directions of persuasive communications, termed "warnings," and those in which the subjects are specifically told that the experiment deals with persuasion, dubbed "persuasion contexts." He hypothesized that persuasion contexts served to reduce the impact of communications about issues with high involvement or those that concern controversial topics or make emotional appeals, whereas for communications about issues with low involvement or those that argue about cultural truisms or make factual appeals, persuasion contexts have no more effect than disguised (no forewarning) contexts when other characteristics of the persuasion situation are neutral.

There are at least two theoretical reasons for expecting different results for issues of high and low involvement: (a) Psychological reactance (Brehm, 1966) should be greater for important issues, and (b) subjects are probably less able and/or motivated (e.g., Vinokur & Burnstein, 1978) to think of counterarguments in the case of cultural truisms or esoteric topics. Dean, Austin, and Watts (1971) found no support for these conjectures; however, their studies were limited in the sense that each employed only two issues—one high and the other low in involvement. Naturally, the topics varied in a number of other respects that may have influenced the obtained results.

In the present study, eight issues were used in an attempt to minimize the aforementioned problem: Four dealt with familiar, controversial issues; and the others, with esoteric topics. An example of the former would be the proliferation of nuclear weapons, and of the latter, the increasing length of the geological day. If the reasoning of Papageorgis (1968) and perhaps that of Apsler and Sears (1968) is correct, the immediate effects of forewarning should be to inhibit change on the familiar topics and to facilitate it or have no effect on subjects' opinions concerning the esoteric issues.

The first study was designed to test these conjectures. Since each subject read messages on four different topics, it was not feasible to warn them of the positions advocated in each case. Rather, subjects were warned of the persuasive intent (Papageorgis, 1968) of the materials they were about to read. This manipulation has been used in several previous studies (e.g., Hass & Grady, 1975; Kiesler & Kiesler, 1964). Under these circumstances there is no opportunity for anticipatory counterarguing (Petty & Cacioppo, 1977); consequently, there is no reason for a delay between forewarning and receipt of the messages (e.g., Hass & Grady, 1975). Similarly, there is little opportunity for anticipatory belief change (e.g., McGuire & Millman, 1965), since the subjects are unaware of the topics and positions advocated until the message are received.

Distraction as a Variable Influencing Persistence of Opinion Change

A second study focused on the single and joint effects of forewarning and distraction. Whereas forewarning has been shown previously to facilitate production of counterarguments (e.g., Petty & Cacioppo, 1977), distraction apparently interferes with this process (e.g., Baron, Baron, & Miller, 1973; Festinger & Maccoby, 1964; Osterhouse & Brock, 1970; Petty, Wells, & Brock, 1976). Therefore, distraction should influence positively the yielding component of persuasion (e.g., McGuire, 1968, 1969; Wyer, 1974, chap. 7), thus rendering the person more susceptible to influence if the message is adequately comprehended.

However, there is considerable evidence that distraction also interferes with comprehension (e.g., Haaland & Venkatesan, 1968; Petty et al., 1976; Regan & Cheng, 1973; Zimbardo, Snyder, Thomas, Gold, & Gurwitz, 1970). Whenever a variable influences the reception and yielding components of persuasion in opposite directions, the resultant effect often may be nonmonotonic (e.g., McGuire,

1968, 1969; Wyer, 1974, chap. 7). That is, a moderate amount of distraction will facilitate opinion change because the reduction in counterarguing more than offsets the decrement in comprehension. As distraction increases, however, the loss in comprehension should become greater, thus leading to a decrease in opinion change, since, in the extreme case, one cannot conform to the position advocated in the communication without understanding the side taken. Even though the immediate effect of distraction may be to facilitate opinion change, there are at least two reasons to expect the gain to be short-lived. First, several prior studies indicate that experimental treatments affecting initial learning influence persistence of induced opinion change as well (see Cook & Flay, 1978). Thus, the decay in opinion change may be quite rapid, depending on the extent that distraction interferes with comprehension. Note that in the case of distraction, consideration of its effects on comprehension lead to the exact opposite prediction of that derived for forewarning, namely, that the induced opinion change would dissipate more rapidly for distracted subjects because of poorer learning of the message content.

Another reason for more rapid decay of opinion change under conditions of distraction hinges on postexperimental counterarguing. Once the subject has left the experimental room, he or she is free of distraction and in a position to reflect on the persuasive messages. It seems likely that in the process, the individual would be able to think of a greater number of opposing arguments that would serve to dampen his/her newly formed opinion. Naturally, the people who were not distracted may also cogitate upon the issues afterward and think of additional counterarguments. However, the latter should be smaller in number and have less impact because these individuals would already have considered many of the counterarguments during the experimental session and have taken them into account at that time (e.g., Vinokur & Burnstein, 1978; Vinokur, Trope, & Burnstein, 1975). Actually, the distracted subjects may be more motivated to think about the topics after the experimental session, since reading communications under conditions of distraction may require greater effort (e.g., Baron et al., 1973), and if subjects have worked harder on a task, it would seem likely that they would feel increased involvement.

To test these conjectures, in the second study, forewarning versus no forewarning plus distraction versus no distraction were varied orthogonally so that their single and joint effects on immediate and delayed opinion change could be examined.

Experiment 1

Method

Subjects and Design

The subjects consisted of 96 students from four senior-level high school classes who participated in both sessions of the experiment, separated by a 1-week interval, during their normal class periods. There were approximately equal numbers of males and females.

Eight messages were randomly divided into two subsets of four each with the restriction that they had to contain two familiar and two esoteric topics. These forms of the materials were alternated. Subjects read one of the subsets of persuasive messages and then stated their opinions on all eight issues. Hence, each person served as an experimental subject on four issues and as a control on the remaining four topics. A $2 \times 2 \times 2 \times 2$ factorial design was used involving one between-subjects variable (forewarning vs. no forewarning) and three within-subjects variables (message type: esoteric vs. familiar, time of measurement, and experimental vs. control). The order of presentation of the two types of messages was counterbalanced. Since subjects were randomly assigned to experimental treatments, a posttest-only design was employed, inasmuch as a pretest itself would serve to some extent as a forewarning of persuasion.

Procedure

The study was represented as an investigation of factors related to learning and retention of written communications. It was explained that such factors as the controversiality of the topics and whether essays were written in an emotional or factual manner were thought to play important roles in determining learning efficiency and recall of message content. In the first session, subjects read four persuasive communications, each averaging about 300 words in length. Two of the messages in each set dealt with

familiar, somewhat controversial topics, whereas the other two concerned esoteric issues.[1]

The forewarning manipulation was incorporated into the written materials to facilitate random assignment of this variable to subjects within a given class. Approximately one half of the subjects received instructions indicating that the messages were designed to change their opinions on certain issues. The directions were worded as follows:

> On the following pages you will find four passages; they are designed to persuade you. Each message will attempt to change your opinion about a particular topic. To ensure a careful reading, please pick out and underline the shortest phrase, or phrases, in each paragraph which convey the idea expressed. As you know, we are interested in the manner in which subject matter affects information processing; and its persuasive intent is thought to be one important factor. Your close attention to the task will contribute to a more adequate understanding of the process. Thank you.

In the instructions for the nonforewarned subjects, all references to the persuasive intent of the materials were deleted, and the previous verbal statement regarding the influence of familiarity and controversiality in information processing was repeated. The time interval between warning and exposure to the persuasive materials was no more than a few seconds, which Hass and Grady (1975) have shown is quite adequate.

After the four messages were read and the materials collected, subjects' opinions were assessed on all eight issues, thus providing no message control data for four topics. The justification of the opinion measurement was in terms of obtaining indices of controversiality of the issues for the population being studied. Opinions were measured on 100-point probability-of-truth scales calibrated in units of 10, ranging from 0 (Very improbable) to 100 (Very probable). The rating scale was presented immediately below each statement. One opinion item pertained to each topic. Examples include "The number of countries producing nuclear weapons is increasing rapidly" and "A substantial falling off in consumer purchasing power in the United States is presently occurring." After the opinion ratings had been collected, each subject completed a multiple-choice test of comprehension, with three items pertaining to each issue. In addition, subjects rated, on 7-point scales, how fair the articles were, how interesting they found the subject matter, and to what extent they were thinking of opposing arguments as they read the communications. Afterward, all subjects were thanked for their participation, and no mention was made of any follow-up testing.

In the second session, after a 1-week interval, subjects again supplied opinion ratings and completed the comprehension test. Thereafter, the true purpose of the study was explained. Since the experiment was represented as one dealing with comprehension and retention of message content as functions of controversiality and other characteristics of the material, the second request for opinion ratings was justified on the basis of providing indices of temporal fluctuations, since controversial issues are often quite changeable.

Results and Discussion

The opinion data were analyzed with a $2 \times 2 \times 2 \times 2$ factorial analysis of variance consisting of one between-subjects variable (forewarning vs. no forewarning) and three within-subjects variables (message type: esoteric vs. familiar, time of measurement, and experimental vs. control). To facilitate analysis, data from two subjects in the forewarned condition were randomly discarded in order to obtain equal cell frequencies. The overall persuasive impact of the messages and the effects of the experimental treatments were determined by comparing the final opinion scores for the message topics with the no-message control scores. Within the familiar and esoteric message conditions, subjects' opinion scores were averaged across the two issues.

Message Effects

The overall persuasive effect of the messages, without regard to experimental conditions, was impressive, $F(1, 92) = 62.92$, $p < .01$, for the main effect of experimental versus control treatments. As expected, subjects changed their opinions considerably more on the low-involvement, esoteric issues than on the familiar, more involving ones; $F(1, 92) = 13.75$, $p < .01$, for the interaction effect of message type in the experimental versus control conditions. Although the interaction between forewarning and experimental versus control conditions was significant, $F(1, 92)$

---

[1] The four familiar topics dealt with the proliferation of nuclear weapons, economic aid to developing African nations, the increasing number of professionally trained ministers in the United States, and the current decline in consumer purchasing power. The esoteric issues concerned the increasing length of the geological day, the effectiveness of a new drug on the peripheral vascular circulation in humans, the amount of fuel and power in Zambia, and the influence of New York City banks on the yield of U.S. government bonds.

Table 1
Mean Opinion Scores for Each of the Experimental and Control Conditions

	Familiar issues				Esoteric issues			
	Not forewarned		Forewarned		Not forewarned		Forewarned	
Condition	Immediate measurement	Delayed measurement	Immediate measurement	Delayed measurement	Immediate measurement	Delayed measurement	Immediate measurement	Delayed measurement
Experimental	72.76	61.92	61.38	65.74	73.40	63.19	57.34	63.62
Control	55.32	55.21	63.62	60.32	44.04	49.36	45.42	49.68

Note. These means are averaged across the two similar issues in each condition and are based on 100-point scales ranging from 0 (Very improbable) to 100 (Very probable). Cell *n*s = 47.

= 9.93, $p < .01$, the second-order interaction involving forewarning, message type, and experimental versus control conditions that would be expected from the work of Papageorgis (1968) was trivial ($F = .10$). That is, although subjects changed their opinions much more when the topics were low rather than high in involvement, compared to no-message controls, the inhibitory effects of forewarning were the same for both types of issues. While the data offered no support for Apsler and Sears' (1968) multiplier hypothesis, one should keep in mind that these investigators warned subjects of the specific topics and sides to be taken, whereas we simply forewarned them of persuasive intent without specifying the topic or side. Finally, in retrospect, it seems that perhaps even the familiar issues were relatively low in personal involvement, particularly for the population studied. This may explain the weakness of some of the anticipated effects.

Persistence of Forewarning Effects

The primary hypothesis tested in the present study involved the interaction between forewarning and time of measurement. It was predicted that forewarning would inhibit the immediate change resulting from the persuasive communications, but that this effect would be short-lived, and over time, the forewarned group would show an increment as the forewarning is forgotten or spontaneously dissociated, whereas those subjects who were not forewarned would show a typical decay. The test for this hypothesis is the Forewarned × Time × Experimental versus Control interaction, which was significant beyond the .01 level, $F(1, 92) = 22.12$. The pattern of means displayed in Table 1 shows that the direction of this interaction was as predicted, with subjects who were forewarned showing an initial decrement of 13.72 points, averaged across message type, compared to those individuals who were not forewarned, and, indeed, changing only trivially more than the no-message controls. After a week elapsed, however, the forewarned subjects had shown an absolute increase of 5.32 points and were now slightly higher than the nonforewarned group and substantially higher than their controls. Although the data show a relative sleeper effect (Cook & Flay, 1978), the increase did not reach significance when tested by Dunn's (1961) method; thus, an absolute sleeper effect was not obtained.

It is interesting to note the similarity of the delayed measurement means for the experimental treatments, indicating that regardless of the immediate effects of forewarning, the long-term (1-week interval) results are about the same as for subjects who were not forewarned. The findings cannot be attributed to all persons reverting to the control level during the time interval, since the overall delayed experimental mean, aggregated across treatments, was much higher than for the similar control conditions (e.g., 63.62 versus 53.64).

Forewarning Effects on the Control Issues

As stated in the introduction, it would have been virtually impossible for anticipatory belief change (such as that observed by McGuire & Millman, 1965, and others) to have occurred in the present study, since neither the topics nor the positions advocated were known prior to being read by the subjects. It is quite conceivable, however, that the forewarning might have led to some general reaction at the time of completing the postexperimental test, particularly since each subject served in both the experimental and control conditions. It can be seen in Table 1 that control subjects in the forewarned conditions were more favorable toward the familiar issues, both immediately afterward and during the delayed testing, than those subjects who were not forewarned; but no such effect occurred for the esoteric topics.

Dinner, Lewkowicz, and Cooper (1972) found greater change in high-self-esteem subjects anticipating communications concerning familiar topics, presumably so they would avoid appearing gullible. While Dinner et al. announced the topics and sides to be taken, it is possible that in the present study, subjects surmised the directions that would be advocated on the control issues and responded accordingly to avoid subsequent change should they later receive communications on these topics. However, considering the relative lack of sophistication of the high school subjects, it seems far more probable that this was just a chance occurrence, particularly since it did not appear for the esoteric topics and was not replicated in the second study for familiar issues.

Comprehension

These data were analyzed in the same manner as the opinion scores, with subjects' responses averaged across the two topics within each experimental condition. Naturally, there were no control scores for this variable, since subjects could only be asked to recall what they had read. The data indicated a slight superiority across time for students in the forewarned conditions, who obtained a mean of 1.40 correct responses out of a possible 3, compared to 1.17 correct for those who were not forewarned. The difference between these means is of borderline significance $F(1, 92) = 3.85$, $p < .06$. Forgetting during the 1-week interval was trivial ($F = 1.02$), and the only other significant effect was markedly superior memory for the familiar, compared to the esoteric, issues, $F(1, 92) = 24.60$, $p < .01$.

Reactions to the Persuasive Communication

Three questions were included in the immediate posttest to measure subjects' reactions to the persuasive communication: "How fair and unbiased did you find the communications?"; "How interesting did you find the messages?"; and "As you read the messages, to what extent did you find yourself thinking of arguments on the other side?" Subjects responded to each question by checking a 7-point scale ranging from 1 (Not at all) to 7 (Very).

Only the differences in ratings of fairness reached significance ($t = 3.07$, $p < .01$), with subjects who had been forewarned of the persuasive nature of the communications rating the contents as considerably less fair and unbiased ($M = 3.20$) than their nonforewarned counterparts ($M = 4.28$).[2] These results are similar to those found by Dean et al. (1971) with regard to messages attributed to positive sources and provide further corroboration for the discounting cue interpretation of the obtained sleeper effect in the forewarned condition. Although the obtained means for counterarguing were in the expected direction (4.00 for forewarned subjects vs. 3.46 for those persons not forewarned), the difference did not approach significance and offered no support for the conjecture that subjects whose defenses had been aroused by forewarning would think of more counterarguments while reading the communications.

In summary, these data provide strong

[2] Since these last three questions appeared separately on the last page of the booklet, some subjects apparently overlooked them and failed to respond. Consequently, the cell *n*s were reduced to 45 for the forewarned individuals and 36 in the case of those persons who were not forewarned.

evidence for the predicted sleeper effect due to forewarning of persuasive intent. They are readily interpreted in terms of forewarning serving as a discounting cue to acceptance without interfering with learning—a cue that over time (1 week) subjects are likely to forget or spontaneously dissociate, thus allowing the persuasive materials to reach their full impact. Indeed, this method of studying the sleeper effect may be superior to the classic approach of attributing a compelling message to a negative source, since the latter often must arouse feelings of incredulity on the part of the subjects.

Experiment 2

The second study was designed to replicate the first and, in addition, to test the conjecture that while moderate distraction may facilitate immediate opinion change, the latter will rapidly dissipate, because once removed from the distraction conditions, the person will think of additional counterarguments to the communications, and, to some extent, comprehension of the persuasive messages will have been impaired. Each of these factors should operate to shorten any immediate advantages realized.

Method

Subjects and Design

One hundred four high school students participated in both sessions of the experiment, which were held 1 week apart, during their normal class periods. The number of males and females was divided about equally. Each subject read two out of four persuasive messages and then stated his or her opinion on all four, thus providing control scores for two issues. Only one of the issues was read under conditions of distraction, and subjects were either forewarned or not for both issues.

A posttest-only design was employed, with subjects randomly assigned to the experimental conditions. Since distraction was a within-subjects variable, the order of presentation of the two messages (under conditions of distraction and nondistraction) was counterbalanced. The two levels of forewarning and distraction and the two times of measurement constituted a $2 \times 2 \times 2$ factorial design with one between-subjects (forewarning) and two within-subjects treatments. An equal number of subjects (52) served in each condition.

Procedure

As in Experiment 1, the study purported to investigate the effects of types of reading material on retention of message content.

In the first session, each subject read two (randomly selected from a set of four) persuasive messages averaging about 300 words in length. These communications dealt with current topics of moderate familiarity, namely, economic aid to African nations, consumer purchasing power in the United States, the increasing number of trained ministers, and proliferation of nuclear weapons. One message was read under normal circumstances and the other under distraction conditions. The distraction was designed to be relatively neutral and consisted of printing the messages in white type on a black background, rather than the conventional black on white. Pretests had shown that this reverse negative procedure was indeed distracting without being annoying, at least for the relatively short periods of time involved.[3] Furthermore, the distraction was quite consistent with the rationale given that the study investigated the influence of such factors on comprehension.

As in the previous study, the forewarning manipulation was incorporated into the written materials in order to facilitate random assignment of subjects within a given class to the different experimental conditions. For one half of the subjects, the instructions forewarned that the communications were designed to change their opinions on certain issues. For the other half, these statements were deleted; otherwise the same directions were issued to the forewarned and nonforewarned groups. All other aspects, including the purported rationale for the second opinion ratings and comprehension testing were identical to those in Experiment 1. Of course, as before, the true purpose of the experiment was eventually revealed.

Results

In the first study, the types of messages introduced the possibility that subjects could hold different initial opinions for the esoteric and familiar topics, as indeed was the case. Therefore, inclusion of the control data was necessary for the main analysis. Since no such variables were included in the second study, and there was no counterpart for the distraction manipulation under the control

[3] The materials were pretested by asking subjects to rate the extent to which they found themselves distracted or annoyed. In addition, immediate comprehension was found to be poorer for the distracted subjects—a characteristic that has been associated with distraction in several other studies.

conditions, the data for the experimental conditions were analyzed separately; and comparisons were made with the controls only when they were theoretically interesting.

Persistence of Forewarning and Distraction Effects

The opinion means for each of the experimental conditions are presented in Table 2, where it can be seen that the data generally conform to the predicted pattern.

An analysis of variance with repeated measures indicated a main effect of time of measurement, $F(1, 102) = 39.02$, $p < .01$, with subjects being less favorable after a week had elapsed; and significant interactions appeared between both forewarning, $F(1, 102) = 6.77$, $p < .01$, and distraction, $F(1, 102) = 34.55$, $p < .01$, and time of measurement. Furthermore, the second-order interaction involving all three variables was significant beyond the .01 level, $F(1, 102) = 7.69$.

The directions of the first-order interactions were as predicted: Forewarning initially inhibited opinion change, but this disadvantage vanished over time; and distraction facilitated immediate opinion change, but the initial advantage rapidly dissipated. The significant second-order interaction appears to be due primarily to the fact that in the immediate-measurement condition, the distraction completely nullified the effects of forewarning, yielding a difference of more than 20 points between the distracted and nondistracted subjects' means. This fact is particularly interesting, since it is exactly what would be expected if the major mediating variable was the number of counterarguments produced. That is, if the inhibiting effect of forewarning of persuasive intent is primarily due to the increased production of counterarguments while reading the messages, then, when subjects are distracted from counterarguing, the usual effects produced by forewarning should be nullified.

It is of interest to determine whether the data replicate the earlier finding of a sleeper effect interaction between forewarning and time of opinion measurement when subjects are not distracted. The relevant data are contained in the no-distraction conditions of Table 2 (identical to the earlier study). The immediate effect of forewarning of persuasive intent was to reduce substantially the impact of the messages (Ms = 53.85 vs. 70.77 in the nonforewarned conditions). After a week's delay, however, this initial difference had all but disappeared, with the forewarned group showing an increase of 7.88 points. Planned comparisons (Dunn, 1961) indicated that the interaction and the aforementioned increase were significant beyond the .05 level. Hence, this study showed both relative and absolute sleeper effects for the forewarned groups.

Table 2
Mean Opinion Scores for Each of the Experimental and Control Conditions

Condition	Not forewarned		Forewarned	
	Measurement		Measurement	
	Immediate	Delayed	Immediate	Delayed
No distraction	70.77	63.85	53.85	61.73
Distraction	75.38	59.42	74.42	57.12
Controls	54.42	54.23	47.12	44.52

Note. These means are based on 100-point scales ranging from 0 (Very improbable) to 100 (Very probable). Cell ns = 52.

In regard to the effects of distraction, the data for the no-forewarning conditions, which would be most similar to the typical distraction study, showed the predicted interaction. Distraction facilitated opinion change immediately after reading the communications ($M = 75.38$ for the distracted individuals, compared to 70.77 for those persons in the no-distraction conditions), but after a week, this pattern had reversed, with the nondistracted subjects showing somewhat superior retention of opinion change. This interaction was also significant beyond the .05 level as tested by Dunn's (1961) method.

In summary, there appears to be considerable support for the conjecture that moderate distraction may have a facilitating, but short-lived, effect on opinion change. There was no evidence in the present study that forewarning increased the favorability of sub-

jects' responses to the familiar control issues. In contrast, when only familiar issues were presented, subjects who were forewarned agreed somewhat less strongly with the opinion statements on both the immediate and delayed measurements. There seems to be no ready explanation for the different results obtained with the control issues in the two experiments.

Comprehension

The prediction that forewarning would serve as a discounting cue and lead to a sleeper effect was based on the assumption that it would not interfere with learning of the message content. In contrast, the predictions involving distraction were based in part on the assumption that distraction would interfere with learning and that, consequently, induced change would be quite ephemeral. The mean comprehension scores for each of the experimental conditions are presented in Table 3.

Analysis of variance indicated that the main effect of distraction was significant, $F(1, 102) = 28.33$, $p < .01$, and in the predicted direction. The main effect of forewarning was in the same direction as that obtained in the first study, $F(1, 102) = 2.57$, $p \approx .11$. In contrast to the earlier study, time of measurement had a significant effect, $F(1, 102) = 52.36$, $p < .01$, with subjects remembering less of the messages' contents at the time of delayed measurement. The only other significant effect was the interaction between fore warning and distraction, $F(1, 102) = 4.1$ $p < .05$. There was a greater facilitating e fect of forewarning on learning when subjec were not distracted. Hence, in the prese study, subjects learned significantly less u der conditions of distraction even thoug their opinions were changed much mor These results are quite similar to those ob tained by Insko, Turnbull, and Yande (1974) and Petty et al. (1976), in whos studies, at least under some conditions, di traction interfered with recall but facilitate opinion change. Together, these studies sug gest some limitations of the generalizatio advanced by previous researchers (e.g., Fes inger & Maccoby, 1964; Osterhouse & Broc 1970; Regan & Cheng, 1973) that distractio facilitates opinion change only in cases whe it does not interfere with reception of th message. Naturally, if the interference wit comprehension were severe, the distractio would inhibit opinion change, since one ca only change his or her opinion in the direc tion advocated if he or she knows what tha position is.

Table 3
Mean Comprehension Scores for Each of the Experimental Conditions

Condition	Not forewarned		Forewarned	
	Measurement		Measurement	
	Immediate	Delayed	Immediate	Delayed
No distraction	1.92	1.38	2.19	1.83
Distraction	1.58	1.23	1.65	1.25

Note. These means are based on the number of correct answers to three multiple-choice questions. Cell *ns* = 52.

General Discussion

Effects of Forewarning

Both studies indicated that forewarning o of persuasive intent of a communication pro duced a sleeper effect whereby subjects were more influenced by the message after a week' interval than they were immediately after read ing it. There were at least three reasons fo expecting such a delayed-action effect. The first pertained to the number of counterargu ments evoked by forewarning. It was assumed that if forewarning increased counterarguing less immediate opinion change would result, but that over time, the subjects' reactions would be forgotten or dissociated from the messages, thus allowing the information to have a greater impact. In the first study, forewarned subjects showed a slightly greater tendency toward counterarguing; but the effect fell short of significance. This may have been due, in part, to the insensitivity of the response rating scale employed.

The second reason for expecting a delayed-action effect rested on the assumption that forewarning would increase the subjects' tendencies to derogate the communications. As found in earlier studies (e.g., Dean et al., 1971; Hass & Grady, 1975), subjects rated the messages as less fair and unbiased when forewarned of their persuasive intent. Thus, forewarning apparently served as a discounting cue (much like a negative source), thereby reducing the immediate impact of the communication. However, with the passage of time, subjects tended to forget or dissociate the discounting cue, thus allowing the full impact of the persuasive material to emerge, providing that forewarning did not interfere with comprehension. Consequently, the forewarned subjects would be expected to reach the same level as the nonforewarned group over time. Obviously, whether an absolute increase occurred, as witnessed in the second study, would depend on such factors as the period of time elapsed and the decay rate for the nonforewarned group. This reasoning parallels the original interpretation of the sleeper effect (Hovland et al., 1949), in which it was presumed that the propaganda-type context led to an initial dampening of the film's persuasive impact, but that as this discounting cue was forgotten over time, the subjects accepted the message to a greater extent. Despite the fact that Gillig and Greenwald (1974) have written its obituary, and Capon and Hulbert (1973) have concluded that there is no strong evidence for a generalized sleeper effect, the data in these studies fully supported an opposing stance. Indeed, Gruder et al. (1978) have recently shown that absolute sleeper effects can be obtained when certain conditions are met for strong tests: (a) when a persuasive message has a substantial initial impact on attitudes, (b) when this change is totally inhibited by a discounting cue, (c) when the cue and message are dissociated over time, and (d) when this dissociation occurs quickly enough so that the message still has some impact. The present studies generally would appear to meet these criteria.

As previously mentioned, forewarning may be a better method of studying the sleeper effect than attributing a negative source to a compelling message, since this must often arouse feelings of incredulity, and furthermore, the negative source may be overshadowed by the mention of a number of positive sources in the communication itself. For example, a health message may be attributed to a low-prestige source, such as a high school student, but within the text, various facts may be presented that presumably come from medical journals, learned professors, or prominent physicians. The result could conceivably end in confusion.

The third possibility, that forewarning increases reactance, was not tested in these studies; but again, this view would lead to the predicted delayed-action effect. As Hass and Grady (1975) pointed out, despite the presence of reactance, well-written communications containing compelling arguments often lead to persuasion. The arguments should have greater influence on the subject once the reactance has dissipated.

Actually, all three of the interpretations mentioned above can be viewed as variations of the discounting cue hypothesis, with counterarguments, derogation of source and message, and feelings of reactance each serving as an initial rejection cue.

Effects of Distraction

More rapid decay of opinion change, induced under conditions of distraction, was expected for two reasons. First, distraction has been shown to interfere with comprehension of message content (e.g., Petty et al., 1976; Zimbardo et al., 1970). While correlational studies of memory and persistence have produced somewhat inconsistent results (e.g., Cook & Flay, 1978; Miller & Campbell, 1959; Watts & McGuire, 1964), Cook and Flay point out that experimental manipulations affecting learning have usually influenced persistence of opinion change. Thus, while the immediate effects of distraction may be to increase opinion change (presumably because more is gained from the reduction in counterarguing than is lost through poor comprehension), the long-term prognosis would be a rapid reversion to the level of the nondis-

tracted subjects, if not to a lower level. The data in the present study are consistent with this view, inasmuch as distracted subjects changed more immediately afterward, but not during the delayed measures of opinions; and they scored significantly lower on the comprehension test at both time intervals.

The second reason for expecting the obtained temporal effects depends on postexperimental counterarguing. Although this variable was not assessed in the present study, the fact that distraction completely nullified the immediate effects of forewarning strongly suggests that counterarguing is a major mediating process. After the subject has left the experimental room, it is only reasonable to assume that he or she would be able to think of a greater number of counterarguments to the communications when free to reflect on them without being distracted. This phenomenon should, in turn, serve to move the person's opinion back toward the preexperimental level. It is regrettable that both immediate and delayed counterarguing were not measured in the second study. The rating scale that was used in the first study seemed inadequate, and the more thorough thought-listing procedure (e.g., Brock, 1967) would have been too complicated for our within-subjects design.

References

Apsler, R., & Sears, D. O. Warning, personal involvement, and attitude change. *Journal of Personality and Social Psychology*, 1968, *9*, 162–166.

Baron, R. S., Baron, P. H., & Miller, N. The relation between distraction and persuasion. *Psychological Bulletin*, 1973, *80*, 310–323.

Brehm, J. W. *A theory of psychological reactance*. New York: Academic Press, 1966.

Brock, T. C. Communication discrepancy and intent to persuade as determinants of counterargument production. *Journal of Experimental Social Psychology*, 1967, *3*, 296–309.

Capon, N., & Hulbert, J. The sleeper effect—An awakening. *Public Opinion Quarterly*, 1973, *37*, 333–358.

Cook, T. D., & Flay, B. R. The temporal persistence of experimentally induced attitude change. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 11). New York: Academic Press, 1978.

Dean, R. B., Austin, J. A., & Watts, W. A. Forewarning effects in persuasion: Field and classroom experiments. *Journal of Personality and S Psychology*, 1971, *18*, 210–221.

Dinner, S. H., Lewkowicz, B. E., & Cooper, J. ticipatory attitude change as a function of esteem and issue familiarity. *Journal of Perso ity and Social Psychology*, 1972, *24*, 407–412.

Dunn, O. J. Multiple comparisons among me *Journal of the American Statistical Associat* 1961, *56*, 52–64.

Festinger, L., & Maccoby, N. On resistance to suasive communications. *Journal of Abnormal Social Psychology*, 1964, *68*, 359–366.

Freedman, J. L., & Sears, D. O. Warning, distracti and resistance to influence. *Journal of Personal and Social Psychology*, 1965, *1*, 262–266.

Gillig, P. M., & Greenwald, A. G. Is it time to the sleeper effect to rest? *Journal of Personali and Social Psychology*, 1974, *29*, 132–139.

Gruder, C. L., Cook, T. D., Hennigan, K. M., Fl B. R., Alessis, C., & Halamaj, J. Empirical te of the absolute sleeper effect predicted from t discounting cue hypothesis. *Journal of Person ity and Social Psychology*, 1978, *36*, 1061–1074.

Haaland, G. A., & Venkatesan, M. Resistance to pe suasive communications: An examination of t distraction hypotheses. *Journal of Personality an Social Psychology*, 1968, *9*, 167–170.

Hass, R. G., & Grady, K. Temporal delay, type forewarning, and resistance to influence. *Journal Experimental Social Psychology*, 1975, *11*, 459–4

Hovland, C. I., Lumsdaine, A. A., & Sheffield, F. *Experiments on mass communication*. Princeto N.J.: Princeton University Press, 1949.

Insko, C. A., Turnbull, W., & Yandell, B. Facilita tive and inhibiting effects of distraction on a titude change. *Sociometry*, 1974, *37*, 508–528.

Kiesler, C. A., & Kiesler, S. B. Role of forewarni in persuasive communications. *Journal of Abno mal and Social Psychology*, 1964, *68*, 547–549.

McGuire, W. J. Attitudes and opinions. *Annual Re view of Psychology*, 1966, *17*, 475–514.

McGuire, W. J. Personality and susceptibility to so cial influence. In E. F. Borgatta & W. W. Lambe (Eds.), *Handbook of personality theory and re search*. Chicago: Rand McNally, 1968.

McGuire, W. J. The nature of attitudes and at titude change. In G. Lindzey & E. Aronson (Eds.) *The handbook of social psychology* (Vol. 3) Reading, Mass.: Addison-Wesley, 1969.

McGuire, W. J., & Millman, S. Anticipatory belief lowering following forewarning of a persuasive at tack. *Journal of Personality and Social Psychology*, 1965, *2*, 471–479.

McGuire, W. J., & Papageorgis, D. Effectiveness of forewarning in developing resistance to persua sion. *Public Opinion Quarterly*, 1962, *26*, 24–34.

Miller, N., & Campbell, D. T. Recency and primacy in persuasion as a function of the timing of speeches and measurements. *Journal of Abnormal and Social Psychology*, 1959, *59*, 1–9.

Osterhouse, R. A., & Brock, T. C. Distraction increases yielding to propaganda by inhibiting coun-

terarguing. *Journal of Personality and Social Psychology*, 1970, *15*, 344–358.

Papageorgis, D. Warning and persuasion. *Psychological Bulletin*, 1968, *70*, 271–282.

Petty, R. E., & Cacioppo, J. T. Forewarning, cognitive responding, and resistance to persuasion. *Journal of Personality and Social Psychology*, 1977, *35*, 645–655.

Petty, R. E., Wells, G. L., & Brock, T. C. Distraction can enhance or reduce yielding to propaganda: Thought disruption versus effort justification. *Journal of Personality and Social Psychology*, 1976, *34*, 874–884.

Regan, D. T., & Cheng, J. B. Distraction and attitude change: A resolution. *Journal of Experimental Social Psychology*, 1973, *9*, 138–147.

Vinokur, A., & Burnstein, E. Depolarization of attitudes in groups. *Journal of Personality and Social Psychology*, 1978, *36*, 872–885.

Vinokur, A., Trope, Y., & Burnstein, E. A decision-making analysis of persuasive argumentation and the choice-shift effect. *Journal of Experimental Social Psychology*, 1975, *11*, 127–148.

Watts, W. A., & McGuire, W. J. Persistence of induced opinion change and retention of the inducing message content. *Journal of Abnormal and Social Psychology*, 1964, *68*, 233–241.

Wyer, R. S., Jr. *Cognitive organization and change: An information processing approach*. Potomac, Md.: Erlbaum, 1974.

Zimbardo, P. G., Snyder, M., Thomas, J., Gold, A., & Gurwitz, S. Modifying the impact of persuasive communications with external distraction. *Journal of Personality and Social Psychology*, 1970, *16*, 669–680.

Received May 24, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 5, 790-808

The Role of Assumed Reciprocation of Sentiment and Assume Similarity in the Production of Attraction and Agreement Effects in *p-o-x* Triads

Chester A. Insko
University of North Carolina at Chapel Hill

Akanbi Adewole
University of Lagos
Lagos, Nigeria

Two experiments were conducted to explore the adequacy with which balance theory can account for attraction and agreement effects in *p-o-x* triads. The first experiment provided evidence for the importance of assumed reciprocated sentiment in the production of attraction effects, and the second experiment provided evidence for the importance of assumed *p-o* similarity in the production of agreement effects. In addition, both experiments replicated past results concerning the importance of contact for attraction and agreement effects and also found that attraction and agreement effects were more apparent with affective scales (pleasantness and harmony) than with more cognitive scales (expectancy, consistency, and stability). An unexpected finding of the second experiment was that the *p-o* similarity (or social comparison) interpretation of the agreement effect held only for the more cognitive scales.

The present article addresses the general issue of the extent to which balance theory can be elaborated to account for subjects' reactions to hypothetical social situations, like Heider's (1946, 1958) *p-o-x* (or self–other–inanimate object) triad. What is the reason for studying the subjects' reactions to hypothetical social situations? There are at least two possible answers to such a question. One is that such studies are important because they are role-playing simulations of recurrent social situations. (See the Summer, 1977, issue of the *Personality and Social Psychology Bulletin* for a series of articles arguing the pros and cons of such role-playing methodology.) A second, and more compelling, answer to the question asked above relates to the fact that such studies provide an arena for theory development.

There is a tendency to downgrade the importance of research in hypothetical situations because it is "irrelevant." To some extent such a reaction is based on a negative reaction to balance theory, and to some exte it is based on the failure to recognize th from a balance theory perspective, phenom nology is phenomenology—whether it relat to actual or hypothetical situations. There thus the possibility of working out principl that may generally apply. Research on hyp thetical triads can be conceived of as playin a role for the balance theorist that is analo gous to the role played by the Skinner bo for the operant psychologist. The Skinne box, like the hypothetical triad, holds litt intrinsic interest or "relevance." On the oth hand, the Skinner box does provide a situ tion in which variables believed to be theoretical importance can be easily manipu lated and behavior can be easily measured Thus, principles may be worked out that ca be applied to other situations.

Insko, Songer, and McGarvey (1974) hav argued that Heider's *p-o-x* triad can best b conceptualized from the perspective of a three-factor analysis of variance model. The first factor is the positive or negative sign of the *p-o* relation, the second is the positive or negative sign of the *p-x* relation, and the third is the positive or negative sign of the

Requests for reprints should be sent to Chester A. Insko, Department of Psychology, University of North Carolina, Chapel Hill, North Carolina 27514.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3705-0790$00.75

o-x relation. Within-subjects ratings of the eight possible triads (typically for degree of pleasantness) give seven possible orthogonal effects (three main effects, three double interactions, and a triple interaction). Numerous studies (see Zajonc, 1968, for a review) have found evidence for three of the seven effects. These are the main effect for the *p-o* relation, or attraction effect; the interaction between the *p-x* and *o-x* relations, or agreement effect; and the triple interaction, or three-sign balance effect.[1] The attraction effect is a tendency for subjects to rate *p*-likes-*o* triads as more pleasant than *p*-dislikes-*o* triads. The agreement effect is a tendency for subjects to rate triads in which *p* and *o* both like or both dislike *x* as more pleasant than triads in which *p* and *o* do not agree about *x*. The three-sign balance effect is a tendency for subjects to rate the triads in which the product of the three signs is positive as more pleasant than triads in which the product of the three signs is negative. Stated another way, it is the tendency to rate triads in which *p* likes an *o* with whom there is agreement, or dislikes an *o* with whom there is disagreement, as more pleasant than the triads in which *p* likes an *o* with whom there is disagreement, or dislikes an *o* with whom there is agreement.

It can be argued that attraction and agreement effects are embarrassing to balance theory (cf. Zajonc, 1968). Insko et al. (1974), however, have pointed out that various balance interpretations are possible. These interpretations are based on the general assumption that subjects supply additional cognitive bands beyond the three that are experimentally given. For example, according to Aderman (1969), when subjects consider the *p-o-x* triad, there is a tendency for some of them to think about a specific *o* with whom they interact or have contact. In the language of balance theory, such social interaction is a positive unit relation between *p* and *o*. Insko et al. argued that the addition of this positive unit relation to the *p-o-x* triad (or cycle) produces two additional cycles that provide balance interpretations of attraction and agreement effects.

Counting the research in Insko et al. as one experiment, and including the two experiments reported by McGarvey (1974) as well as an additional unpublished experiment (LaTour et al., Note 1), there are four experiments that have studied the important contact conditions. There are four such conditions: standard (with no mention of contact), future contact, no contact, and breaking contact. According to the unit-relation interpretation, the algebraic magnitudes of the attraction and agreement effects should be as follows: future contact > standard > no contact > breaking contact. Furthermore, the direction of the effects should be positive in the future contact and standard conditions, neither positive nor negative in the no-contact condition, and negative in the breaking-contact condition. With regard to the attraction effect, results have been generally consistent with the predictions given above. The exception is a significant attraction effect in the no-contact condition in one of the four experiments. With regard to the agreement effect, results have been less supportive of the unit-relation interaction. All experiments have obtained the same rank order of conditions: future contact $\simeq$ standard > no contact > breaking contact. Furthermore, three of the four experiments obtained a significant agreement effect in the no-contact condition, and one failed to find a significant negative (or reversed) agreement effect in the breaking contact condition. Clearly the contact interpretation does not adequately account for all of the reliable variance in the agreement effect.

One further complication relates to the fact that some of the studies mentioned above had two future contact conditions: future contact with discussion of *x* and future contact without discussion of *x*. The difference between these two conditions does not obviously reflect either contact or similarity, and the reason for the previously obtained finding of a greater agreement effect with than without discussion (e.g., Insko et al., 1974) is not

[1] Insko, Songer, & McGarvey (1974) refer to attraction effects as "positivity" effects. Since usage seems to be following Zajonc's precedent, we will conform.

clear. Perhaps the explicit elimination of discussion regarding x decreases the importance or weight of the agreement cycle, or perhaps discussion of x produces greater concern with the consequences of disagreement.

As previously indicated, there are several possible interpretations of the attraction effect. One relates to contact, another to reciprocated sentiment. Existing evidence (McGarvey, 1974) indicates that at least some subjects do assume same-sign reciprocation of the *p-o* relation, and research by McGarvey (1974) indicates that a manipulation of reciprocated sentiment alters the magnitude of the attraction effect. Why should this be? Reciprocated positive sentiment produces a balanced cycle, as does also reciprocated negative sentiment. Thus, there would seemingly be no basis for predicting an attraction effect. On the other hand, to the extent that the self, or self-concept, is positively valued, reciprocated positive sentiment produces agreement regarding the worth of the self, and reciprocated negative sentiment produces disagreement regarding the worth of the self. Thus, from the perspective of reciprocated sentiment, the attraction effect is an agreement effect. A crucial issue, therefore, is the extent to which there are balance theory explanations of agreement effects. These explanations may depend on various types of *p-o* unit relations (contact, assumed similarity, etc.).

Experiment 1 was primarily designed to explore the reciprocated sentiment interpretation of attraction effects. Thus, one of the factors was a manipulation of reciprocated sentiment. Beyond this, however, this experiment was also designed to explore the relationship between reciprocated sentiment and other factors of possible theoretical importance. One of these factors relates to the type of rating scale. Existing research by Crockett (1974); Gutman and Knox (1972); Gutman, Knox, and Storm (1974); and Miller and Norman (1976) suggests that pleasantness scales reveal larger attraction effects than do consistency and/or expectancy ratings. Experiment 1 contains a scales factor that includes five different rating scales: pleasantness, harmony, expectancy, consistency, and stability.

It is reasonable to assume that affect is self-related and that an affective rating makes the self more salient than does a more cognitive rating, such as consistency. Pleasure and pain are self-related and cannot exist apart from persons (or organisms). Consistency, however, need not be self-related and certainly can exist apart from persons (e.g., between logical symbols). Thus, to the extent that the attraction effect results from assumed reciprocated sentiment (with its implication of agreement or disagreement regarding the worth of the self), affective ratings will produce a larger attraction effect than will cognitive ratings. In terms of the present design this is a prediction that a contrast effect comparing affective (pleasantness and harmony) scales with cognitive (expectancy, consistency, and stability) scales will be significant.

Recall that McGarvey (1974) found a main effect of reciprocation on attraction using just a pleasantness scale. We predicted that in the present study there would be an interaction between reciprocation and rating scales such that the effect of reciprocation would be more apparent for the affective scales (pleasantness and harmony) than for the relatively more cognitive scales (expectancy, consistency, and stability). The theoretical rationale for this prediction was that reciprocated sentiment produces agreement or disagreement regarding the self-concept, and the affective scales are more directly relevant to cycles including the self-concept as an element (e.g., p interacts with o, p approves of his/her self-concept, o disapproves of p's self-concept). We are assuming that the affective scales are most relevant to cycles that include the self-concept as an element (i.e., cycles in which p's self-evaluation provides one of the signs that determines balance or imbalance in the cycle).

A third factor having a bearing on the reciprocated sentiment interpretation of attraction effects is involvement, or a comparison of *p-o-x* and *q-o-x* triads. With *p-o-x* triads, the subject is told that p, or "I" is the subject himself/herself. With *q-o-x* triads, the subject is asked to assume that q is some person

who is known but has never been met. Subjects are asked to rate the *q-o-x* triads according to how they as *outside observers* would find the situation. In view of the argument detailed above regarding the role of self-relevancy in the production of the attraction effect, high-involvement (*p-o-x*) triads should produce a greater attraction effect than low-involvement (*q-o-x*) triads. However, a study by LaTour et al. (Note 1) did not support this prediction. It is possible, however, that subjects tend to judge *q-o-x* triads by vicariously identifying with *q*. Some evidence for vicarious identification in the context of *p-o-q* (three-person) triads was obtained by Insko et al. (1974). In that study, *o*-to-*q* attraction was found under the special circumstances of *q*-to-*o* same-sign reciprocation. Since same-sign reciprocation also increased the magnitude of *p-q* attraction, it was argued that *o-q* attraction resulted from *p*'s vicarious identification with *o*. Thus, it is plausible that subjects judge the interpersonal situation through at least partial vicarious identification.

We believed it unlikely that vicarious identification would be sufficient to explain all of the subtleties present in *p-o-x* triads. If this is true, it is reasonable to suppose that the prediction given above regarding the greater effect of reciprocation with affective than cognitive scales would be more apparent for *p-o-x* than *q-o-x* triads. This is a triple interaction: Involvement × Reciprocation × Scales. To the extent that this interaction occurs, we would have further support for the supposition that the self-concept is important in an adequate conceptualization of attraction effects.

Finally, we come to the three-sign balance effect. Past results have uniformly indicated a greater effect in the standard condition than in the remaining contact conditions. Such results possibly indicate an attentional shift away from the three-sign balance cycle (that does not include the unit relation) to the attraction and agreement cycles (that do include the unit relation). No further predictions regarding three-sign balance were made.

Experiment 1 contained four between-subjects factors: reciprocation (standard no mention, or explicitly mentioned), scales (pleasantness, harmony, expectancy, consistency, or stability), involvement (*p-o-x* or *q-o-x* triads), and contact (standard, future contact with discussion of *x*, future contact without discussion of *x*, no contact, or breaking contact). The contact factor was included in order to examine the extent to which contact results would generalize across scales.

Experiment 1 did not include a manipulation of triad order. This was not done for two reasons. First, the existing design of four between-subjects factors generates 100 cells, and that is the capacity of the available multivariate analysis of variance program. Second, past research (Insko et al., 1974; McGarvey, 1974) has not found any interactive effects between triad order and other manipulated factors.

Experiment 1

Method

Subjects

The subjects were 1,000 male and female students from the introductory psychology course at the University of North Carolina.

Independent and Dependent Variables

Each subject rated eight triads. These ratings can be conceived of as generating three within-subjects factors: positive or negative sign of the first relation (*p* to *o* or *q* to *o*), positive or negative sign of the second relation (*p* to *x* or *q* to *x*), and positive or negative sign of the third relation (*o* to *x*). These within-subjects factors were used to generate seven orthogonal difference scores corresponding to the main effects and interactions of these factors. Three of these difference scores, the attraction, agreement, and three-sign balance effects, were used as the main dependent variables in a multivariate analysis of variance that included four between-subjects factors. The four between-subjects factors were involvement (*p-o-x* or *q-o-x* triads), reciprocation (standard no mention of reciprocation, or same-sign reciprocation of *p*- or *q*-to-*o* relation), contact (standard, future contact with discussion of *x*, future contact without discussion of *x*, no contact, or breaking contact), and rating scales (pleasantness, expectancy, harmony, stability, or consistency).

Procedure

Subjects were tested in small groups of up to approximately 20 people. In any given group, every

subject was in a different one of the 100 conditions or cells. After subjects had been briefly instructed regarding the procedure for rating each hypothetical situation, they were given a nine-page booklet. Subjects were asked to regard each situation as independent of the others and not to alter a previous response after a page had been turned. It was emphasized that they should rate each of the eight situations and not leave any of the rating scales blank. Remarkably, this instruction was followed by all 1,000 subjects, and no subjects were eliminated from the final sample.

The first page of the booklet contained instructions regarding the experimental procedure and introduced two of the between-subjects factors, involvement and rating scales. For example, the self-involved–consistency instruction was as follows:

> On each of the following eight pages is a description of a situation involving two persons, I and O, and an impersonal object or issue, X. As you read each situation, try and imagine yourself as Person I. Think about what it would be like to be involved in the situation described. Then rate each situation according to the extent to which *you as a person, I,* would find such an interpersonal situation consistent or inconsistent. Make the rating by circling the number on the scale corresponding to how *you as Person I* find the situation. Below is an example of the scale.
>
> Most inconsistent 1 2 3 4 5 6 7 Most consistent

For the self-uninvolved–harmonious-scale condition, the instruction was the following:

> On each of the following eight pages is a description of a situation involving two persons, Q and O, and an impersonal object or issue, X. Imagine that you have never met Q and O but that you know of them and how they feel about X. Rate each situation according to the extent to which *you as an outside observer* would find such an interpersonal situation harmonious or unharmonious. Make the rating by circling the number on the scale that corresponds to how *you as the outside observer* find the situation. Below is an example of the scale.
>
> Most discordant 1 2 3 4 5 6 7 Most harmonious

As previously indicated, five different rating scales were used (most unpleasant – most pleasant, most unexpected – most exepected, most inconsistent – most consistent, most discordant – most harmonious, most changeable – most stable). These rating scales were repeated on each of the eight pages. Each page stated one of the triadic situations, indicated reciprocation (in the appropriate condition), stated assumptions regarding contact (in the nonstandard conditions), and finally gave a rating scale. Here is an example.

> Situation: I like O. I like X. O dislikes X. O likes I (me).
>
> Assume: I and O have not had personal contact in the past and will never have personal contact in the future. However, I and O know enough about each other to form some feelings about each other.
>
> Rating of situation:
>
> Most unpleasant 1 2 3 4 5 6 7 Most pleasant

Reciprocation was manipulated by stating O's feeling for I (or Q) as the same as I (or Q's) feeling for O, or by not mentioning anything about O's feelings. Contact was manipulated by stating (in all but the standard condition) an assumption regarding contact. The assumptions were as follows: I and O have had contact in the past and will have contact in the future and will discuss X; I and O have had personal contact in the past and will have personal contact in the future but will not discuss X; I and O have not had personal contact in the past and will never have personal contact in the future. However, I and O know enough about each other to form some feelings about each other; and I and O have had personal contact in the past but will never have personal contact in the future.

At the conclusion of the experiment, subjects were told about the purposes of the study.

Dependent Variables

It is worth emphasizing that the dependent variables were the seven orthogonal difference scores created by taking the difference between the mean of the ratings for a given four triads and the ratings of the remaining four triads. It is true that the subjects rated the triads for pleasantness, consistency, and so forth. The dependent variables, however, were not pleasantness and consistency, but rather the difference scores. From this perspective the rating scales constituted an independent variable.

Results

Grand Mean Tests

The initial analysis involved a test of the difference from zero for each of the seven orthogonal difference scores (or contrasts). This analysis revealed the usual significant effects for attraction, $F(1, 900) = 179.36$, $p < .01$; agreement, $F(1, 900) = 171.43$, $p < .01$; and three-sign balance, $F(1, 900) = 288.18$, $p < .01$. In addition there was also a significant effect for the second sign (p-x or q-x), $F(1, 900) = 11.36$, $p < .01$. This effect indicates a tendency for the p-likes-x and q-likes-x triads to be rated as more favorable (pleasant, etc.) than the triads in which p dislikes x and q dislikes x. Insko et al. (1974)

Table 1
Multivariate Analysis of Variance for Experiment 1

Source	Attraction F	Agreement F	Three-sign balance F	Multivariate F
Involvement (I)	13.48**	.22	4.17**	4.18**
Reciprocation (R)	4.79*	2.47	.87	2.01*
Contact (C)	80.39**	37.68**	22.44**	21.36**
Scales (S)	36.60**	22.44**	6.73**	9.70**
I × R	.34	.15	.30	.55
I × C	2.03	1.60	3.19*	1.29
I × S	.96	1.90	1.46	.99
R × C	1.57	3.84**	1.69	1.55*
R × S	6.90**	1.40	.72	1.79**
C × S	4.32**	.73	2.25**	1.77**
I × R × C	.11	1.06	.85	.84
I × R × S	2.90*	.66	1.57	1.63*
I × C × S	1.17	1.11	1.44	1.09
R × C × S	1.02	1.21	1.57	1.25
Quadruple interaction	1.70*	.96	1.53	1.18

* $p < .05$. ** $p < .01$.

found a similar effect. A possible balance interpretation of this effect would involve a conceptualization of an x with which p or q has a positive unit relation—a recently purchased bicycle, for example. However, since none of the between-subjects factors were designed to explore this matter, since the effect was small relative to the other three effects, and since the analysis is already complex, reported results will relate solely to the attraction, agreement, and three-sign balance effects.

Multivariate Analysis of Variance

A four-factor multivariate analysis of variance of the attraction, agreement, and three-sign balance effects is presented in Table 1. This analysis revealed significant multivariate Fs for all four main effects, as well as for four of the interactions. Discussion of the significant univariate Fs accompanying the significant multivariate Fs is complex. This material will be presented under three subheadings: rating scales main effect, involvement and reciprocation, and contact.

Rating Scales Main Effect

There was a significant main effect for rating scales. Table 2 contains the marginal means. A contrast comparing the affective scales (pleasantness and harmony) with the more cognitive scales (expectancy, stability, and consistency) was significant for both attraction and agreement. As expected, attraction, $F(1, 900) = 162.12$, $p < .01$, and

Table 2
Rating Scale Means for Experiment 1

Condition	Pleasantness	Harmony	Expectancy	Consistency	Stability
Attraction	5.71**	5.90**	.49	1.02*	.67
Agreement	3.59**	3.84**	.05	.94**	1.78**
Three-sign balance	2.28**	2.08**	1.84**	2.14**	3.92**

Note. Each subject rated eight triads on one of five 7-point scales. These ratings were then converted to attraction, agreement, and three-sign balance scores by taking the difference between the mean of the appropriate four triads and the mean of the remaining four triads. These difference scores (with a possible +7 to −7 range) were then averaged across subjects to give the tabled means.
* $p < .05$, $df = 1, 900$. ** $p < .01$, $df = 1, 900$.

agreement, $F(1, 900) = 89.06$, $p < .01$, effects were greater with the affective scales than with the cognitive.

Grand mean tests were conducted on each of the cell means in Table 2. The significance of these tests of differences from zero is indicated by the asterisks above each mean. All of the means except three (expectancy ratings for attraction and agreement and stability ratings for attraction) differ from zero. The cognitive scales tend to show smaller attraction and agreement effects, with almost no effects at all appearing in ratings of expectancy.

Involvement and Reciprocation

Involvement main effect. There was a significant involvement effect for both attraction and three-sign balance. The marginal means indicated that the attraction effect was greater with self-uninvolved triads (*q-o-x*) than with self-involved (*p-o-x*) triads (M *q-o-x* = 3.52, M *p-o-x* = 2.00), while just the opposite was true for the three-sign balance effect (M *p-o-x* = 2.75, M *q-o-x* = 2.16). This finding suggests a tendency for subjects to react more to the simpler aspects of the triads when they are not personally involved and to react more to the complex aspects of the triads when they are personally involved. Attraction involves one relation, and three-sign balance involves three relations.

Reciprocation main effect. The only significant reciprocation main effect was for attraction. Consistent with McGarvey's (1974) results (which related only to pleasantness ratings of *p-o-x* triads), same-sign reciprocation of the *p-o* or *q-o* relation increased the magnitude of the attraction effect (3.21) over standard (no stated) reciprocation (2.31). As previously indicated, assuming a positively evaluated self-concept, reciprocated positive sentiment produces agreement regarding the self, and reciprocated negative sentiment produces disagreement regarding the self. Thus, from the perspective of reciprocated sentiment the attraction effect can be regarded as an agreement effect.

Reciprocation × Scales interaction. The main effect for reciprocation on attraction was qualified by several interactions, one of which was the Reciprocation × Scales interaction. The cell means and the significance of the grand mean tests for the Reciprocation × Scales × Involvement interaction are given in Table 3. The Reciprocation × Scales interaction for attraction was explored by testing the interaction involving reciprocation and the contrast between affective scales (pleasantness and harmony) and cognitive scales (expectancy, stability, and consistency). The contrast was significant, $F(1, 900) = 23.85$, $p < .01$. The interaction involving this contrast indicates that the main effect for reciprocation is a resultant of the affective scales. The pattern can be seen in Table 3.

An alternative view of this interaction is a qualification of the tendency for affective scales to produce a larger attraction effect than cognitive scales. This aspect of the main effect for rating scales is more apparent

Table 3
Involvement × Reciprocation × Scales Attraction Means for Experiment 1

Involvement and reciprocation	Pleasantness	Harmony	Expectancy	Consistency	Stability
p-o-x triad					
Standard (no mention of) reciprocation	3.28**	2.06*	.60	1.56	−.34
Reciprocation	7.04**	7.74**	.18	−1.22	−.86
q-o-x triad					
Standard (no mention of) reciprocation	4.90**	6.32**	−.46	1.86*	3.32**
Reciprocation	7.60**	7.50**	1.64	1.90*	.58

Note. *p-o-x* triad indicates high personal involvement; *q-o-x* triad indicates low personal involvement.
* $p < .05$, $df = 1, 900$. ** $p < .01$, $df = 1, 900$.

Table 4
Contact Means for Experiment 1

Condition	Standard (no mention of) contact	Future contact with discussion	Future contact without discussion	No contact	Breaking contact
Attraction	2.86*	7.45*	4.20*	3.19*	−3.88*
Agreement	2.75*	4.79*	−.21	2.94*	−.06
Three-Sign Balance	4.58*	2.06*	.86*	3.43*	1.35*

* $p < .01$, $df = 1, 900$.

with stated reciprocation than with standard (no stated) reciprocation. Even with standard reciprocation, however, the effect was still significant, $F(1, 900) = 30.49$, $p < .01$.

The Reciprocation × Scales interaction is consistent with our theoretical argument regarding the effect of reciprocation on attraction. One of the arguments advanced to explain the existence of attraction effects relates to the assumption of reciprocated sentiment. Reciprocated positive sentiment produces agreement regarding the worth of the self, whereas reciprocated negative sentiment produces disagreement regarding the worth of the self. Given that the affective scales are most self-related, it is reasonable that they would most clearly reveal an attraction effect and would also most clearly show an alteration in this effect as a function of the reciprocation manipulation.

Involvement × Reciprocation × Scales interaction. The reasoning outlined above implies that the double interaction pattern shown in Table 3 would reveal itself most clearly in self-involved (p-o-x) triads rather than in self-uninvolved (q-o-x) triads. This, in fact, occurred. The reciprocation main effect was strongest with affective scales and self-involved triads. This interpretation of the Involvement × Reciprocation × Scales interaction was supported by a significant triple interaction involving the affective versus cognitive scales contrast, $F(1, 900) = 5.12$, $p < .05$.

Contact

The remaining results all relate to contact. Consistent with past results, there was a significant contact main effect on attraction, agreement, and three-sign balance. Table 4 contains the marginal means. Four contrasts were planned for both attraction and agreement. For attraction, all contrasts were significant. The first contrast indicated that the three conditions in which contact was either implicitly or explicitly positive showed a greater attraction effect than the two conditions in which contact was either null or negative, $F(1, 900) = 77.12$, $p < .01$. The second contrast indicated that the standard condition, in which contact was implicit, showed a smaller attraction effect than the two future-contact conditions, in which contact was explicit, $F(1, 900) = 281.73$, $p < .01$. The third contrast indicated that attraction was greater with than without discussion, $F(1, 900) = 26.82$, $p < .01$, and the fourth contrast indicated that attraction was algebraically greater in the no-contact condition than in the breaking-contact condition, $F(1, 900) = 130.51$, $p < .01$.

The asterisks in Table 4 indicate the significance for grand mean tests (tests of difference from zero) for each of the cell means. As Table 4 indicates, all of the cell means for attraction were significantly different from zero. Except for the significant mean in the no-contact condition, these results are consistent with the contact interpretation of attraction.

The same planned contrasts that were used for attraction were used for agreement. The first contrast indicated that the three conditions in which contact was either implicitly or explicitly positive showed more agreement than the two conditions in which contact was either null or negative, $F(1, 900)$

Table 5
Contact × Rating Scales Attraction Means for Experiment 1

Contact	Pleasantness	Harmony	Expectancy	Consistency	Stability
Standard (no mention)	7.18**	5.50**	.30	1.70	−.40
Future contact with discussion	7.85**	8.95**	10.20**	6.63**	3.70**
Future contact without discussion	8.00**	7.60**	.30	1.50	3.70**
No contact	7.70**	7.30**	.63	.93	−.60
Breaking contact	−2.10*	.22	−8.93**	−5.55**	−3.05**

* $p < .05$, $df = 1, 900$. ** $p < .01$, $df = 1, 900$.

$= 122.50$, $p < .01$. This contrast was significant in spite of the unexpectedly low mean for future contact without discussion. Furthermore, this unexpectedly low mean appears responsible for the nonsignificance of the second contrast comparing the standard and two future-contact conditions. The third and fourth contrasts revealed the expected results. There was a greater agreement effect with than without discussion, $F(1, 900) = 115.03$, $p < .01$, and agreement was algebraically greater with no contact than with breaking contact, $F(1, 900) = 78.51$, $p < .01$. The grand mean tests indicated that all of the cell means, except the ones for breaking contact and future contact without discussion, were significantly different from zero.

The only planned contrast for three-sign balance was a comparison of the standard condition with the four remaining conditions. Consistent with past results, this contrast was significant, $F(1,900) = 16.1$, $p < .01$. The effect was larger in the standard condition.

As Table 1 indicates, the large contact main effect was qualified by two descriptively smaller interactions: the Reciprocation × Contact interaction for agreement only and the Contact × Rating Scales interaction for attraction and three-sign balance.

The Reciprocation × Contact interaction indicates that the contact predictions for agreement were more strongly supported in the standard condition than in the reciprocation. Theoretically, reciprocation relates most directly to attraction, not agreement. This unexpected interaction may have resulted from an attention shift away from agreement and the bearing of contact on agreement when there was explicit reciprocation.

The potentially most important interaction, Contact × Rating Scales, was not significant for agreement, and the effects for attraction and three-sign balance were not replicated in Experiment 2. The attraction means are contained in Table 5.[2] One of the more interesting aspects of this interaction relates to the means in the no-contact condition. Note that the pleasantness and harmony means in the no-contact condition both differed significantly from zero—contrary to the contact interpretation. Furthermore, they were significantly greater than the means for the cognitive scales, $F(1, 900) = 6.20$, $p < .01$. Such a result makes sense, if it is recalled that it is the affective scales that make the self-concept most salient and thus provide for the production of a positive attraction effect through reciprocated sentiment. The effect was not predicted, but it is consistent with the initially stated theoretical orientation.

Discussion

Experiment 1 had two general purposes, one descriptive and one theoretical. The descriptive purpose was to look simultaneously at the differential effects of five scales (pleasantness, harmony, consistency, expectancy, and stability). Although there is no previous literature for the harmony scale (Heider's term for balance), the existing studies were interpreted as indicating that affective scales (pleasantness and harmony)

[2] An analysis of this interaction by a set of orthogonal contrasts is available from the authors.

would produce larger attraction and agreement effects than cognitive scales (consistency, expectancy, and stability). This, in fact, occurred. Such results may suggest that only the cognitive scales are accurate measures of balance. It should be recognized, however, that all five scales produced significant three-sign balance effects, that the consistency scale produced significant attraction and agreement effects, and that the stability scale produced a significant agreement effect (see Table 2). Only the expectancy scale exhibited a total absence of attraction and agreement effects. On the other hand, the means in the final column in Table 2 descriptively indicate that the expectancy scale produced the smallest three-sign balance effect.

What is the meaning of the term *balance?* Heider (1958) described balanced relations as "harmonious" (p. 180). Clearly he was expressing himself metaphorically. Nonetheless, the metaphor does have an interesting double meaning. For the layman, harmony means affectively pleasing, but for the musician, harmony has more of a cognitive, rational meaning. To the extent that balance has both rational and affective connotations, the term *harmony* may be an appropriate metaphor.

The theoretical purpose of Experiment 1 was to provide support for the reciprocated sentiment interpretation of the attraction effect (the greater pleasantness, consistency, and so forth of the *p*-likes-*o* than the *p*-dislikes-*o* relations). When McGarvey (1974) interviewed a small sample of subjects after they had rated the eight triads, he found that some of them did in fact report that they had considered what the other person might think of them (i.e., considered the nonspecified *o*-to-*p* relation). Furthermore, many of these subjects assumed same-sign reciprocation of sentiment (i.e., *o* likes *p* when *p*-likes-*o* was specified, and *o* dislikes *p* when *p*-dislikes-*o* was specified). McGarvey went on to show that a direct manipulation of reciprocated sentiment increased the magnitude of the attraction effect (on pleasantness ratings).

From a balance theory perspective, the problem is to understand why reciprocated same-sign sentiment should increase the magnitude of the attraction effect. Insofar as attention is directed solely toward the *p*-*o*, *o*-*p* cycle or dyad, both reciprocated liking and reciprocated disliking are balanced. Thus, for this one cycle, there is no way for balance theory to predict McGarvey's results. However, Heider assumed that *p* likes *p*: that the typical individual holds himself/herself in high regard. If the second *p* is regarded as a self-concept with which the first *p* has a positive unit relation, it is balanced for *p* likes *p* (Wiest, 1965). Thus, it follows that reciprocated positive sentiment produces agreement regarding the worth of the self and that reciprocated negative sentiment produces disagreement regarding the worth of the self. This means, of course, that from the perspective of reciprocated sentiment, the attraction effect is an agreement effect. For balance theory the problem is one of whether agreement effects can be reduced to cycles that are consistent with the multiplicative rule. These matters will be discussed in greater detail in the context of Experiment 2, which was more directly concerned with the agreement effect.

Whatever is the correct account of the agreement effect, the discussion above does make obvious that reciprocated sentiment relates directly to self-evaluation. For that reason, the finding in the current study that the main effect of reciprocation on attraction only occurred with more affective scales (pleasantness and harmony) is hardly surprising. The means for this Reciprocation × Scales interaction are contained in Table 3. The occurrence of this interaction provides assurance that the effect of reciprocated sentiment on attraction is indeed mediated by self-relevancy considerations. (This assertion rests on the assumption that pleasure and pain are more obviously self-related than are consistency, stability, and expectancy.) The self-relevancy perspective also provides a partial explanation for the finding that the affective scales produced a larger attraction effect than did the cognitive scales. (Note that this scales effect was present even in the standard condition, in which there was no explicit statement of reciprocation.) To the extent that some subjects assumed reciprocated sentiment with its implications for self-evaluation, it

makes sense that affective scales would be more sensitive to attraction.

One potential problem with this line of argument is the finding of LaTour et al. (Note 1) that the attraction effect in a high-involvement (*p-o-x*) triad was not greater than in a low-involvement (*q-o-x*) triad. The present study, in fact, found that the attraction effect was larger with *q-o-x* than *p-o-x* triads. This result is exactly opposite to what one would expect on the basis of a simple self-relevancy perspective on attraction effects.

It is possible that subjects did not realize that *q* was someone other than themselves. We were, however, sensitive to the problem, and made the instructions as explicit as possible. Subjects were twice instructed to rate the situations according to how "*you as the outside observer*" would find the situation. We initially theorized that subjects judged *q-o-x* triads through partial, vicarious identification. To the extent that such identification does occur, the comparison of *p-o-x* and *q-o-x* triads is less obviously a manipulation of self-involvement. It was further theorized that vicarious identification would not be sufficiently complete to detect all of the subtleties in *p-o-x* triads, and thus the interaction of reciprocation and affective versus cognitive scales should be more apparent with *p-o-x* than with *q-o-x* triads. This triple interaction occurred. The means in Table 3 indicate that the tendency for reciprocation to increase the magnitude of the attraction effect was most apparent with affective scales and *p-o-x* triads. The occurrence of this triple interaction is the heart of the present study and strongly implies that some of the variance in attraction effects is mediated by self-relevancy considerations.

A further finding that in retrospect appears consistent with the argument detailed above is that the three-sign balance effect was larger with *p-o-x* than with *q-o-x* triads. The attraction effect was larger with *q-o-x* triads, but the three-sign balance effect was larger with *p-o-x* triads. The vicarious identification tendency does rather well in reproducing the simpler effects, but less well in reproducing the more complex effects. (Attraction is a main effect and three-sign balance a triple interaction.)

The remaining results all related to contact. Past studies of these matters had used only pleasantness and expectancy scales (LaTour et al., Note 1) or only pleasantness scales (Aderman, 1969; Insko et al., 1974; McGarvey, 1974). The present results add generality to the contact-cycle analysis of Insko et al. Despite the fact that half of the cells involved an explicit statement of reciprocation that appears to have partially distracted attention from the contact cycles, the results by and large were consistent with past results. As the means in Table 4 indicate, attraction was maximized by the explicit statement of future contact and was reversed, or negative, in the breaking-contact condition. The major finding that was *not* consistent with the contact-cycle analysis was the fact that the no-contact mean of 3.19 (see Table 4) was significantly greater than zero. According to the contact-cycle analysis, this mean should be within error variance of zero. This theoretical analysis relates most directly to *p-o-x* triads in which there is no stated reciprocation. However, even within this condition, the mean of 2.42 still differed from zero, $F(1, 900) = 5.25$, $p < .01$.

A further possibility is that some of the subjects in the standard-reciprocation–no-contact condition assumed reciprocation that produced an attraction effect. Some circumstantial evidence for this possibility comes from the finding that the affective scales (pleasantness and harmony) were responsible for the attraction effect under no contact. This pattern can be clearly seen along the no-contact row of Table 5.

Experiment 2

The general purpose of the present line of research was to examine the extent to which balance theory is capable of providing interpretations of attraction and agreement effects in the perception of hypothetical situations. Experiment 1 provided support for the hypothesis that contact and cycles containing the contact unit relation can account for at least some of the variance. Previous research (Aderman, 1969; Insko et al., 1974; McGarvey, 1974; LaTour et al., Note 1) has consistently demonstrated the importance of contact when

subjects respond on pleasantness scales. Experiment 1 found that although there were differences among various scales (see Table 5), the effects of contact were evident for all scales. Experiment 1 went further and produced evidence that reciprocated sentiment and self-relevancy considerations were also important in explaining attraction effects. The theoretical account provides a possible explanation for the fact that affective scales revealed a larger attraction effect than did more cognitive scales.

The major remaining problem relates not to attraction, but to agreement. The above-cited studies, and also Experiment 1, all found that the agreement effect was altered in theoretically predictable ways by the manipulation of contact. In general, however, the contact predictions for the agreement effect have been confirmed less frequently than has been the case for the attraction effect. Experiment 1 is a case in point. Note that the contact prediction of a reversed (negative) attraction effect with the breaking-contact condition was supported, but the analogous prediction of a reversed agreement effect with breaking contact was not supported. The agreement mean in this condition was −.06 (see Table 4). Even within the more appropriate standard-reciprocation condition, the breaking-contact mean was −.73— still not significantly less than zero. Another way of making this point is to note that the contact F for agreement (37.68) was descriptively smaller than the contact F for attraction (80.39).

It is clear that contact does affect agreement. It is, of course, intuitively compelling that disagreement with an *o* with whom there is direct contact has more impact than disagreement with an *o* with whom there is no direct contact. Milgram's (1965) "immediacy"-of-the-victim finding is a possible illustration of an analogous effect. On the other hand, it is also intuitively compelling that disagreement has an impact even when there is no contact with *o*. Why is this? Experiment 2 was designed to explore one possible explanation.

The best known theory of social influence is, of course, social comparison theory (Festinger, 1954). Social comparison posits a self-evaluative drive that results in a motivational force to validate opinions. Validation of opinions (or more specifically, nonverifiable opinion) is sought through agreement with others. Thus, social comparison theory provides a possible explanation of the agreement effect.

Acceptance of the social comparison theory explanation of agreement effects would involve an admission that balance theory could not account for the no-contact variance in the agreement effect. This would not mean that balance theory was incorrect; but it would mean that balance theory was of more limited generality. Balance theory explanations proceed by identifying some implicit aspect of *p*'s phenomenology that can be reduced to a balanced cycle. The theory itself does not specify the important aspect of the phenomenology, and the subject himself/herself may not be explicitly aware of the process that is causally important (cf. Nisbett & Wilson, 1977). This matter is illustrated by previously obtained results regarding contact. When McGarvey interviewed a sample of 18 subjects after they had rated the standard eight triads for pleasantness, he found that 13 of the subjects reported thinking about an *o* with whom there was interaction. Heider's (1958) statement of balance theory does not indicate that subjects will so concretize the situation, and it is rather doubtful that subjects recognized the importance of assumed contact in the production of attraction and agreement effects (although McGarvey did not specifically ask about this).

The current challenge is to identify some implicit aspect of *p*'s phenomenology. Examination of McGarvey's interview results indicates that in 8 out of 18 instances, the subjects imagined *o* to be a specific acquaintance, usually a friend or a peer. It is interesting that these subjects conceptualized *o* as someone who was *similar*—similar in cultural experience and in reasoning capacity. They did not conceptualize *o* as a "man from Mars," a member of another culture or subculture, a resident of a mental institution, or a child. From a balance theory perspective, similarity is a unit relation. Therefore, even under conditions of no contact, there would be an agreement cycle—*p* agrees with a

similar *o* or *p* disagrees with a similar *o*—and also an attraction cycle—*p* likes a similar *o* or *p* dislikes a dissimilar *o*. The analysis parallels Insko et al.'s (1974, p. 56) cycles interpretation of contact, except that in this instance the unit relation involves similarity and not contact.

It is intriguing that social comparison theory appears to take an analogous position regarding similarity. According to Festinger (1954), "a person who believes that Negroes are the intellectual equals of whites does not evaluate his opinions by comparison with the opinion of a person who belongs to some very anti-Negro group" (p. 21). Why does social comparison theory make this assumption? Social comparison theory itself does not answer this question. Festinger appears to have intuited a balance implication without being explicit about it.

The manipulation of similarity in Experiment 2 had three levels: a standard condition (with no mention of similarity), a similar condition ("O generally has interests and attitudes similar to yours"), and a dissimilar condition ("O generally has interests and attitudes dissimilar to yours"). In addition, there were manipulations of scales (pleasantness, harmony, consistency, expectancy, and stability) and contact (standard, future contact, no contact, and breaking contact). Since the theoretical import of future contact with and without discussion of *x* is not clear, reference to discussion of *x* was simply omitted. Altogether the design involved 60 between-subjects cells ($3 \times 5 \times 4$).

The scales factor was included to gain some information regarding the generality of the similarity manipulation, and the contact factor was included to examine the extent to which similarity produced effects beyond those resulting from positive, absent, or negative contact. The factorial combinations produced by the present design present a theoretical dilemma regarding the effects of contact across the levels of similarity. It is possible that the effects of contact may summate (or average) with those of similarity, or that the effects of contact will be more apparent with standard similarity (when there is no explicit mention of similarity). The latter possibility could be the result of attention shifts and/or of information overload.

Previous research has obtained results that appear to be the result of attention shifts. First, Experiment 1, in agreement with previous research (Aderman, 1969; Insko et al., 1974; McGarvey, 1974; LaTour et al., Note 1), found that the explicit mention of contact reduced the magnitude of the three-sign balance effect from its magnitude in the standard condition (with no mention of contact). Second, Experiment 1 found that the explicit statement of reciprocation reduced the magnitude of the agreement effect from its magnitude in the standard-reciprocation condition in which there was no mention of reciprocation. Third, manipulations of triad order by Insko et al., and also McGarvey, found that the magnitude of the three-sign balance effect was greater if the first triad was three-sign imbalanced (and thus more atypical and attention getting) rather than three-sign balanced.

Method

Subjects

The subjects were 600 male and female students from the introductory psychology course at the University of North Carolina.

Independent and Dependent Variables

As in Experiment 1, each subject rated the standard eight triads, and these ratings were used to generate seven orthogonal effects. The major analyses, however, concerned the attraction, agreement, and three-sign balance effects.

There were three between-subjects factors: similarity (standard, similar, or dissimilar), contact (standard, future contact, no contact, or breaking contact), and rating scales (pleasantness, expectancy, harmony, stability, or consistency).[3] The basic design involved 60 between-subjects cells and three dependent variables with 10 subjects per cell.

Procedure

The procedure closely paralleled that of Experiment 1. Subjects were tested in groups of approxi-

[3] Through an oversight the negative endpoint of the stability scale was changed from "Most changeable" to "Most unstable," and the negative endpoint of the harmony scale was changed from "Most discordant" to "Most unharmonious."

Table 6
Multivariate Analysis of Variance for Experiment 2

Source	Attraction F	Agreement F	Three-sign balance F	Multivariate F
Similarity (S)	6.65**	21.60**	3.18*	5.15**
Contact (C)	65.21**	9.50**	5.99**	11.05**
Scales (L)	17.93**	24.19**	1.50	6.40**
S × C	3.17*	2.45*	1.67	1.92**
S × L	4.10**	2.20*	1.21	1.96**
C × L	1.24	.52	1.06	1.20
S × C × L	.84	.75	.80	1.29

* $p < .05$. ** $p < .01$.

mately 25 people, and no two subjects in any group were in the same experimental condition. Each subject received a nine-page booklet, the first page of which gave general instructions and the remaining pages of which listed each of the eight triads (one per page), along with a statement of one or more assumptions and a rating scale. The similarity assumptions were stated as follows.

Assume: O generally has interests and attitudes similar to yours.

Or

Assume: O generally has interests and attitudes dissimilar to yours.

As in Experiment 1, no subjects failed to rate all eight triads, and thus no subjects were eliminated from the analysis.

Results

Grand Mean Tests

Grand mean tests of the seven analysis of variance contrasts for the eight triads revealed the usual attraction, $F(1, 540) = 110.88$, $p < .01$; agreement, $F(1, 540) = 149.80$, $p < .01$; and three-sign balance effects, $F(1, 540) = 252.67$, $p < .01$. This analysis also found an unexpected First × Second Sign interaction, $F(1, 540) = 6.70$, $p < .01$. Triads in which the p-o and p-x relation were of the same sign were rated higher than triads in which these two relations were of unlike sign. Subsequent analyses focused on the descriptively larger attraction, agreement, and three-sign balance effects.

Multivariate Analysis of Variance

A three-factor multivariate analysis of variance of the attraction, agreement, and three-sign balance effects is presented in Table 6. This analysis revealed significant multivariate Fs for all three main effects, as well as for two of the interactions. Since both of the interactions involve similarity, we will begin with a description of the main effects for rating scales and contact.

Rating Scales Main Effect

There was a significant rating scales main effect for both attraction and agreement. Table 7 contains the marginal means. A contrast comparison of the affective scales (pleasantness and harmony) with the more cognitive scales (expectancy, stability, and consistency) was significant for both attraction, $F(1, 540) = 60.16$, $p < .01$, and agreement, $F(1, 540) = 71.50$, $p < .01$, indicating a greater effect with the affective scales and replicating a similar finding in Experiment 1.

A second contrast comparing expectancy and consistency scales with the stability scale indicated a greater effect with stability for both attraction, $F(1, 540) = 15.66$, $p < .01$, and agreement, $F(1, 540) = 25.10$, $p < .01$. This contrast, which was significant for agreement but not attraction in Experiment 1, possibly means that stability has a more affective connotation than do expectancy and stability.

Grand mean tests were conducted on each of the cell means in Table 7, which also includes the means for three-sign balance. Except for the failure of consistency to reveal significant attraction and agreement effects and the presence of a significant attraction effect for stability, the results replicated the findings of Experiment 1.

Table 7
Rating Scale Means for Experiment 2

Condition	Pleasantness	Harmony	Expectancy	Consistency	Stability
Attraction	4.34**	5.52**	−.26	.78	3.02*
Agreement	4.21**	4.44**	−.74	1.24	2.97*
Three-sign balance	3.44**	2.39*	2.79*	3.31**	2.40*

* $p < .05$, $df = 1, 540$. ** $p < .01$, $df = 1, 540$.

Contact Main Effect

In agreement with Experiment 1 and all past results, there was a significant contact main effect on attraction, agreement, and three-sign balance. Table 8 contains the marginal means. Three contrasts were planned for attraction and agreement. First, attraction was greater in the standard and future-contact conditions than in the no-contact and breaking-contact conditions, $F(1, 540) = 55.54$, $p < .01$, as was also agreement, $F(1, 540) = 18.57$, $p < .01$. Second, attraction was greater in the future-contact than in the standard condition, $F(1, 540) = 84.23$, $p < .01$, but the contrast was not significant for agreement. Third, attraction was greater with no contact than with breaking contact, $F(1, 540) = 55.86$, $p < .01$, as was also agreement, $F(1, 540) = 7.83$, $p < .01$.

The asterisks in Table 8 indicate the significance of the grand mean tests (tests of differences from zero) for each of the cell means. The failure of the attraction mean in the standard condition to reach significance is inconsistent with the results of Experiment 1 (see Table 4) and is not in agreement with the general assumption that some subjects in a standard condition implicitly assume contact. As is indicated below, however, the standard contact mean was significantly greater than zero within the standard similarity condition (see Table 10).

The only planned contrast for three-sign balance was a comparison of the standard condition with the four remaining conditions. In agreement with Experiment 1, the contrast was significant, $F(1, 560) = 10.24$, $p < .01$.

Similarity Main Effect

There was a significant similarity main effect for attraction, agreement, and three-sign balance; the marginal means for attraction and agreement are in the right-hand column of Table 9. A contrast of the standard condition (with implicit similarity) and similar condition with the dissimilar condition was significant for both attraction, $F(1, 540) = 13.25$, $p < .01$, and agreement, $F(1, 540) = 27.69$, $p < .01$. A second contrast of the standard and similar conditions was significant for agreement only, $F(1, 540) = 15.51$, $p < .01$. According to the similarity interpretation of attraction and agreement, all four of these contrasts should have been significant. The agreement results were more in accord with prediction than were the attraction results (in contrast to the contact results, in which the opposite occurred). Grand mean tests (see

Table 8
Contact Means for Experiment 2

Condition	Standard (no mention of) contact	Future contact	No contact	Breaking contact
Attraction	1.27	7.88**	3.47**	−1.91*
Agreement	2.87*	3.68**	2.35*	.78
Three-sign balance	3.65**	2.23*	3.60**	1.99*

* $p < .05$, $df = 1, 540$. ** $p < .01$, $df = 1, 540$.

Table 9
Similarity × Rating Scales Means

Condition and similarity	Pleasantness	Harmony	Expectancy	Consistency	Stability	Marginal
Attraction						
Standard (no mention)	2.88*	6.08**	.85	2.60*	3.93**	3.27**
Similarity	4.00**	5.60**	−.40	3.05**	4.78**	3.41**
Dissimilarity	6.15**	4.90**	−1.23	−3.33**	.35	1.37
Agreement						
Standard (no mention)	3.53**	4.63**	−.60	1.50	1.98*	2.21*
Similarity	5.00**	5.00**	1.00	4.00**	5.58**	4.12**
Dissimilarity	4.10**	3.70**	−2.63**	−1.78*	1.35	.95

* $p < .05$, $df = 1, 540$. ** $p < .01$, $df = 1, 540$.

Table 9) indicated that significant attraction and agreement effects occurred in all but the dissimilar condition. According to the similarity interpretation, the attraction and agreement effects should have been negative in the dissimilar condition. In reality, of course, contact and reciprocation also have some bearing on these means.

The results for three-sign balance indicated that only the first contrast was significant, $F(1, 540) = 5.66$, $p < .05$. Three-sign balance was greater in the standard and similar conditions than in the dissimilar condition.

Similarity × Rating Scales Interaction

Unfortunately for the sake of both initial prediction and simplicity of discussion, the main effect for similarity mentioned above was qualified by a Similarity × Rating Scales interaction. The attraction and agreement means are contained in Table 9. A series of contrasts was calculated to explore this pattern of results.[4] These contrasts indicated that greater attraction and agreement effects with the standard and similar conditions than with the dissimilar condition were only found for the cognitive scales (expectancy, consistency, and stability). For attraction the effect for pleasantness and harmony was even in the wrong direction, although only slightly. For agreement, the effect was in predicted direction, although not significantly so. These results indicate that the social comparison theory perspective is only supported by the more cognitive scales. Put less technically, the subjects indicated that they regarded agreement with a dissimilar other as inconsistent but not unpleasant.

Similarity × Contact Interaction

There was a significant Similarity × Contact interaction for both attraction and agreement. As the means in Table 10 indicate, the interaction implies that the contact predictions were more strongly supported in the standard-similarity condition. Such a result is in agreement with the anticipation that the factorial combination of similarity and contact might result in attention shifts and/or information overload. It is, furthermore, reassuring to find that the attraction effect with standard contact is significantly greater than zero when there is standard similarity. Across all levels of similarity the standard contact mean was not significantly greater than zero (see Table 8).

Scales Within the No-Contact Condition

One further finding should be mentioned. This finding is relevant to the argument that assumed reciprocation and self-relevancy considerations are important in accounting for attraction variance in the no-contact condition. To the extent that self-relevancy con-

[4] A detailed summary of these analyses is available from the first author.

Table 10
Similarity × Contact Means

Condition and similarity	Standard (no mention of) contact	Future contact	No contact	Breaking contact
Attraction				
Standard (no mention)	2.76*	9.92**	4.12**	−3.74**
Similarity	1.48	7.36**	4.04**	.74
Dissimilarity	−.42	6.36*	2.26*	−2.72*
Agreement				
Standard (no mention)	3.96**	3.32**	2.04*	−.50
Similarity	4.80**	5.56**	3.56**	2.54*
Dissimilarity	−.14	2.16*	1.46	.32

* $p < .05$, $df = 1, 540$. ** $p < .01$, $df = 1, 540$.

siderations are important in the no-contact condition, the affective scales should produce greater attraction than the more cognitive scales. This happened in Experiment 1 (see Table 5). A planned test of the same contrast in Experiment 2 was also significant, $F(1, 540) = 35.60$, $p < .01$. Mean attraction effects for the pleasantness, harmony, expectancy, consistency, and stability scales were 4.47, 8.60, −.03, 1.53, and 2.80, respectively.

Postexperimental Questionnaire

Finally, we should like to mention that Experiment 2 also collected some self-report data—data that will only be briefly described. Since the one variable that appears to be most consistently important in a balance account of attraction and agreement effects is contact, and since McGarvey's informal postexperimental interviews on this matter were only done with 18 subjects, we were interested in collecting data from a somewhat larger sample. The first question, which was asked only in the standard contact condition, inquired whether the subjects thought of a specific *o* (another person) with whom they had contact. The results were even more striking than McGarvey's since 136 of 150 subjects (90.7%) answered in the affirmative. The second question, which was also asked in the nonstandard conditions, inquired about the time duration of the contact. All but one subject in the standard condition specified present contact. In the nonstandard conditions, the question constituted a manipulation check on past and future contact and overwhelmingly yielded the expected results.

The final set of questions was directed at a matter with which we were most concerned. This is the potential criticism that the previously obtained contact results were an artifact of choice. Perhaps when subjects were told to assume no more contact or continued contact in the future they imagined that this change was the result of their own free choice. A set of questions designed to get at this problem revealed that some subjects imagined that their contact was the result of *p*'s choice; some, of *o*'s choice; some, of mutual choices; some, of no-choice circumstances; and some, of a combination of choice and circumstances. Since the modal response category was "circumstances," this category was compared with all of the other choice categories. Except in two instances, this analysis revealed no significant effects. In the future-contact condition, the agreement effect was greater for no-choice than for choice subjects; and in the breaking-contact condition, the attraction effect was more negative for no-choice than choice subjects. In general, the results do not seem to indicate that the usual contact results are an artifact of choice.

Discussion

Experiment 2 generally replicated the Experiment 1 results relating to rating scales and contact. The two experiments agreed in the major finding that the affective scales (pleasantness and harmony) produced larger

attraction and agreement effects than did the cognitive scales (expectancy, consistency, and stability). It is also interesting that in neither Experiment 1 nor Experiment 2 did expectancy ratings reveal significant attraction or agreement effects. The means in Experiment 2 are, in fact, slightly negative. LaTour et al. (Note 1) also did not find that expectancy ratings produced significant attraction or agreement effects. All three experiments found that expectancy ratings did not reveal significant attraction or agreement effects, but did reveal significant three-sign balance effects. These results possibly indicate that when subjects make expectancy judgments, they rely to a great extent on the relative frequencies of past social experiences (i.e., they rely on something analogous to the long-run relative frequency interpretation of probability). From this perspective, subjects do not regard *p*-dislikes-*o* or *p*-disagrees-with-*o* situations as infrequent. On the other hand, three-sign imbalanced situations in which *p* likes someone and disagrees with that person or dislikes someone and agrees with him/her may have a relatively low frequency of past occurrence —possibly because past group relationships tend to be ones in which there is both attraction and agreement (cf. Newcomb, 1978).

The results for contact were generally in accord with the findings of Experiment 1. It is again apparent that more of the contact predictions were supported for attraction than for agreement. In view of the fact that the explicit mention of similarity and/or dissimilarity appeared to lessen the magnitudes of some of the predicted contact effects, this general replication of basic results is gratifying.

The main focus of Experiment 2 was on the anticipated effect of similarity on attraction and agreement. Experiment 2 can be regarded as a test of the social comparison theory perspective on agreement and attraction effects. This perspective is that we seek agreement with and are attracted to similar others. The anticipated main effect of similarity on both attraction and agreement occurred. Both attraction and agreement effects were greater in the standard and similar conditions than in the dissimilar condition. Furthermore, the agreement effect was greater in the similar condition than in the standard condition. Beyond this, however, the attraction and agreement effects were not significantly less than zero in the dissimilar condition, and, most important of all, the main effect of similarity was qualified by the rating scales factor. With the affective scales, the manipulation of similarity had no significant effect on either attraction or agreement (see Table 9). Agreement with similar and dissimilar others was equally pleasant, but not equally consistent. Although this finding is intuitively plausible, it was not anticipated. There does not seem to be anything in social comparison theory that would have indicated such results. Indeed, the basic assumption of a drive for self-evaluation might be interpreted as implying a greater effect on the affective than on the cognitive scales—certainly not a lesser effect.

From the perspective of our initial expectations, Experiment 2 was a qualified success. Some additional agreement variance, beyond that explained by contact, was accounted for, but there was obviously unexplained variance, particularly on the affective scales. The problem becomes acute when it is recognized that it was the affective scales that produced the most agreement variance. Also it is important to note that a balance account of the reciprocated sentiment results for attraction is dependent on the adequacy with which balance theory can account for agreement variance. At the present time it is apparent that the balance theory is better able to account for variance on the cognitive scales than for variance on the affective scales.

We regard the results of the two experiments as sufficiently encouraging to warrant continued investigation of the possibility that balance theory can account for the reliable variance in the phenomenology of hypothetical social situations. As previously indicated, the rationale for studying hypothetical social situations is not necessarily to learn about actual social situations. Although the phenomenology of actual social situations may bear some resemblance to the phenomenology of hypothetical social situations (cf. Sampson & Insko, 1964, for a study of nonhypothetical

p-o-x triads), there are some important differences. At the very least the phenomenology of actual situations will be more complex and span a greater period of time. Hypothetical social situations are studied because the ease of manipulating variables makes these situations a convenient arena for testing specific balance hypotheses. Balance theory should be able to account for certain aspects of human thought, whatever their basis in reality.

Reference Note

1. LaTour, et al. *Pleasantness and expectancy ratings in* p-o-x *and* q-o-x *triads: The partially overlapping predictions of psychological hedonism and balance theory.* Unpublished manuscript, University of North Carolina, 1975.

References

Aderman, D. Effects of anticipating future interaction on the preference for balanced states. *Journal of Personality and Social Psychology,* 1969, *11*, 214–219.

Crockett, W. H. Balance, agreement, and subjective evaluations of the *P-O-X* triads. *Journal of Personality and Social Psychology,* 1974, *29*, 102–110.

Festinger, L. A theory of social comparison processes. *Human Relations,* 1954, 7, 117–140.

Gutman, G. M., & Knox, R. E. Balance, agreement, and attraction in pleasantness, tension, and consistency ratings of hypothetical social situations. *Journal of Personality and Social Psychology,* 1972, *24*, 351–357.

Gutman, G. M., Knox, R. E., & Storm, T. F. Developmental study of balance, agreement, and attraction effects in the ratings of hypothetical social situations. *Journal of Personality and Social Psychology,* 1974, *29*, 201–211.

Heider, F. Attitudes and cognitive organization. *Journal of Psychology,* 1946, *21*, 107–112.

Heider, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

Insko, C. A., Songer, E., & McGarvey, W. Balance, positivity, and agreement in the Jordan paradigm: A defense of balance theory. *Journal of Experimental Social Psychology,* 1974, *10*, 53–83.

McGarvey, W. E. *Beyond three-sign balance: Balance theory predictions for more than three bands in the Jordan paradigm.* Unpublished doctoral dissertation, University of North Carolina, 1974.

Milgram, S. Some conditions of obedience and disobedience to authority. *Human Relations,* 1965, *18*, 57–76.

Miller, C. E., & Norman, R. M. G. Balance, agreement, and attraction in hypothetical social situations. *Journal of Experimental Social Psychology,* 1976, *12*, 109–119.

Newcomb, T. M. The acquaintance process: Looking mainly backward. *Journal of Personality and Social Psychology,* 1978, *36*, 1075–1076.

Nisbett, R. E., & Wilson, T. D. Telling more than we can know: Verbal reports on mental processes. *Psychological Review,* 1977, *84*, 231–259.

Sampson, E. E., & Insko, C. A. Cognitive consistency and performance in the autokinetic situation. *Journal of Abnormal and Social Psychology,* 1964, *68*, 184–193.

Wiest, W. M. A quantitative extension of Heider's theory of cognitive balance applied to interpersonal perception and self-esteem. *Psychological Monographs,* 1965, *79*(14, Whole No. 607).

Zajonc, R. B. Cognitive theories in social psychology. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (Vol. 1). Reading, Mass.: Addison-Wesley, 1968.

Received June 19, 1978

Editors Named for New *Journal of Personality and Social Psychology*

The Publications and Communications Board has named Melvin Manis, Ivan D. Steiner, and Robert Hogan to edit the *Journal of Personality and Social Psychology* (JPSP), which will appear in a new format in January 1980. The affiliations of the new editors are, respectively, University of Michigan and Veteran's Hospital, Ann Arbor; University of Massachusetts; and Johns Hopkins University. Manis will edit the section on ATTITUDES AND SOCIAL COGNITION, Steiner the section on INTERPERSONAL RELATIONS AND GROUP PROCESSES, and Hogan the section on PERSONALITY PROCESSES AND INDIVIDUAL DIFFERENCES.

An announcement on the restructuring of JPSP appeared on p. 668 of the June 1978 issue of JPSP. As indicated at that time, the three editors will work independently with separate groups of advisory editors. Their acceptances will appear as distinct sections in monthly issues of JPSP. Detailed descriptions of each section's coverage appear on pages 468–469 of the March issue of JPSP.

Authors should submit manuscripts to the appropriate section editor as follows:

ATTITUDES AND SOCIAL COGNITION

April 1 through June 15:

Melvin Manis
c/o Anita DeVivo
Suite 700
1400 North Uhle Street
Arlington, Virginia 22201

After June 15:

Melvin Manis
Research Center for Group Dynamics
Institute for Social Research
University of Michigan
Ann Arbor, Michigan 48104

INTERPERSONAL RELATIONS AND GROUP PROCESSES

Ivan D. Steiner
Department of Psychology
University of Massachusetts
Amherst, Massachusetts 01003

PERSONALITY PROCESSES AND INDIVIDUAL DIFFERENCES

Robert Hogan
Johns Hopkins University
Department of Psychology
33rd and Charles Streets
Baltimore, Maryland 21218

Ad Hoc Reviewers

The Editor acknowledges with thanks the advice and consultation provided by the following persons, who served recently as ad hoc reviewers of manuscripts. (The number of manuscripts reviewed is in parentheses, if more than one.)

David F. Aderman
Teresa Amabile
Albert Bandura
Andrew Baum
Gordon Bear
Paul A. Bell
Michael Berbaum
Ellen Berscheid
Jack Block (2)
Jeff Bryson
Nancy Cantor
Jerry L. Cohen
William Crano
Robyn Dawes (2)
Edward F. Diener
Anthony Doob
Stanley Feldstein (2)
Ira J. Firestone
R. C. Gardner
Lewis Goldberg
William Graziano
Steve Harkins
John Harvey
Robert Hogan
Ted L. Huston
Alice Isen
Martin Kaplan
H. H. Kelley
Ellen Langer
Richard Lempert
Mark Lepper
Howard Leventhal
David McMillen
Dale Miller
Norman Miller
Walter Mischel (2)
Bert S. Moore
Gregory Moschetti
Stanley Mulaik
Richard E. Nisbett
Kent Norman
Thomas Ostrom
Richard Page
Sidney Perloe
Harry Reis (2)
Marc Riess
T. B. Rogers
Lee Ross
Michael Ross
Myron Rothbart
Diane Ruble
Stanley Schachter
Jerome Singer
Paul Slovic
Edward Smith
Mark Snyder (2)
Arthur Staats
Toby Tetenbaum
Ewart Thomas
Richard Tsujimoto
Stephen Worchel
Paul H. Wright
Mark Zanna
Miron Zuckerman

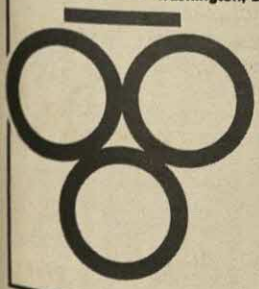
APA EMPLOYMENT BULLETIN

Possibly your key to greater career opportunities.

Each month 125 to 150 position openings are listed by geographical area and employment field for psychologists with a minimum of a Master's degree. Members may also publish availability notices. Coded identification numbers assure confidentiality. Ideal for employers, students, and Master's and Ph.D graduates.

A one year subscription is as little as $9.00 for members, $16.00 for nonmembers. If you're in a hurry, First Class delivery is available for an additional $4.00.

For more information, contact:
The Editor,
Employment Bulletin,
or send your order to:
American Psychological Association,
Subscription Dept.-Employment Bulletin
1200 Seventeenth St., N.W.
Washington, D.C. 20036



POSITIVE SOCIAL BEHAVIOR AND MORALITY VOLUME 1
SOCIAL AND PERSONAL INFLUENCES

By ERVIN STAUB

What leads human beings to helping, generosity, cooperation, sensitivity to others, and the reciprocation of others' kindness? What values and beliefs guide them to help others in need, to extend effort to increase others' well-being, to try to do what is "right" in their conduct toward others? When do they experience sympathy, empathy, and other emotions that tie them to other people and lead to altruistic acts? What influences positive behavior toward strangers, friends and intimates, members of ingroups and outgroups? This book concerns itself with these questions. It examines and integrates the extensive research and lively theorizing that exists about how personal characteristics and the nature of specific circumstances and of the surrounding social world affect positive conduct. It provides a new theoretical conception and suggests new areas for future research.

1978, 490 pp., $19.50 ISBN: 0-12-663101-8

VOLUME 2
SOCIALIZATION AND DEVELOPMENT

By ERVIN STAUB

It is of great importance for the welfare of individual human beings and for the harmonious functioning of society that people engage in conduct that benefits others—that they help others in need and attempt to contribute to others' welfare. The tendency to behave positively, and the personal characteristics that make positive behavior more (or less) likely develop through socialization, through the influence of peers, and through the child's varied experiences are investigated in this book. How do these influences, together with the nature of the social world, affect children's awareness of and concern with "right" and "wrong"? How are children's values and beliefs, their sense of connectedness to other human beings, their capacity for empathy, sympathy, and identification with others affected? By what principles of learning and development are the varied influences incorporated into the child's personality and behavior?

This book explores and integrates the extensive research and theorizing in these domains, identifies previously unrecognized or unexplored influences on children's positive tendencies, and points to future directions for the field. It suggests to all those concerned with the optimal social development of children, ways of promoting a positive orientation towards other human beings and positive social behavior.

Volume I and II of *Positive Social Behavior and Morality* are related and enlarge upon each other. The material in them is so organized that each can be meaningfully read independently, and each represents an independent whole.

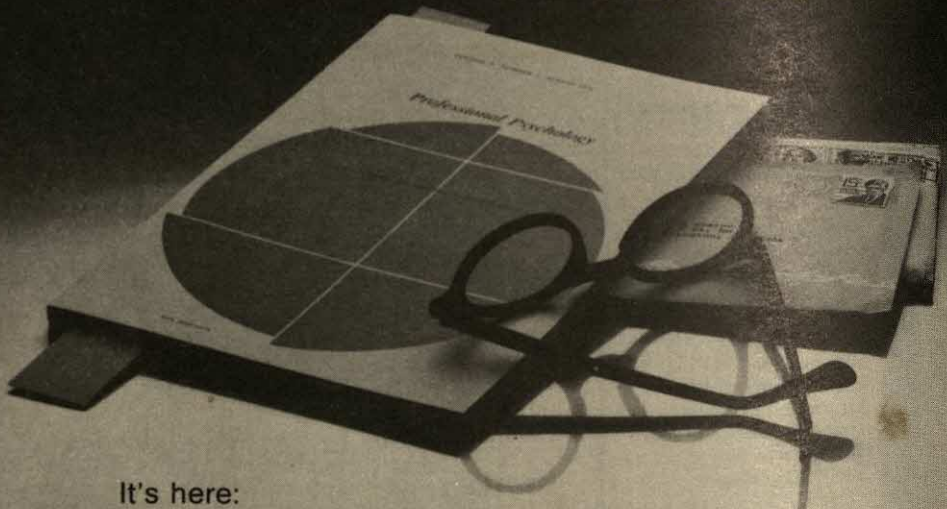
1979, 328 pp., $14.50 ISBN: 0-12-663102-6

Send payment with order and save postage and handling charge.
Prices are subject to change without notice.

ACADEMIC PRESS, INC.

A Subsidiary of
Harcourt Brace Jovanovich, Publishers
111 FIFTH AVENUE, NEW YORK, N.Y. 10003
24-28 OVAL ROAD, LONDON NW1 7DX

NOW: INFORMATION ABOUT THE LAW AND PSYCHOLOGY.



It's here:
Law and Professional Psychology. This special August 1978 issue of **Professional Psychology** brings you thirteen original articles full of fresh thought and constructive recommendations on the issues that vitally affect your day-to-day practice. You'll find precautions you may take to protect yourself as a professional and a plan for improving investigations of malpractice claims. Expert witness testimony, jury selection, civil commitment, confidentiality and privilege, minor's consent to treatment, and the use of psychological devices are critically examined.

Send your order today. And soon you'll receive this special issue.

Single copies of the issue are available for $5 each. All orders $25 or less must be prepaid. Make and send checks payable to:

American Psychological Association
Order Department
1200 17th Street, N.W.
Washington, D.C. 20036

Enclosed is $_____ for ___ copies of **Professional Psychology's** special issue—Law and Professional Psychology—at $5.00 each.

NAME _____

ADDRESS _____

CITY _____

STATE _____ ZIP CODE _____

PP-8

(Continued from front cover)

The Role of Assumed Reciprocation of Sentiment and Assumed Similarity in the Production of Attraction and Agreement Effects in *p*-*o*-*x* Triads 790
Chester A. Insko and Akanbi Adewole

Ad Hoc Reviewers 810

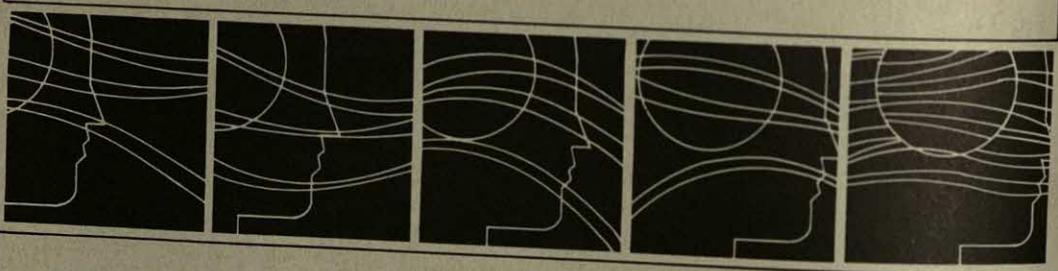
Corrections to Koretzky, Kohn, and Jeger 727

Editors Named for New *Journal of Personality and Social Psychology* 809

79

Willo P. White, editor

Resources in Environment and Behavior



NEW FROM APA IN 1979

An invaluable sourcebook for students, instructors, and researchers in the new field of environment and behavior. This solid reference includes:

- Overview and history of this emerging field
- Graduate programs—both formal and informal
- Teaching innovations introduced in the United States, Canada, and Great Britain
- Funding sources
- Career opportunities
- Directory of key individuals currently working in the field
- Annotated bibliography
- Listing of relevant journals

Resources in Environment and Behavior is available from the American Psychological Association for $10 (soft cover only). To order, make your check payable to APA.

American Psychological Association
Order Department
1200 17th Street, NW
Washington, DC 20036

Please include full remittance for orders of $25 or less.

JOURNAL OF

Personality AND Social Psychology

VOLUME 37 NUMBER 6 · JUNE 1979

Special Section on Interpretive Controversies in Personality and Social Psychology
See Contents on Inside Back Cover

Feeling More Than We Can Know: Exposure Effects Without Learning 811
William Raft Wilson

Repeated presentation of stimuli in an unattended channel produced increased liking, despite only chance recognition.

Many Hands Make Light the Work: The Causes and Consequences of Social Loafing 822
Bibb Latané, Kipling Williams, and Stephen Harkins

People exert themselves less when working collectively than when working alone—and this has negative implications for social institutions.

Template Matching: A Proposal for Probing the Ecological Validity of Experimental Settings in Social Psychology 833
Daryl J. Bem and Charles G. Lord

The technique is to match nonlaboratory Q-sort descriptions of subjects to Q sorts judged prototypical for their laboratory-task behavior.

Factors Affecting the Evaluation of Improvement: The Role of Normative Standards and Allocator Resources 847
André deCarufel

Cardiovascular Changes During Social Competition in a Mixed-Motive Game 858
Lawrence F. Van Egeren

Vocal Interruptions in Dyadic Communication as a Function of Speech and Social Anxiety 865
Michael Natale, Elliot Entin, and Joseph Jaffe

Multidimensional Input in Equity Theory 879
Arthur J. Farkas and Norman H. Anderson

Race, Sex, and the Expression of Self-Fulfilling Prophecies in a Laboratory Teaching Situation 897
Marylee C. Taylor

Attitude Organization and the Attitude–Behavior Relationship 913
Richard P. Bagozzi and Robert E. Burnkrant

(Continued on inside back cover)

Published monthly by the
AMERICAN PSYCHOLOGICAL ASSOCIATION
ISSN 0022-3514

Editor

Anthony G. Greenwald, *Ohio State University*

Topic Editors

Irwin Altman, *University of Utah*
Robert A. Baron, *Purdue University*
Daryl Bem, *Stanford University*
Eugene Burnstein, *University of Michigan*
Robert B. Cialdini, *Arizona State University*
Reid Hastie, *Harvard University*
Douglas N. Jackson, *University of Western Ontario*
Leslie Z. McArthur, *Brandeis University*

Consulting Editors

Icek Ajzen, *University of Massachusetts—Amherst*
Robert S. Baron, *University of Iowa*
Gordon Bear, *Ramapo College of New Jersey*
Ellen Berscheid, *University of Minnesota*
Michael Birnbaum, *University of Illinois at Urbana-Champaign*
Philip Brickman, *Institute for Social Research*
J. Merrill Carlsmith, *Stanford University*
John S. Carroll, *Loyola University of Chicago*
Jerry L. Cohen, *University of Illinois at Urbana-Champaign*
Sheldon Cohen, *University of Oregon*
Rick Crandall, *Texas Christian University*
William D. Crano, *Michigan State University*
Robyn M. Dawes, *University of Oregon*
Kay K. Deaux, *Purdue University*
Marshall Dermer, *University of Wisconsin—Milwaukee*
Edward Diener, *University of Illinois at Urbana-Champaign*
Anthony Doob, *University of Toronto, Toronto, Canada*
Susan V. Eisen, *McLean Hospital, Belmont, Massachusetts*
Baruch Fischhoff, *Decision Research, Eugene, Oregon*
Irene H. Frieze, *University of Pittsburgh*
David Funder, *Stanford University*
David L. Hamilton, *University of California, Santa Barbara*
Ranald D. Hansen, *Oakland University*
Richard J. Harris, *University of New Mexico*
Edward Helmes, *University of Western Ontario, London, Canada*
Donald S. Hiroto, *VA Hospital, Brentwood, Los Angeles, CA*
Edward E. Jones, *Princeton University*
Martin F. Kaplan, *Northern Illinois University*
Douglas T. Kenrick, *Montana State University*
Norbert L. Kerr, *University of California, San Diego*
Samuel Komorita, *University of Illinois at Urbana-Champaign*
Vladimir J. Konečni, *University of California, San Diego*
Robert M. Krauss, *Columbia University*
Ellen J. Langer, *Harvard University*
Mark R. Lepper, *Stanford University*
Melvin Manis, *University of Michigan (Ann Arbor)*
Hazel Markus, *University of Michigan (Ann Arbor)*
Margaret Matlin, *SUNY College at Geneseo*
Dale T. Miller, *University of British Columbia, Vancouver*
Thomas M. Ostrom, *Ohio State University*
Paul B. Paulus, *University of Texas at Arlington*
Richard Petty, *University of Missouri—Columbia*
Harry T. Reis, *University of Rochester*
Judith Rodin, *Yale University*
Seymour Rosenberg, *Livingston College, Rutgers*
Ralph L. Rosnow, *Temple University*
Lee Ross, *Stanford University*
Michael Ross, *University of Waterloo, Waterloo, Canada*
Diane Ruble, *Princeton University*
Michael J. Saks, *Nat'l Center for State Courts, Williamsburg, VA*
Michael F. Scheier, *Carnegie-Mellon University*
Barry R. Schlenker, *University of Florida*
Philip Shaver, *New York University*
Steven J. Sherman, *Indiana University*
Jerome E. Singer, *University of the Health Services, Bethesda, Maryland*
Eliot R. Smith, *University of California, Riverside*
Mark Snyder, *University of Minnesota*
Melvin L. Snyder, *University of Texas at Austin*
Lawrence Stricker, *Educational Testing Service*
Shelley E. Taylor, *Harvard University*
Abraham Tesser, *University of Georgia*
Bernard Weiner, *University of California, Los Angeles*
Gary L. Wells, *University of Alberta, Edmonton, Canada*
Stephen Worchel, *University of Virginia*
Robert S. Wyer, *University of Illinois at Urbana-Champaign*
Mark Zanna, *University of Waterloo, Waterloo, Canada*
Miron Zuckerman, *University of Rochester*

Assistant to the Editor
Holly Wong

APA Journal Staff

Anita DeVivo, *Executive Editor*
Ann I. Mahoney, *Manager, Journal Production*
Barbara R. Richman, *Production Supervisor*
Marjorie M. Henry, *Production Editor*
Robert J. Hayward, *Advertising Representative*
Juanita Brodie, *Subscription Manager*

The Journal of Personality and Social Psychology publishes original reports of research, methodology, theory, criticism, and literature review in the fields of personality and social psychology of normal processes and populations. Submitted manuscripts should delineate an important question and provide a clear answer to it. In the case of research reports, this will often require demonstration that the reported results are generalizable beyond a single research setting. Consult the editorial in the January 1976 issue (Vol. 33, pp. 1–7) for a detailed statement of policies and instructions to authors.

Manuscripts: Submit manuscripts in quadruplicate, prepared according to the *Publication Manual of the American Psychological Association* (2nd ed.). Copies should be highly legible and will not be returned to authors unless they are original typewritten copies. Mail manuscripts according to instructions on p. 809 of the May issue.

Required supplementary information: Manuscripts cannot be reviewed for publication unless the author provides supplementary information in the following categories: (a) a cover letter that indicates the relation (or the lack of relation) of the submitted manuscript to other papers by the same author(s) that are currently submitted, in press, or recently published; (b) one copy of each such related paper; (c) one copy of a statement of the procedures of the research reported in the submitted manuscript in sufficient detail to serve as a guide for conduct of an exact replication; and (d) one copy of the summary tables (e.g., correlation tables, analysis of variance summaries) for the *major* analyses reported. If it appears that the supplementary materials in categories (c) or (d) might be bulky, the author should exercise judgment and restaint in selecting what to send. It is not necessary to send supplementary material that repeats verbatim manuscript content.

After a manuscript is accepted, authors will be expected to provide assurance that their data are available in unanalyzed form for reanalyses by other investigators and that their research procedures were consistent with the American Psychological Association's published principles of research ethics, except insofar as may be detailed otherwise in the manuscript.

Copyright and Permission: All rights reserved. Written permission must be obtained from the American Psychological Association for copying or reprinting text of more than 500 words, tables, or figures. Permission is normally granted contingent upon like permission of the author, inclusion of the APA copyright notice on the first page of reproduced material, and payment of a fee of $10 per page, table, or figure. Abstracting is permitted with credit to the source. Permissions and fees are waived for authors who wish to reproduce their own material for personal use; fees only are waived for authors who wish to use their own material commercially. Permission and fees are waived for the photocopying of isolated articles for nonprofit classroom or library reserve use by instructors and educational institutions. Libraries are permitted to photocopy beyond the limits of U.S. copyright law: (1) those post-1977 articles with a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through the Copyright Clearance Center, P.O. Box 765, Schenectady, NY 12301; (2) pre-1978 articles, provided that the per-copy fee stated in the Publishers' Fee List is paid through the Copyright Clearance Center, P.O. Box 765, Schenectady, NY 12301. Address requests for reprint permission to the Permissions Office, APA, 1200 Seventeenth Street, N.W., Washington, D.C. 20036.

Subscriptions: Subscriptions are available on a calendar year basis only (January through December). Nonmember rates for 1979: $60 domestic, $64 foreign, $6 single issue. APA member rate: $20. Write to Subscription Section, APA.

Change of Address: Send change of address notice and a recent mailing label to the attention of the Subscription Section, APA, 30 days prior to the actual change of address. APA will not replace undelivered copies resulting from address changes; journals will be forwarded only if subscribers notify the local post office in writing that they will guarantee second-class forwarding postage.

Reprints: Authors may order reprints of their articles from the printer when they receive proofs.

Back Issues and Back Volumes: For information regarding back issues or back volumes write to Order Department, APA.

Microform Editions: For information regarding microform editions write to any of the following: Johnson Associates, Inc., P.O. Box 1017, Greenwich, Connecticut 06830; University Microfilms, Ann Arbor, Michigan 48106; or Princeton Microfilms, Princeton, New Jersey 08540.

Published monthly in one volume per year by the American Psychological Association, Inc., 1200 Seventeenth Street, N.W., Washington, D.C. 20036. Printed in the U.S.A. Second-class postage paid at Arlington, Va., and at additional mailing offices.

Diary No. 98
Date 24.12.79
librarian
Bureau Educ Psy Research.

JOURNAL OF
Personality
AND
Social Psychology

Volume 37 June 1979 Number 6

Feeling More Than We Can Know: Exposure Effects Without Learning

William Raft Wilson
School of Nursing
University of Michigan

Given the joint occurrence of positive attitudinal and learning effects that typically result from repeated stimulus exposure, most researchers have concluded that what an individual knows and what he or she feels about the familiar stimulus are not independent. In fact, it is generally assumed that some form of learning, such as recognition, both precedes and mediates the affective response. This study questioned that assumption by seeking to obtain positive exposure effects in the absence of stimulus recognition. In order to unconfound affect and learning effects, two experiments with female subjects employed a dichotic-listening procedure for stimulus exposure, in which the critical stimuli were exposed on the unattended channel. Reliable attitudinal enhancement toward objectively familiar stimuli was obtained in both experiments, even though recognition levels were low (Experiment 1) or at chance levels (Experiment 2). Subjective stimulus familiarity was not found to be a significant predictor of affect in either experiment. Positive feelings toward a previously encountered object are not dependent on consciously knowing or perceiving that the object is familiar. Implications for the role of awareness in processes of attitude formation under conditions of mere exposure are discussed.

Research on the cognitive and affective changes that an individual experiences as a function of repeated stimulus exposure has consistently obtained two results. At a cognitive level, the individual's ability to recognize the stimulus increases (e.g., Martin, 1968). At an affective level, the individual's attitude toward the stimulus becomes increasingly positive (e.g., Zajonc, 1968). Given the joint occurrence of these two effects, most researchers have concluded that changes in what an individual knows and in what he or she feels about the familiar stimulus are highly related. In fact, it is generally assumed that some form of learning that improves the individual's ability to identify, classify, or

This research was supported by National Science Foundation Grant GS-35495 to R. B. Zajonc.

The author wishes to thank Robert B. Zajonc for his comments on earlier versions of this manuscript, and Lily Huang for her advice on the data analyses.

Reprint requests should be sent to William R. Wilson, Research Area, School of Nursing, University of Michigan, Ann Arbor, Michigan 48109.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3706-0811$00.75

recognize the stimulus both precedes and mediates the affective response (Berlyne, 1970; Harrison, 1968; Stang, 1975).

Although the above assumption has been incorporated in all major explanations of the attitudinal effects of stimulus exposure, little research has been designed to test it experimentally. Furthermore, clear support for stimulus recognition or feelings of familiarity as a necessary condition for obtaining exposure effects has not emerged in past research. For example, Matlin (1971) reported that objectively (de facto) familiar stimuli, which were correctly recognized and subjectively familiar, were rated more positively than objectively familiar stimuli that were incorrectly identified and not subjectively familiar. Surprisingly, however, feelings of subjective familiarity did not lead to more positive ratings of stimuli if they were indeed novel. Confidence in these findings should be limited, since comparisons were based on two very infrequent recognition errors. It is possible that these judgment errors may have resulted from certain biases elicited by particular features of the stimuli that also influence affect.

In a direct attempt to connect learning to exposure effects, Stang (1975) performed a number of experiments in which he varied exposure duration and measured subjects' affective ratings of, and ability to recognize or recall, stimuli. Learning and affect curves were found to vary, almost identically, as a function of exposure duration. Stang concluded that repeated stimulus exposure leads to learning that is intrinsically rewarding, and that this reward value is associatively linked to the stimulus. Thus, when asked to rate the familiar stimulus, an individual will reflect this accumulation of reward value by giving a positive rating.

The conclusions drawn from Stang's data are questionable. First, subjects were told that they were participating in a learning study. This may have inflated the relation between stimulus recognition and affect rating, since "good" performance was dependent on accuracy of learning. Other related research suggests that subjects respond differently to stimuli during exposure, once they are informed that recognition will later be measured (e.g., Berlyne, Craw, Salapatek, & Lewis, 1963). Secondly, the joint occurrence of stimulus recognition and positive affect ratings does not in itself demonstrate that learning is a necessary condition for obtaining exposure effects. That learning is not a necessary condition was suggested by recent research that reported a significant partial correlation between exposure and affect, independent of the effects of subjective familiarity (Moreland & Zajonc, 1977).

The fact that, under typical conditions of exposure, subjective familiarity judgments and affective ratings are highly correlated can not be disputed. The critical question that remains to be investigated, however, is whether these effects are partially independent products of repeated stimulus presentations, or whether the enhanced affective response is dependent on increments in subjective familiarity. What is needed is an alternative method for stimulus exposure that can unconfound attitudinal and learning effects. One obvious approach is to establish exposure conditions in which the form of perceptual learning that commonly enhances stimulus recognition is severely attenuated or reduced to chance levels, even after repeated stimulus presentations.

Although repeated presentation of a stimulus to an individual usually does increase the probability that it will be recognized or recalled, studies of selective attention suggests that this is not an inevitable result when the information is not attended to during presentation. For example, the dichotic-listening procedure involves simultaneous transmission of different messages to each of the subject's ears. The subject is required to attend to the information being transmitted to one ear and to ignore the information coming to the other ear. Under these conditions, information in the unattended channel rarely enters consciousness, and its later recognition or recall rarely exceeds chance levels, even when processing up to a semantic level is evident (Corteen & Wood, 1972; Moray, 1959). Thus, the dichotic-listening procedure appears to be an excellent method for establishing conditions of *mere* exposure, that is, for making the stimulus accessible to the individual's percep-

tion (Zajonc, 1968) without increasing the probability of stimulus recognition.

The purpose of the present study was to examine the contribution of objective and subjective familiarity to exposure effects obtained in the absence of stimulus recognition. It was assumed that objectively familiar stimuli would be rated more positively than objectively unfamiliar stimuli, independent of how familiar these stimuli may have appeared to the individual. At the extreme, I sought to obtain the exposure effect in the absence of stimulus recognition.

Experiment 1

A modified dichotic-listening procedure for stimulus exposure was employed. Subjects were required to attend to a distractor message in one ear while the critical stimuli were being transmitted to the unattended ear. The purpose of this experiment was to determine if attitudinal enhancement effects of mere exposure could be obtained when the critical stimuli were presented on the unattended channel.

Method

Subjects

The subjects were 24 undergraduate women enrolled in introductory psychology courses at the University of Michigan. They were randomly drawn from the Department of Psychology voluntary subject pool. Female students were selected for reasons of convenience.

Materials and Apparatus

The critical stimuli were six tone sequences. Each sequence was approximately 10 sec long and appeared, to the listener, to be a simple melody. Each tone sequence consisted of 40 tone units, each unit having a given frequency selected at random from among 10 frequencies varying from 282 Hz to 631 Hz, having a given duration selected at random from among five durations varying from 100 msec to 600 msec, and having one of three loudness levels. The six sequences were of intermediate complexity level, and were chosen from among 18 created by Vitz (1964; 1966). Under ordinary listening conditions, these sequences were found to be clearly discriminable when judged for pleasantness, amount of variation, and complexity (Crozier, 1974; Vitz, 1964; 1966).

Six tapes were prepared for exposure purposes, each of which contained three of the six critical stimuli repeated five times each, for a total of 15 exposure trials per tape with 5-sec intertrial intervals. The order in which the tone sequences were played on any given tape was randomly determined. For purposes of postexposure ratings, a test tape, comprised of all six critical stimuli, randomly ordered, was prepared. Use of each tone sequence for exposure and/or testing was balanced so that half of the subjects were exposed to it prior to testing, while the other half encountered the stimulus only during testing.

The distractor message, to which subjects were asked to attend during the exposure phase, was a taped passage of a story segment taken from Daphne DuMaurier's *The Birds*. This passage was read in a monotone female voice at about 150 words per min. A typewritten version of this passage was prepared which contained a number of "errors." For example, the subject would hear "Autumn was best for this, better than spring," whereas the written passage would read, "*Author* was *been* for this, better *them* spring." There were either 3 or 4 randomly determined errors per line. Whether any particular line had 3 or 4 errors was also randomly determined.

As in previous research (e.g., Zajonc, Markus, & Wilson, 1974), a 7-point bipolar scale was employed to measure affect. The extreme points on this scale were labeled "like" and "dislike." The recognition responses were recorded on a separate sheet. On the left side of the page, the subject checked whether the test stimulus was "OLD" or "NEW." To the right of this judgment were spaces where the subject indicated whether she was sure, half-sure, or guessing when making the recognition judgment.

Both the critical stimuli and the distractor message were recorded and played back on a stereo tape recorder. All stimuli were heard by the subjects over stereo headphones. The subjects were seated at library-type desks, which prevented them from seeing the experimenter, or any of the equipment, during the experiment. Finally, volume levels for both the distractor message and the critical stimuli were measured by a sound meter.

Procedure

Exposure phase. The subjects were asked to sit down immediately upon arriving and were read the following instructions:

> Before you is a copy of a literary passage. I would like you to do two things: (1) listen to someone recite that passage while you simultaneously read the text and (2) put a slash through all words in the written text which do not correspond to the words you hear. This is like a proofreading task where all noncorresponding words are considered errors. In order to do your best at this task, you will have to listen very closely to what the speaker is saying. There will be many errors in the written text and your job

is to put a slash through as many as you can discover. If you have no questions, we will now have a short practice session.

The subject was then given eight lines of practice. Then the entire passage was played to the subject's right ear without any interruptions. Approximately 30 sec after the passage had started, the critical stimuli were played to the subject's left ear.[1] The distractor passage continued for 30 sec after the exposure phase was completed. The subjects were not forewarned about the occurrence of the critical stimuli, but they were instructed to ignore any extraneous sounds, regardless of the source, and to focus their complete attention on the literary passage. The distractor passage was played at a volume of approximately 60 dB (B)[2] and the critical stimuli at approximately 55 dB.

Testing phase. After the exposure phase, the subjects were given the following instructions for the recognition test:

I am now going to play six tone sequences that sound much like simple melodies. Three of the tone sequences you will hear were played to you during the proofreading task. Three of the tone sequences you will hear will be ones played to you for the first time. What I would like you to do is to indicate, for each tone sequence, whether or not you think it was played during the proofreading task.

The subjects were then shown how to use the recognition and confidence scales. The same set of instructions was provided for the affective ratings, with the exception that the subjects were asked to indicate how they felt toward each tone sequence. Up to 30 sec per stimulus was allowed for making a recognition or affective judgment.

To control for possible order effects, judgment of affect and recognition were counterbalanced. Half of the subjects made recognition judgments first and then gave affect ratings, while the opposite was true for the other subjects.

Results

Recognition

On the average, the subjects were able to identify correctly 3.54 of the test stimuli as OLD or NEW.[3] Analysis of this average, in comparison with the average recognition that would be expected by chance (3.00), reveals that the subjects did better than chance, $t(23) = 2.41$, $p < .05$. Thus, the method of exposure employed in this experiment appears to be relatively effective for attenuating recognition; however, it is not sufficient for eliminating recognition entirely.[4]

Affect

Responses on the affect scale were converted to numbers ranging from 0 to 6, with 6 indicating the most positive response. Each subject's ratings of stimuli in each of the four cells (e.g., "OLD"/NEW) were then averaged. If a subject did not provide ratings for stimuli in a particular cell, for example, if no OLD stimuli were identified as "OLD" by that subject, that cell was given an estimated value in the form of the subject's average rating of stimuli in all other cells combined. This procedure allowed for a straightforward comparison of the effects of objective and subjective familiarity on affective rating of stimuli. The means of both the adjusted and unadjusted ratings appear in Table 1.

A $2 \times 2 \times 2$ analysis of variance for repeated measures, with one between-subjects factor (order of judgment–affect or recognition first) and two within-subjects factors (objective familiarity—OLD vs. NEW; and subjective familiarity—"OLD" vs. "NEW"), was performed on the adjusted affect ratings. The results show no significant main effect for Order of Judgment, $F(1, 22) < 1$. They do reveal, however, that the subjects rated OLD stimuli more positively than NEW stimuli, regardless of whether they were judged "OLD" or "NEW," $F(1, 22) = 11.66$, $p < .005$ (Table

---

[1] The left ear was selected because research suggests that, in general, musical patterns are better processed and better recognized later when played to the right hemisphere (cf. Bever & Chiarello, 1974; Kimura, 1973).

[2] re 20 $\mu N/m^2$.

[3] Stimuli that appeared in the training series are designated as OLD whereas those that appeared in testing only are designated as NEW. The subjects' responses are designated as "OLD" and "NEW."

[4] In earlier pilot research to establish the validity of the recognition measure and to confirm the discriminability of the tone sequences, 12 subjects were asked to attend to the critical stimuli (rather than the distractor passage) during the exposure phase and then to do the recognition test. As might be expected under these conditions, the subjects made no errors (100% accuracy) in classifying previously exposed tone sequences as "OLD" and tone sequences exposed for the first time during testing as "NEW." Thus, there appears to be quite a substantial reduction in recognition accuracy when the tone sequences are presented on the unattended channel.

2). There is also a marginally significant tendency for the subjects to rate stimuli that they judged as "OLD" more positively than stimuli judged as "NEW," regardless of whether or not the stimulus was objectively OLD or NEW, $F(1, 22) = 3.41$, $p < .10$.

Even though the Objective Familiarity × Subjective Familiarity interaction is not significant, it is of interest to examine, separately, the ratings of objective familiar stimuli called "OLD" and "NEW." For stimuli judged "OLD," truly OLD stimuli were liked better than NEW stimuli, although this effect is not significant, $F(1, 22) = 2.23$. This effect, however, is reliable for stimuli judged "NEW": Truly OLD stimuli were liked better than truly NEW stimuli, $F(1, 22) = 7.40$, $p < .05$. Thus the exposure effect was obtained even in the absence of the subject's awareness of his own differential experience with the stimuli.

The relationship between affect ratings and subjective familiarity was further examined by combining recognition and confidence judgments into a single measure of subjective familiarity. Each stimulus was classified as to its degree of oldness, ranging from "OLD-SURE" (6) to "NEW-SURE" (1). Then each subject's average oldness rating and average affect rating were correlated for OLD and NEW stimuli, separately. The correlations obtained for NEW and OLD stimuli were −.29 and .14 ($df = 22$), respectively. In the case of objectively NEW stimuli, a negative correlation indicates that the more certain the subject was that the stimulus was "NEW," the better she liked it. Conversely, in the case of OLD stimuli, a positive correlation indicates that the more certain the subject was that the stimulus was "OLD," the better she liked it. In short, although neither of the above correlations is significant, they are in the direction of a positive relationship between liking and degree of subjective certainty about the true familiarity of the stimulus. Because the correlation was positive for OLD stimuli and negative for NEW stimuli, these results do not indicate that affective ratings depend, to any significant degree, on oldness ratings per se. If the latter were true, we would expect a positive correlation between oldness and affect ratings, for *both* OLD and NEW stimuli.

Intercorrelations among all the measures are provided in Table 2.

Table 1
Average Stimulus Affect Ratings as a Function of Objective Familiarity (OLD–NEW) and Subjective Familiarity ("OLD"–"NEW")

Objective familiarity	Subjective familiarity "OLD"	Subjective familiarity "NEW"	*M*
OLD	4.22 (4.16)	4.00 (4.11)	4.12 (4.12)
NEW	3.77 (3.33)	3.04 (3.19)	3.30 (3.23)
M	4.02 (3.83)	3.52 (3.60)	

Note. Figures without parentheses include estimated values for some subjects. The number of subjects for which values are estimated is 2, 4, 6, and 1 for the "OLD"/OLD, "NEW"/OLD, "OLD"/NEW, and "NEW"/NEW cells, respectively. Responses on the affect scale range from 0 to 6, with 6 indicating the most positive evaluation.

Table 2
Intercorrelations Among the Measures

Measure	1	2	3	4
1. Affect	1.00			
2. Objective familiarity	.26**	1.00		
3. Subjective familiarity	.07	.18*	1.00	
4. Oldness	.09	.15	.88***	1.00

Note. Objective familiarity was controlled by the experimenter and refers to the true degree of contact that the subject had with the stimulus. The other variables were under the subject's control. The oldness response is comprised of both the subject's familiarity ("OLD" or "NEW") judgment and the confidence ("SURE," "HALF-SURE," or "GUESS") rating of that judgment.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

Experiment 2

The procedure employed in the first experiment proved to be relatively successful in attenuating recognition accuracy, but it did not reduce it to a chance level. Nevertheless, under those conditions, objective familiarity was found to influence affective ratings, inde-

pendently of other processes. To some extent, subjective familiarity and subjective certainty were found to be related to affective judgments, as well. Whether these same effects would be observed under conditions in which subjects could not recognize the critical stimuli at better than chance levels remained to be demonstrated. Therefore, in this second experiment, a modified procedure was employed in order to eliminate all reliable recognition; during stimulus exposure, the subjects were required to verbally shadow the distractor message and simultaneously do the proofreading task.

To gain a better understanding of the cognitive and affective changes that result when the critical stimuli are exposed in this unobtrusive fashion, two additional measures, found in past research to be sensitive to levels of stimulus familiarity, were employed. The first measure was latency of recognition judgments. In general, familiar stimuli are recognized not only better but also faster than less familiar stimuli are (cf. Matlin, 1971). It is possible that subjects respond faster to objectively familiar stimuli, even when they do not recognize them as familiar. This finding would suggest that, even in the absence of stimulus recognition, there may be some form of differential cognitive processing of de facto familiar and unfamiliar stimuli.

The second new measure was an indicator of the subject's level of arousal when evaluating each stimulus, namely, galvanic skin response (GSR). Both Berlyne et al. (1963) and Zajonc (1968) have reported that novel stimuli evoke a greater drop in skin resistance than do familiar stimuli. Even in the absence of stimulus recognition, such physiological differences in response to de facto familiar and unfamiliar stimuli may occur, given that changes in skin resistance are a function of information (about the stimulus) that is not used in making the recognition judgment (see Eriksen, 1960).

Method

Subjects

The subjects were 24 undergraduate women enrolled in introductory psychology courses at the University of Michigan. They were randomly drawn from the Department of Psychology voluntary subject pool.

Materials and Apparatus

The critical stimuli and the distractor message were the same as those employed in Experiment 1. All the equipment used in the first experiment was employed in this experiment. In addition, a response panel was constructed to register the subject's recognition and confidence judgments. The subject made her recognition and confidence judgments by pressing the appropriate buttons on the response panel "OLD" or "NEW" for recognition, and "SURE," "HALF-SURE," or "GUESS" for confidence judgments. The subject's response activated a light that indicated to the experimenter which response had been made. The latency of the subject's response was recorded by a clock, outside of the subject's vision.

Changes in skin resistance (GSR) were measured on an event recorder, connected to a control unit and to Lykken zinc-zinc sulphate electrodes. The skin resistance was recorded for each subject on a scale calibrated at 500 k-ohms, with a 600 k-ohms limit. This system relies on a 10-μV current (see Kaplan & Hobart, 1964). Change in skin resistance was arbitrarily defined as the distance in ohm units between the lowest and highest point of resistance within a 20-sec period immediately following stimulus onset. Affect ratings were obtained in the same way as in Experiment 1.

Procedure

Exposure phase. When the subject arrived, the experimenter connected the electrodes to the first and third fingers of the subject's nonpreferred hand. The subject was told that the electrodes would cause no inconvenience, other than that due to their presence, and that the purpose of the electrodes would be explained after the experiment. The subject was then seated at the experimental desk and read the following instructions:

> You are about to hear, over these headphones, a literary passage read by a young woman. I would like you to do two things. First, repeat each word that this person says, immediately after you hear it. She will be speaking at a fairly fast rate, so you will have to work hard to keep up with her. If you should fall behind by more than two or three words, start again with the last word that she has just spoken. I will be recording your repetition of the passage on the microphone sitting before you, so be sure to speak loudly and clearly. Secondly, while you are listening to and repeating what the woman is saying, I would like you to read the text before you. When you see a word that does not correspond to what you are hearing and saying, put a slash through it. This is like

proofreading task where all noncorresponding words are considered errors. In order to do your best at this task, you will have to listen very closely to what the speaker is saying. There will be many errors in the written text, and your job is to put a slash through as many as you can discover. If you have no questions, we will now have a short practice session.

The subject was then given eight lines of practice, followed by a restatement of the task requirements. The rest of the exposure phase was the same as in Experiment 1.

Testing phase. The instruction and procedure employed for testing was comparable to that used in Experiment 1. In this experiment, however, the subject was instructed on how to use the apparatus for indicating recognition judgments; also, the fact that half of the stimuli were truly OLD was emphasized again, just prior to testing.

Results

Recognition

The subjects identified stimuli as "OLD" or "NEW," independently of their true exposure history, with an overall average level of accuracy of only 3.17, $t(23) < 1$. Thus, the method of exposure employed in this experiment achieved the desired condition of chance recognition.

Table 3
Average Stimulus Affect Ratings as a Function of Objective Familiarity (OLD–NEW) and Subjective Familiarity ("OLD"–"NEW")

Objective familiarity	Subjective familiarity "OLD"	Subjective familiarity "NEW"	*M*
OLD	3.51 (3.61)	3.85 (3.87)	3.66 (3.72)
NEW	3.03 (3.08)	3.02 (3.06)	3.03 (3.07)
M	3.29 (3.36)	3.40 (3.42)	

Note. Figures without parentheses include estimated values for some subjects. The number of subjects for which values are estimated is 2, 5, 2, and 2 for the "OLD"/OLD, "NEW"/OLD, "OLD"/NEW, and "NEW"/NEW cells, respectively. Responses on the affect scale range from 0 to 6, with 6 indicating the most positive evaluation.

Table 4
Average Recognition Judgment Latencies (in sec) as a Function of Objective Familiarity (OLD–NEW) and Subjective Familiarity ("OLD"–"NEW")

Objective familiarity	Subjective familiarity "OLD"	Subjective familiarity "NEW"	*M*
OLD	12.12 (12.34)	11.76 (11.53)	11.97 (11.99)
NEW	13.35 (13.52)	11.17 (11.61)	12.26 (12.57)
M	12.70 (12.89)	11.44 (11.57)	

Note. Figures without parentheses include estimated values for some subjects. The number of subjects for which values are estimated is 2, 5, 2, and 2 for the "OLD"/OLD, "NEW"/OLD, "OLD"/NEW, and "NEW"/NEW cells, respectively.

Affect

Liking ratings were scored and analyzed as described in the first experiment. Again, no significant effects were obtained for Order of Judgment, $F(1, 22) < 1$. As we can see in Table 3, OLD stimuli were rated more positively than NEW stimuli, regardless of whether they were judged "OLD" or "NEW," $F(1, 22) = 5.71$, $p < .05$, a finding that is consistent with the one obtained in Experiment 1. Now, however, subjective familiarity has no influence on affect ratings, $F(1, 22) < 1$, and once again, no significant effects are observed for the Objective Familiarity × Subjective Familiarity interaction, or for any other interaction.

When the main effect of objective familiarity is decomposed, the pattern of results obtained is similar to that found in the first experiment. OLD stimuli were liked somewhat more than NEW stimuli, even when both were identified as "OLD"; but the effect was not significant, $F(1, 22) = 1.51$. However, the exposure effect was again obtained without the subject having been aware of her differential experience with the tone sequences, in that OLD stimuli were liked significantly more than NEW stimuli, even when both were judged as "NEW," $F(1, 22) = 6.89$, $p < .05$.

As in Experiment 1, a 6-point oldness measure of subjective familiarity was obtained by combining recognition and confidence judgments. The average oldness ratings and the average affect ratings were correlated for OLD and NEW stimuli, separately. The coefficients suggest little relation between affect and subjective familiarity or subjective certainty—for OLD stimuli $r(22) = .09$, and for NEW stimuli $r(22) = -.02$.

GSR and Latency

GSR and recognition latency scores were subjected to the same analysis as affect ratings. Little systematic variation in skin resistance was obtained among the different cells. Analysis revealed no significant effect for order of judgment, objective familiarity, subjective familiarity, or the OLD × "OLD" interaction.

Table 4 presents average latency responses. Neither order of judgment, $F(1, 22) = 1.49$, nor objective familiarity, $F(1, 22) < 1$, emerged as a significant factor. There is a reliable difference between latency of "OLD" and "NEW" judgments, $F(1, 22) = 8.36$, $p < .01$, the average difference being 1.25 sec. There were no significant interactions.

Intercorrelations among all the measures are provided in Table 5.

General Discussion

When the attitudinal and learning effects produced by repeated stimulus exposure are unconfounded, an interesting pattern of results emerges. First, reliable exposure effects are obtained, that is, previously encountered stimuli are rated more positively than novel stimuli. Second, subjective judgments of stimulus familiarity are not reliably related to affective evaluations, even when confidence ratings are considered. Together, these results suggest that an individual's positive feelings toward a previously encountered object are not dependent on his consciously knowing or perceiving that the object is familiar. Such findings have important implications for the role of awareness in processes of attitude formation under conditions of mere exposure.

Explanations of the Exposure Effect

The first major explanation of the attitudinal effects of exposure is the *response competition* hypothesis proposed by Harrison (1968). He suggested that when a stimulus is first encountered, it elicits a number of response tendencies that compete for emission. This induction of response competition leads to a state of tension in the individual, characterized by "negative affect and a motivational impetus to eliminate the tension" (p. 363). With repeated exposure of the stimulus, however, only a limited number of compatible response tendencies achieve ascendancy. The end result of repeated stimulus presentations is a reduction in response competition, which is reflected by an increment of positive affect

Table 5
Intercorrelations Among the Measures

Measure	1	2	3	4	5	6
1. Affect	1.00					
2. Objective familiarity	.18*	1.00				
3. Subjective familiarity	−.02	.07	1.00			
4. Oldness	−.05	.10	.85**	1.00		
5. Latency	.06	−.06	.13	.08	1.00	
6. Galvanic skin response	−.05	−.01	.05	.03	−.10	1.00

Note. Objective familiarity was controlled by the experimenter and refers to the true degree of contact that the subject had with the stimulus. The other variables were under the subject's control. The oldness response is comprised of both the subject's familiarity ("OLD" or "NEW") judgment and the confidence ("SURE," "HALF-SURE," or "GUESS") rating of that judgment.

* $p < .05$.

** $p < .001$.

for the familiar stimulus. Although the point is never made explicitly by Harrison, it is clear that stimulus recognition is a sine qua non for the reduction of response competition, that is, access to, and retrieval of, previously stored associations is contingent on the elicitation of a stable encoding response by the individual over successive exposures (cf. Höffding, 1891). Thus, an individual must correctly identify the stimulus in order to have access to past associations (response tendencies) that have been linked to it (cf. Lawrence, 1963; Martin, 1968).

A second major explanation of exposure effects is a *two-factor* learning theory proposed by Berlyne (1970). He postulated that individuals have a preference for some optimal level of arousal. When extremely novel or complex stimuli are encountered, arousal is increased above the optimal level and the stimuli elicit low or even negative affect in the individual. But with repeated encounters, the arousal potential of the stimuli declines because they are processed with greater ease and are better understood. Thus, familiar stimuli will be favorably evaluated following exposure, because the individual's arousal level is modified back toward the optimum. Berlyne reasoned that "opportunities for further acquaintance with the pattern will provide scope for the perceptual and ideational processing that enables uncertainty and conflict to be resolved as elements are discriminated, classified, recognized, and grouped together as sub-wholes. This can evidently be a source of pleasure, presumably dependent on the arousal-reduction mechanism" (p. 285). Clearly, the role of stimulus recognition is central to this theory as well.

Obviously, the present results raise questions about the validity of the above theories or any other explanations of exposure effects that implicitly include overt stimulus recognition as a prerequisite for obtaining enhanced affect (e.g., Burgess & Sales, 1971; Crandall, 1967). In the current experiments, it is unlikely that the subjects were able to engage in strategies to reduce uncertainties, form associations, and/or systematically explore the critical stimuli in order to promote subsequent recognition ability. Furthermore, the data clearly suggest that subjective familiarity, veridical or otherwise, does not influence affective responses to previously encountered stimuli in a simple fashion, in such a way that the former always precedes and mediates the latter. Thus, one must question any theory that assumes that individuals must be aware of their past experiences with the stimulus as a sine qua non for obtaining the frequency–affect relationship.

In a related vein, some researchers (e.g., Stang, 1974) have suggested that exposure effects may result from demand characteristics, because the subject is "aware" of the familiarity–affect hypothesis and responds accordingly. Even granting that such awareness was present in these experiments, it would be of little use to the subjects, since they could not reliably determine which stimuli they had heard previously. Thus it is difficult to interpret the present results as artifactual. One might argue, however, that if subjects were aware of the familiarity–affect hypothesis and were conforming to demand characteristics, they would consistently assign high affect ratings to stimuli that they judged to be familiar and low ratings to stimuli judged unfamiliar. The data reveal no such effects.

The Role of Awareness

Although the results from this study do not provide a sufficient basis for generating an alternative model to account for exposure effects, they do suggest that if learning processes underlie these effects, those processes may not require conscious awareness or effort by the subject.[5] In other words, one might argue that the subjects did engage in some form of learning during exposure (i.e., developed expectancies, reduced uncertainty, alleviated response competition, and the like) but not at the level of conscious awareness. Such lack of awareness during exposure may have prevented the subjects from later re-

[5] It is interesting to note that frequency information appears to be encoded with little effort or conscious intent by individuals (Hasher & Chromiak, 1977; Howell, 1973a, 1973b).

liably retrieving, from memory, information about the occurrence of the stimulus (cf. Melton & Martin, 1972), but the affective response was not impaired because learning without awareness had occurred. The plausibility of such an explanation is suggested by recently developed models of human information processing that question the necessity (Erdelyi, 1974) and even the efficacy (Nisbett & Wilson, 1977) of conscious awareness of processes underlying affective reactions.

Related demonstrations of such effects have been reported in studies of perception and learning without awareness. For example, pictures presented too briefly (1 msec) to obtain correct verbal identification or recognition responses nevertheless produced reliable differences in evoked potentials (average evoked response) and interpretable free associations (Shevrin & Fritzler, 1968; Shevrin, Smith, & Fritzler, 1971). Employing a dichotic-listening procedure, Konecni and Slamecka (1972) found that verbal conditioning effects could be obtained even when subjects were not aware of the verbal reinforcer presented on the unattended channel.

The merit of a learning-without-awareness explanation of exposure effects needs, however, to be demonstrated. No evidence in the present results can be invoked to support or deny such a process. Support for a theory that postulates that the exposure effect is mediated by a completely covert learning process must demonstrate that this process occurs prior to affect changes, and must employ measures that are independent of such changes.

Conclusions

If exposure effects need not involve the individual's awareness of the familiarization process, then the influences of proximity and familiarity on attitude formation may be even more pervasive than originally suspected. It has been established that objects and individuals we *knowingly* encounter or interact with (even incidentally) are likely to come to be regarded positively (Festinger, Schachter, & Back, 1950; Saegert, Swap, & Zajonc, 1973; Segal, 1974; Zajonc, 1968). The present results raise the possibility that the objects and individuals we encounter in our environment are likely to engender a positive orientation, regardless of our knowledge of past encounters with them. Thus, in some cases, what we feel about objects or individuals may precede our "awareness" of previous contact! One might speculate that, in the extreme case, unconsciously formed positive attitudes toward "novel" objects or "strangers" in our environment stimulate both an awareness of their existence and conscious attempts to obtain knowledge about them. Whereas prior research has focused largely on how learning mediates affect, we should consider more seriously the Moreland and Zajonc (1977) proposition that changes in affect following exposure may influence learning.[6]

Obviously, the validity of such speculations must be determined by additional research. The results from the present study suggest that future work may require a departure from traditional methods of studying exposure effects and, to some extent, a change in assumptions about the conditions under which these effects can occur. Such research must determine whether familiarity mediates affect directly, without the intervention of other processes, and, if so, whether changes in affect influence learning processes. Finally, we need to examine more explicitly the role of awareness in exposure effects, in order to better understand how familiarity breeds liking.

---

[6] In a recent article, Birnbaum and Mellers (1979) are critical of the partial correlation technique employed by Moreland and Zajonc (1977) to determine that affect and recognition are two independent outcomes of stimulus exposure (i.e., that these responses are mediated by two different mechanisms). They argue that a single mediator, labeled *subjective recognition*, could account for both the affect and the recognition responses obtained by Moreland and Zajonc under the very likely conditions that measures of these two responses are not perfectly correlated. The present results appear to make some aspects of their thesis questionable. In contrast to Moreland and Zajonc's technique, the recognition factor in the present study was partialed out of the effects experimentally rather than statistically, but affect was still found to be reliably related to objective frequency.

References

Berlyne, D. E. Novelty, complexity, and hedonic value. *Perception and Psychophysics,* 1970, *8,* 279–286.

Berlyne, D. E., Craw, M. A., Salapatek, P. H., & Lewis, J. L. Novelty, complexity, incongruity, extrinsic motivation, and GSR. *Journal of Experimental Psychology,* 1963, *66,* 560–567.

Bever, T. G., & Chiarello, R. J. Cerebral dominance in musicians and nonmusicians. *Science,* 1974, *185,* 537–539.

Birnbaum, M. H., & Mellers, B. A. Stimulus recognition may mediate exposure effects. *Journal of Personality and Social Psychology,* 1979, *37,* 391–394.

Burgess, T. D., & Sales, S. M. Attitudinal effects of mere exposure: A reevaluation. *Journal of Experimental Social Psychology,* 1971, *7,* 461-472.

Corteen, R. S., & Wood, B. Autonomic responses to shock-associated words in an unattended channel. *Journal of Experimental Psychology,* 1972, *94,* 308–313.

Crandall, J. E. Familiarity, preference, and expectancy arousal. *Journal of Experimental Psychology,* 1967, *73,* 374–381.

Crozier, J. B. Verbal and exploratory responses to sound sequences varying in uncertainty level. In D. E. Berlyne (Ed.), *Studies in the new experimental aesthetics.* Washington, D.C.: Hemisphere, 1974.

Erdelyi, M. H. A new look at the new look: Perceptual defense and vigilance. *Psychological Review,* 1974, *81,* 1–25.

Eriksen, C. W. Discrimination and learning without awareness: A methodological survey and evaluation. *Psychological Review,* 1960, *67,* 279–300.

Festinger, L., Schachter, S., & Back, K. *Social pressures in informal groups: A study of human factors in housing.* New York: Harper & Row, 1950.

Harrison, A. A. Response competition, frequency, exploratory behavior, and liking. *Journal of Personality and Social Psychology,* 1968, *9,* 363–368.

Hasher, L., & Chromiak, W. The processing of frequency information: An automatic mechanism. *Journal of Verbal Learning and Verbal Behavior,* 1977, *16,* 173–184.

Höffding, H. *Outlines of psychology.* (M. E. Lowndes, trans.). London: Macmillan, 1891.

Howell, W. C. Representation of frequency in memory. *Psychological Bulletin,* 1973, *80,* 44–53. (a)

Howell, W. C. Storage of events and event frequencies: A comparison of two paradigms in memory. *Journal of Experimental Psychology,* 1973, *98,* 260–263. (b)

Kaplan, S., & Hobart, J. L. A versatile device for the measurement of skin resistance in rats and humans. *American Journal of Psychology,* 1964, *77,* 309–310.

Kimura, D. The asymmetry of the human brain. *Scientific American,* 1973, *228,* 70–80.

Konecni, V. J., & Slamecka, N. J. Awareness in verbal nonoperant conditioning: An approach through dichotic listening. *Journal of Experimental Psychology,* 1972, *94,* 248–254.

Lawrence, D. H. The nature of a stimulus: Some relationships between learning and perception. In S. Koch (Ed.), *Psychology: A study of science* (Vol. 5). New York: McGraw-Hill, 1963.

Martin, E. Stimulus meaningfulness and paired-associate transfer: An encoding variability hypothesis. *Psychological Review,* 1968, *75,* 421–444.

Matlin, M. W. Response competition, recognition, and affect. *Journal of Personality and Social Psychology,* 1971, *19,* 295–300.

Melton, A. W., & Martin, E. *Coding processes in human memory.* Washington, D.C.: Winston, 1972.

Moray, N. Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology,* 1959, *11,* 56–60.

Moreland, R. L., & Zajonc, R. B. Is stimulus recognition a necessary condition for the occurrence of exposure effects? *Journal of Personality and Social Psychology,* 1977, *35,* 191–199.

Nisbett, R. E., & Wilson, T. D. Telling more than we can know: Verbal reports on mental processes. *Psychological Review,* 1977, *84,* 231–259.

Saegert, S., Swap, W. C., & Zajonc, R. B. Exposure, context, and interpersonal attraction. *Journal of Personality and Social Psychology,* 1973, *25,* 234–242.

Segal, M. Alphabet and attraction: An unobtrusive measure of the effect of propinquity in a field setting. *Journal of Personality and Social Psychology,* 1974, *30,* 654–657.

Shevrin, H., & Fritzler, D. E. Visual evoked response correlates of unconscious mental processes. *Science,* 1968, *161,* 295–298.

Shevrin, H., Smith, W. H., & Fritzler, D. E. Average evoked response and verbal correlates of unconscious mental processes. *Psychophysiology,* 1971, *8,* 149–162.

Stang, D. J. Intuition as artifact in mere exposure studies. *Journal of Personality and Social Psychology,* 1974, *30,* 647–653.

Stang, D. J. Effects of "mere exposure" on learning and affect. *Journal of Personality and Social Psychology,* 1975, *31,* 7–12.

Vitz, P. C. Preferences for rates of information presented by sequences of tones. *Journal of Experimental Psychology,* 1964, *68,* 176–183.

Vitz, P. C. Affect as a function of stimulus variation. *Journal of Experimental Psychology,* 1966, *71,* 74–79.

Zajonc, R. B. The attitudinal effects of mere exposure. *Journal of Personality and Social Psychology Monograph,* 1968, *9*(2, Pt. 2).

Zajonc, R. B., Markus, H., & Wilson, W. R. Exposure effects and associative learning. *Journal of Experimental Social Psychology,* 1974, *10,* 248–263.

Received April 7, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 6, 822-832

Many Hands Make Light the Work: The Causes and Consequences of Social Loafing

Bibb Latané, Kipling Williams, and Stephen Harkins
Ohio State University

Two experiments found that when asked to perform the physically exerting tasks of clapping and shouting, people exhibit a sizable decrease in individual effort when performing in groups as compared to when they perform alone. This decrease, which we call social loafing, is in addition to losses due to faulty coordination of group efforts. Social loafing is discussed in terms of its experimental generality and theoretical importance. The widespread occurrence, the negative consequences for society, and some conditions that can minimize social loafing are also explored.

There is an old saying that "many hands make light the work." This saying is interesting for two reasons. First, it captures one of the promises of social life—that with social organization people can fulfill their individual goals more easily through collective action. When many hands are available, people often do not have to work as hard as when only a few are present. The saying is interesting in a second, less hopeful way—it seems that when many hands are available, people actually work less hard than they ought to.

Over 50 years ago a German psychologist named Ringelmann did a study that he never managed to get published. In rare proof that unpublished work does not necessarily perish, the results of that study, reported only in summary form in German by Moede (1927), have been cited by Dashiell (1935), Davis (1969), Köhler (1927), and Zajonc (1966) and extensively analyzed by Steiner (1966, 1972) and Ingham, Levinger, Graves, and Peckham (1974). Apparently Ringelmann simply asked German workers to pull as hard as they could on a rope, alone or with one, two, or seven other people, and then he used a strain gauge to measure how hard they pulled in kilograms of pressure.

Rope pulling is, in Steiner's (1972) useful classification of tasks, maximizing, unitary, and additive. In a maximizing task, success depends on how much or how rapidly something is accomplished and presumably on how much effort is expended, as opposed to an optimizing task, in which precision, accuracy, or correctness are paramount. A unitary task cannot be divided into separate subtasks—all members work together doing the same thing and no division of labor is possible. In an additive task, group success depends on the *sum* of the individual efforts, rather than on the performance of any subset of members. From these characteristics, we should expect three people pulling together on a rope with perfect efficiency to be able to exert three times as much force as one person can, and eight people to exert eight times as much force.

Ringelmann's results, however, were strikingly different. When pulling one at a time, individuals averaged a very respectable 63 kg of pressure. Groups of three people were able to exert a force of 160 kg, only two and a

This research was supported by National Science Foundation Grant GS40194.

The authors would like to thank Lewis Hinkle for technical assistance and Edward Diener, John Harvey, Norbert Kerr, Robert Kidd, George Levinger, Thomas Ostrom, Richard Petty, and Ladd Wheeler for their valuable comments.

S. Harkins is now an assistant professor of psychology at Northeastern University.

Requests for reprints should be sent to Bibb Latané, Behavioral Sciences Laboratory, Ohio State University, 404B West 17th Avenue, Columbus, Ohio 43210.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3706-0822$00.75

half times the average individual performance, and groups of eight pulled at 248 kg, less than four times the solo rate. Thus the collective group performance, while increasing somewhat with group size, was substantially less than the sum of the individual efforts, with dyads pulling at 93% of the sum of their individual efforts, trios at 85%, and groups of eight at only 49%. In a way somewhat different from how the old saw would have it, many hands apparently made light the work.

The Ringelmann effect is interesting because it seems to violate both common stereotype and social psychological theory. Common stereotype tells us that the sense of team participation leads to increased effort, that group morale and cohesiveness spur individual enthusiasm, that by pulling together groups can achieve any goal, that in unity there is strength. Social psychological theory holds that, at least for simple, well-learned tasks involving dominant responses, the presence of other people, whether as co-workers or spectators, should facilitate performance. It is thus important to find out whether Ringelmann's effect is replicable and whether it can be obtained with other tasks.

The Ringelmann effect is also interesting because it provides a different arena for testing a new theory of social impact (Latané, 1973). Social impact theory holds that when a person stands as a target of social forces coming from other persons, the amount of social pressure on the target person should increase as a multiplicative function of the strength, immediacy, and number of these other persons. However, if a person is a member of a group that is the target of social forces from outside the group, the impact of these forces on any given member should diminish in inverse proportion to the strength, immediacy, and number of group members. Impact is divided up among the group members, in much the same way that responsibility for helping seems to be divided among witnesses to an emergency (Latané & Darley, 1970). Latané further suggests that just as psychophysical reactions to external stimuli can be described in terms of a power law (Stevens, 1957), so also should reactions to social stimuli, but with an exponent having an absolute value less than 1, so that the nth person should have less effect than the $(n-1)$th. Ringelmann's asking his workers to pull on a rope can be considered social pressure. The more people who are the target of this pressure, the less pressure should be felt by any one person. Since people are likely to work hard in proportion to the pressure they feel to do so, we should expect increased group size to result in reduced efforts on the part of individual group members. These reduced efforts can be called "social loafing"—a decrease in individual effort due to the social presence of other persons. With respect to the Ringelmann phenomenon, social impact theory suggests that at least some of the effect should be due to reduced efforts on the part of group participants, and that this reduced effort should follow the form of an inverse power function having an exponent with an absolute value less than one.

The Ringelmann effect is interesting for a third reason: If it represents a general phenomenon and is not restricted to pulling on a rope, it poses the important practical question of when and why collective efforts are less efficient than individual ones. Since many components of our standard of life are produced through one form or another of collective action, research identifying the causes and conditions of inefficient group output and suggesting strategies to overcome these inefficiencies is clearly desirable.

For these three and other reasons, we decided to initiate a program of research into the collective performance of individuals in groups.

Experiment 1

Clap Your Hands and Shout Out Loud

One of the disadvantages of Ringelmann's rope pulling task is that the equipment and procedures are relatively cumbersome and inefficient. Therefore, we decided to keep our ears open for other tasks that would allow us to replicate the Ringelmann finding conceptually and would provide the basis for extended empirical and theoretical analysis. We chose cheering and clapping, two activities

that people commonly do together in social settings and that are maximizing, unitary, and additive. As with rope pulling, output can be measured in simple physical units that make up a ratio scale.

Method

On eight separate occasions, groups of six undergraduate males were recruited from introductory psychology classes at Ohio State University; they were seated in a semicircle, 1 m apart, in a large soundproofed laboratory and told, "We are interested in judgments of how much noise people make in social settings, namely cheering and applause, and how loud they seem to those who hear them. Thus, we want each of you to do two things: (1) Make noises, and (2) judge noises." They were told that on each trial "the experimenter will tell you the trial number, who is to perform and whether you are to cheer (Rah!) or clap. When you are to begin, the experimenter will count backwards from three and raise his hand. Continue until he lowers it. We would like you to clap or cheer for 5 seconds as loud as you can." On each trial, both the performers and the observers were also asked to make magnitude estimates of how much noise had been produced (Stevens, 1966). Since these data are not relevant to our concerns, we will not mention them further.

After some practice at both producing and judging noise, there were 36 trials of yelling and 36 trials of clapping. Within each modality, each person performed twice alone, four times in pairs, four times in groups of four, and six times in groups of six. These frequencies were chosen as a compromise between equating the number of occasions on which we measured people making noise alone or in groups (which would have required more noisemaking in fours and sixes) and equating the number of individual performances contributing to our measurements in the various group sizes (which would have required more noisemaking by individuals and pairs). We also arranged the sequence of performances to space and counterbalance the order of conditions over each block of 36 trials, while making sure that no one had to perform more than twice in a row.

Performances were measured with a General Radio sound-level meter, Model 1565A, using the C scale and the slow time constant, which was placed exactly 4 m away from each performer. The C scale was used so that sounds varying only in frequency or pitch would be recorded as equally loud. Sound-level meters are read in decibel (dB) units, which are intended to approximate the human reaction to sound. For our purposes, however, the appropriate measure is the effort used in generating noise, not how loud it sounds. Therefore, our results are presented in terms of dynes/cm^2, the physical unit of work involved in producing sound pressure.

Because people shouted and clapped in full view and earshot of each other, each person's performance could affect and be affected by the others. For this reason, the group, rather than the individual, was the unit of analysis, and each score was based on the average output per person. Results were analyzed in a $4 \times 2 \times 2$ analysis of variance, with Group Size (1, 2, 4, 6), Response Mode (clapping vs. shouting), and Replications (1, 2) as factors.

Results

Participants seemed to adapt to the task with good humor if not great enthusiasm. Nobody refused to clap or shout, even though a number seemed somewhat embarrassed or shy about making these noises in public. Despite this, they did manage to produce a good deal of noise. Individuals averaged 84 dB (C) clapping and 87 dB cheering, while groups of six clapped at 91 dB and shouted at 95 dB (an increment of 6 dB represents a doubling of sound pressure).

As might be expected, the more people clapping or cheering together, the more intense the noise and the more the sound pressure produced. However, it did not grow in proportion to the number of people: The average sound pressure generated *per person* decreased with increasing group size, $F(3, 21) = 41.5$, $p < .001$. People averaged about 3.7 dynes/cm^2 alone, 2.6 in pairs, 1.8 in foursomes, and about 1.5 in groups of six (Figure 1). Put another way, two-person groups performed at only 71% of the sum of their individual capacity, four-person groups at 51%, and six-person groups at 40%. As in pulling ropes, it appears that when it comes to clapping and shouting out loud, many hands do, in fact, make light the work.

People also produced about 60% more sound power when they shouted than when they clapped, $F(1, 7) = 8.79$, $p < .01$, presumably reflecting physical capacity rather than any psychological process. There was no effect due to blocks of trials, indicating that the subjects needed little or no practice and that their performance was not deleteriously affected by fatigue. In addition, there were no interactions among the variables.

Discussion

The results provide a strong replication of Ringelmann's original findings, using a com-

pletely different task and in a different historical epoch and culture. At least when people are making noise as part of a task imposed by someone else, voices raised together do not seem to be raised as much as voices raised alone, and the sound of 12 hands clapping is not even three times as intense as the sound of 2.

Zajonc's (1965) elegant theory of social facilitation suggests that people are aroused by the mere presence of others and are thus likely to work harder (though not necessarily to achieve more) when together. Although social facilitation theory might seem to predict enhanced group performance on a simple task like clapping or shouting, in the present case it would not predict any effect due to group size, since the number of people present was always eight, six participants and two experimenters. Evaluation apprehension theory (Cottrell, 1972) would also not predict any effect as long as it is assumed that coactors and audience members are equally effective in arousing performance anxiety. Therefore, these theories are not inconsistent with our position that an unrelated social process is involved. The results of Experiment 1 also can be taken as support for Latané's (1973) theory of social impact: The impact that the experimenters have on an individual seems to decrease as the number of coperformers increases, leading to an apparent drop in individual performance, a phenomenon we call social loafing.

However, there is an alternative explanation to these results. It may be, not that people exert less effort in groups, but that the group product suffers as a result of group inefficiency. In his invaluable theoretical analysis of group productivity, Steiner (1972) suggests that the discrepancy between a group's potential productivity (in this case n times the average individual output) and its actual productivity may be attributed to faulty social process. In the case of Ringelmann's rope pull, Steiner identifies one source of process loss as inadequate social coordination. As group size increases, the number of "coordination links," and thus the possibility of faulty coordination (pulling in different directions at different times), also increases.

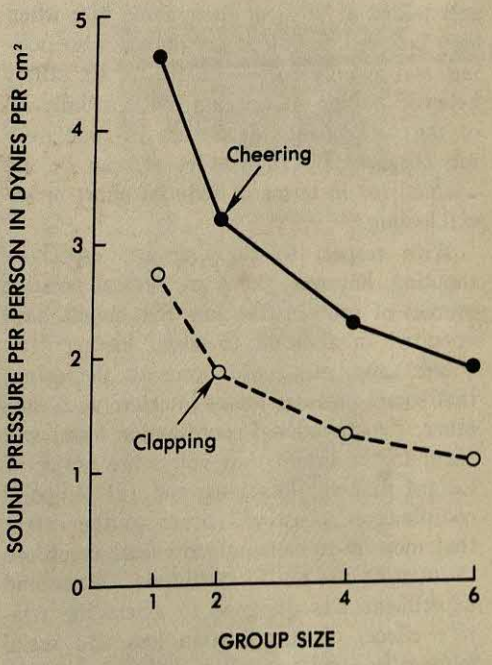


Figure 1. Intensity of noise as a function of group size and response mode, Experiment 1.

Steiner shows that for Ringelmann's original data the decrement in obtained productivity is exactly proportional to the number of coordination links.

Ingham et al. (1974) designed an ingenious experiment to determine whether the process losses found in rope pulling were mainly due to problems of coordinating individual efforts and the physics of the task, or whether they resulted from reductions in personal exertion (what we have called social loafing). First, they conducted a careful replication of Ringelmann's original rope-pulling study and found similar results—dyads pulled at 91% of the sum of their individual capacities, trios at 82%, and groups of six at only 78%.

In a second experiment, Ingham et al. cleverly arranged things so that only the individual's perception of group size was varied. Individuals were blindfolded and led to believe that others were pulling with them, but in fact, they always pulled alone. Under these conditions, of course, there is no possibility of loss due to faulty synchronization. Still there was a substantial drop in output with increases in perceived group size: Individ-

uals pulled at 90% of their alone rate when they believed one other person was also pulling, and at only 85% with two to six others believed pulling. It appears that virtually all of the performance decrement in rope pulling observed by Ingham et al. can be accounted for in terms of reduced effort or social loafing.

With respect to clapping and especially shouting, however, there are several possible sources of coordination loss that might have operated in addition to social loafing: (a) Sound cancellation will occur to the extent that sound pressure waves interfere with each other, (b) directional coordination losses will occur to the extent that voices are projected toward different locations, and (c) temporal coordination losses will occur to the extent that moment-to-moment individual variations in intensity are not in synchrony. Our second experiment was designed to assess the relative effects of coordination loss and social loafing in explaining the failure of group cheering to be as intense as the sum of individual noise outputs.

Experiment 2

Coordination Loss or Reduced Effort?

For Experiment 2 we arranged things so that people could not hear each other shout; participants were asked to wear headphones, and during each trial a constant 90-dB recording of six people shouting was played over the earphones, ostensibly to reduce auditory feedback and to signal each trial. As a consequence, individuals could be led to believe they were shouting in groups while actually shouting alone. Ingham et al. (1974) accomplished this through the use of "pseudosubjects," confederates who pretended to be pulling with the participants but who in fact did not pull any weight at all. That is an expensive procedure—each of the 36 participants tested by Ingham et al. required the services of 5 pseudosubjects as well as the experimenter. We were able to devise a procedure whereby, on any given trial, one person could be led to believe that he was performing in a group, while the rest thought he was performing alone. Thus, we were able to test six real participants at one time.

Additionally, although we find the interpretation offered by Ingham et al. plausible and convincing, the results of their second experiment are susceptible to an alternative explanation. When participants were not pulling the rope, they stood and watched the pseudosubjects pull. This would lead people accurately to believe that while they were pulling the rope, idle participants would be watching (Levinger, Note 1). Thus, as the number of performers decreased, the size of the audience increased. According to Cottrell's evaluation apprehension hypothesis (1972), the presence of an evaluative audience should enhance performance for a simple, well-learned task such as rope pulling, and, although there is little supportive evidence, it seems reasonable that the larger the audience, the greater the enhancement (Martens & Landers, 1969; Seta, Paulus, & Schkade, 1976). Thus, it is not clear whether there was a reduced effort put forth by group members because they believed other people were pulling with them, or an increase in the effort exerted by individuals because they believed other people were watching them. In Experiment 2, therefore, we arranged to hold the size of the audience constant, even while varying the number of people working together.

Method

Six groups of six male undergraduate volunteers heard the following instructions:

> In our experiment today we are interested in the effects of sensory feedback on the production of sound in social groups. We will ask you to produce sounds in groups of one, two, or six, and we will record the sound output on the sound-level meter that you can see up here in front. Although this is not a competition and you will not learn your scores until the end of the experiment, we would like you to make your sounds as loud as possible. Since we are interested in sensory feedback, we will ask you to wear blindfolds and earphones and, as you will see, will arrange it so that you will not be able to hear yourself as you shout.
>
> We realize it may seem strange to you to shout as loud as you can, especially since other people are around. Remember that the room is soundproofed and that people outside the room will not

be able to hear you. In addition, because you will be wearing blindfolds and headsets, the other participants will not be able to hear you or to see you. Please, therefore, feel free to let loose and really shout. As I said, we are interested in how loud you can shout, and there is no reason not to do your best. Here's your chance to really give it a try. Do you have any questions?

Once participants had donned their headsets and blindfolds, they went through a series of 13 trials, in which each person shouted four times in a group of six, once in a group of two, and once by himself. Before each trial they heard the identification letters of those people who were to shout.

Interspersed with these trials were 12 trials, two for each participant, in which the individual's headset was switched to a separate track on the stereophonic instruction tape. On these trials, everybody else was told that only the focal person should shout, but that individual was led to believe either that one other person would shout with him or that all six would shout.

Thus, each person shouted by himself, in actual groups of two and six, and in pseudogroups of two and six, with trials arranged so that each person would have approximately equal rest periods between the trials on which he performed. Each trial was preceded by the specification of who was to perform. The yells were coordinated by a tape-recorded voice counting backwards from three, followed by a constant 90-dB 5-sec recording of the sound of six people shouting. This background noise made it impossible for performers to determine whether or how loudly other people were shouting, or, for that matter, to hear themselves shout. Each trial was terminated by the sound of a bell. This sequence of 25 trials was repeated three times, for a total of 75 trials, in the course of which each subject shouted 24 times.

As in Experiment 1, the data were transformed into dynes/cm^2 and subjected to analyses of variance, with the group as the unit of analysis and each score based on the average output per person. Two separate 3×3 analyses of variance with group size (1,2,6) and trial block (1–3) were run, one on the output of trials in which groups actually shouted together, and one on the pseudogroup trials in which only one person actually shouted.

Results

Overall, participants shouted with considerably more intensity in Experiment 2 than in Experiment 1, averaging 9.22 dynes/cm^2 when shouting alone, as compared to 4.73 dynes/cm^2, $t(12) = 4.05$, $p < .01$. There are several plausible reasons for this difference. The new rationale involving the effects of reduced sensory feedback may have interested

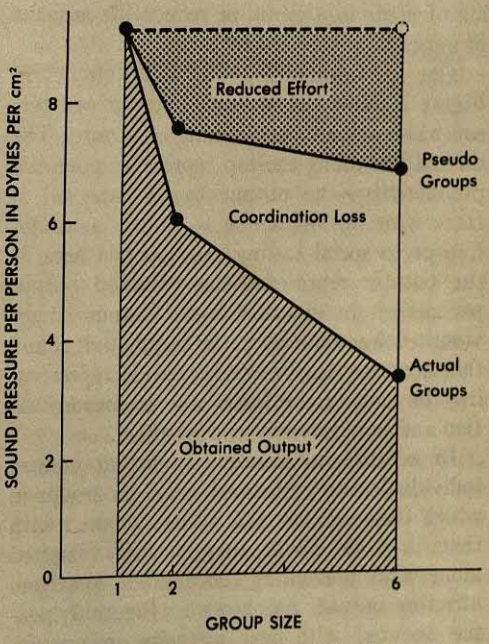


Figure 2. Intensity of sound produced per person when cheering in actual or perceived groups of 1, 2, and 6, as a result of reduced effort and faulty coordination of group efforts, Experiment 2.

or challenged individuals to perform well. The constant 90-dB background noise may have led people to shout with more intensity, just as someone listening to music through headphones will often speak inappropriately loudly (the Lombard reflex). The performers may have felt less embarrassed because the room was soundproof and the others were unable to see or hear them. Finally, through eliminating the possibility of hearing each other, individuals could no longer be influenced by the output of the others, thereby lifting the pressure of social conformity.

As in Experiment 1, as the number of actual performers increase, the total sound output also increased, but at a slower rate than would be expected from the sum of the individual outputs. Actual groups of two shouted at only 66% of capacity, and groups of six at 36%, $F(2, 10) = 226$, $p < .001$. The comparable figures for Experiment 1 are 71% and 40%. These similarities between experiments suggest that our procedural changes, even though they made people unable to hear or see each other, did not eliminate their feel-

ing of being in a group or reduce the amount of incoordination or social loafing.

The line connecting the solid circles in Figure 2 shows the decreased output per person when actually performing in groups. The dashed line along the top represents potential productivity—the output to be expected if there were no losses due to faulty coordination or to social loafing. The striped area at the bottom represents the obtained output per person in actual groups. Output is obviously lower than potential productivity, and this decrease can be considered as representing the sum of the losses due to incoordination and to reduced individual effort.

In addition to shouting in actual groups, individuals also performed in pseudogroups in which they believed that others shouted with them but in which they actually shouted alone, thus preventing coordination loss from affecting output. As shown in Figure 2, people shouted with less intensity in pseudogroups than when alone, $F(2, 10) = 37.0$, $p < .0001$. Thus, group size made a significant difference even in pseudogroups in which coordination loss is not a factor and only social loafing can operate.

When performers believed one other person was yelling, they shouted 82% as intensely as when alone, and when they believed five others to be yelling, they shouted 74% as intensely. The stippled area defined at the top of Figure 2 by the data from the pseudogroups represents the amount of loss due to social loafing. By subtraction, we can infer that the white area of Figure 2 represents the amount of loss due to faulty coordination. Since the latter comprises about the same area as the former, we can conclude that, for shouting, half the performance loss decrement is due to incoordination and half is due to social loafing.

Discussion

Despite the methodological differences between Experiments 1 and 2, both experiments showed that there is a reduction in sound pressure produced per person when people make noise in groups compared to when alone. People in Experiment 1 applauded and cheered in full view of each other, with all the excitement, embarrassment, and conformity that goes along with such a situation. In Experiment 2, no one could see or hear any other person. Only the experimenters could see the people perform. And finally, the rationale changed drastically, from the experimenters' interest in "judgments of how much noise people make in social settings" to their interest in "the effects of sensory feedback on the production of sound in social groups." Yet, despite differences in the task characteristics and supposed purpose, the two studies produced similar results. This points to the robust nature of both the phenomenon and the paradigm.

General Discussion

Noise Production as Group Performance

Although we do not usually think about it that way, making noise can be hard work, in both the physical and the psychological sense. In the present case, the participants were asked to produce sound pressure waves, either by rapidly vibrating their laryngeal membranes or by vigorously striking their hands together. Although superficially similar in consequence, this task should not be confused with more normal outbreaks of shouting and clapping that occur as spontaneous outbursts of exuberant expressiveness. Our participants shouted and clapped because we asked them to, not because they wanted to.

This effortful and fatiguing task resulted in sound pressure waves, which, although invisible, can be easily and accurately measured in physical units that are proportional to the amount of work performed. The making of noise is a useful task for the study of group processes from the standpoint both of production and of measurement—people are practiced and skilled at making noise and can do so without the help of expensive or cumbersome apparatus, and acoustics and audio engineering are sufficiently advanced to permit sophisticated data collection. We seem to have found a paradigm wherein people get involved enough to try hard and become somewhat enthusiastic, yet the task is still effortful

enough so that they loaf when given the opportunity.

The Causes of Social Loafing

The present research shows that groups can inhibit the productivity of individuals so that people reduce their exertions when it comes to shouting and clapping with others. Why does this occur? We suggest three lines of explanation, the first having to do with attribution and equity, the second with submaximal goal setting, and the third with the lessening of the contingency between individual inputs and outcomes.

1. *Attribution and equity.* It may be that participants engaged in a faulty attribution process, leading to an attempt to maintain an equitable division of labor. There are at least three aspects of the physics and psychophysics of producing sound that could have led people to believe that the other persons in their group were not working as hard or effectively as themselves. First, individuals judged their own outputs to be louder than those of the others, simply because they were closer to the sound source. Second, even if everyone worked to capacity, sound cancellation would cause group outputs to seem much less than the sum of their individual performances. Finally, the perception of the amount of sound produced in a group should be much less than the actual amount—growing only as the .67 power of the actual amount of sound, according to Stevens's psychophysical power law (1975).

These factors may have led individuals to believe that the other participants were less motivated or less skillful than themselves—in short, were shirkers or incompetents. Thus, differences in the perception of sound production that were essentially the result of physical and psychophysical processes may have been mistakenly attributed to a lack of either skill or motivation on the part of the others, leading individuals to produce less sound in groups because there is no reason to work hard in aid of shirkers or those who are less competent.

This process cannot explain the results of Experiment 2, since the capacity to judge the loudness of one's own output, much less that of others, was severely impaired by the 90-dB background masking noise used to signal the trials. However, rather than "discovering" social loafing while participating in the experiment, the participants may have arrived with the preexisting notion that people often do not pull their own weight in groups. Thus, despite being unable to hear or see one another, lack of trust and the propensity to attribute laziness or ineptitude to others could have led people to work less hard themselves.

2. *Submaximal goal setting.* It may be that despite our instructions, participants redefined the task and adopted a goal, not of making as much noise as possible, but merely of making enough noise or of matching some more or less well-defined standard. Individuals would clearly expect it to be easier to achieve this goal when others are helping, and might work less hard as a consequence. This, of course, would change the nature of noise production from what Steiner (1972) would term a *maximizing* task to an *optimizing* task. A maximizing task makes success a function of how much or how rapidly something is accomplished. For an optimizing task, however, success is a function of how closely the individual or group approximates a predetermined "best" or correct outcome. If participants in our experiments perceived sound production as an optimizing rather than a maximizing task, they might feel the optimal level of sound output could be reached more easily in groups than alone, thereby allowing them to exert less effort.

The participants in Experiment 2 could hear neither themselves nor others and would not be able to determine whether their output was obnoxious or to develop a group standard for an optimal level. Furthermore, in both experiments, the experimenters reiterated their request to yell "as loud as you can, every time," over and over again. Before the first trial they would ask the group how loud they were supposed to yell. In unison, the group would reply, "As loud as we can!" We think it unlikely that participants perceived the task to be anything other than maximizing.

3. *Lessened contingency between input and*

outcome. It may be that participants felt that the contingency between their input and the outcome was lessened when performing in groups. Individuals could "hide in the crowd" (Davis, 1969) and avoid the negative consequences of slacking off, or they may have felt "lost in the crowd" and unable to obtain their fair share of the positive consequences for working hard. Since individual scores are unidentifiable when groups perform together, people can receive neither precise credit nor appropriate blame for their performance. Only when performing alone can individual outputs be exactly evaluated and rewarded.

Let us assume that group members expect approval or other reward proportional to the total output of a group of n performers, but that since individual efforts are indistinguishable, the reward is psychologically divided equally among the participants, each getting $1/n$ units of reward. Under these assumptions, the average group, if it performed up to capacity and suffered no process loss, could expect to divide up n times the reward of the average individual, resulting in each member's getting $n \times 1/n$, or n/n, units of reward, the same amount as an individual.

Although the total amount of reward may be the same, the contingency on individual output is not. Any given individual under these assumptions will get back only one nth of his own contribution to the group; the rest will be shared by the others. Even though he may also receive unearned one nth of each other person's contribution, he will be tempted, to the extent that his own performance is costly or effortful, to become a "free rider" (Olson, 1965). Thus, under these assumptions, if his own performance cannot be individually monitored, an individual's incentive to perform should be proportional to $1/n$.

Seligman (1975) has shown that animals and people become lethargic and depressed when confronted with tasks in which they have little or no control over the outcomes. Likewise, in our experiments, people may have felt a loss of control over their fair share of the rewards when they performed in groups, leading them also to become, if not lethargic and depressed, at least less enthusiastic about making lots of noise.

Since people were asked to shout both alone and in groups, they may have felt it smart to save their strength in groups and to shout as lustily as possible when scores were individually identifiable, marshalling their energy for the occasions when they could earn rewards. This line of reasoning suggests that if inputs were made identifiable and rewards contingent on them, even when in groups, it would be impossible for performers to get a free ride and they would have an incentive to work equally hard in groups of different sizes.

Social Loafing and Social Impact Theory

Each of these three lines of explanation may be described in terms of Latané's (1973) theory of social impact. If a person is the target of social forces, increasing the number of other persons also in the target group should diminish the pressures on each individual because the impact is divided among the group members. In a group performance situation in which pressures to work come from outside the group and individual outputs are not identifiable, this division of impact should lead each individual to work less hard. Thus, whether the subject is dividing up the amount of work he thinks should be performed or whether he is dividing up the amount of reward he expects to earn with his work, he should work less hard in groups.

The theory of social impact further stipulates the form that the decrease in output should follow. Just as perceptual judgments of physical stimuli follow power functions (Stevens, 1957), so also should judgments of social stimuli, and the exponent of the psychosocial power function should have an exponent of less than one, resulting in a marginally decreasing impact of additional people. Thus, social impact theory suggests that the amount of effort expended on group tasks should decrease as an inverse power function of the number of people in the group. This implication cannot be tested in Experiment 1 or with the actual groups of Experiment 2, inasmuch as coordination loss is confounded with social loafing. However, a power function with an

exponent of −.14 accounted for 93% of the variance for the pseudogroups of Experiment 2. It appears that social impact theory provides a good account of both the existence and the magnitude of social loafing.

The Transsituational and Transcultural Generality of Social Loafing

The present research demonstrates that performance losses in groups occur with tasks other than rope pulling and with people other than prewar German workers. There are, in addition, other instances of experimental research that demonstrate similar cases of social loafing. For example, Marriott (1949) and Campbell (1952) have shown that factory workers produce less per person in larger groups than in smaller ones. Latané and Darley (1970) have found that the likelihood that a bystander will intervene in a situation in which someone requires assistance is substantially reduced by the addition of other bystanders who share in the responsibility for help. Wicker (1969) has found that the proportion of members taking part in church activities is lower in large than in small churches, presumably because the responsibility for taking part is more diffuse. Similarly, Petty, Harkins, Williams, and Latané (1977) found that people perceived themselves as exerting less cognitive effort on evaluating poems and editorials when they were among groups of other unidentifiable evaluators than when they alone were responsible for the task.

These experimental findings have demonstrated that a clear potential exists in human nature for social loafing. We suspect that the effects of social loafing have far-reaching and profound consequences both in our culture and in other cultures. For example, on collective farms (kolkhoz) in Russia, the peasants "move all over huge areas, working one field and one task one day, another field the next, having no sense of responsibility and no direct dependence on the results of their labor" (Smith, 1976, p. 281). Each peasant family is also allowed a private plot of up to an acre in size that may be worked after the responsibility to the collective is discharged. The produce of these plots, for which the peasants are individually responsible, may be used as they see fit. Although these plots occupy less than 1% of the nation's agricultural lands (about 26 million acres), they produce 27% of the total value of Soviet farm output (about $32.5 billion worth) (Yemelyanov, 1975, cited in Smith, 1976, p. 266). It is not, however, that the private sector is so highly efficient; rather, it is that the efficiency of the public sector is so low (Wädekin, 1973, p. 67).

However, before we become overly pessimistic about the potential of collective effort, we should consider the Israeli kibbutz, an example that suggests that the effects of social loafing can be circumvented. Despite the fact that kibbutzim are often located in remote and undeveloped areas on the periphery of Israel to protect the borders and develop these regions, these communes have been very successful. For example, in dairying, 1963 yields per cow on the kibbutz were 27% higher than for the rest of Israel's herds, and in 1960 yields were 75% higher than in England. In 1959, kibbutz chickens were producing 22% of the eggs with only 16% of the chickens (Leon, 1969). The kibbutz and the kolkhoz represent the range of possibilities for collective effort, and comparisons of these two types of collective enterprise may suggest conditions under which per person output would be greater in groups than individually.

Social Loafing as a Social Disease

Although some people still think science should be value free, we must confess that we think social loafing can be regarded as a kind of social disease. It is a "disease" in that it has negative consequences for individuals, social institutions, and societies. Social loafing results in a reduction in human efficiency, which leads to lowered profits and lowered benefits for all. It is "social" in that it results from the presence or actions of other people.

The "cure," however, is not to do away with groups, because despite their inefficiency, groups make possible the achievement of many goals that individuals alone could not

possibly accomplish. Collective action is a vital aspect of our lives: From time immemorial it has made possible the construction of monuments, but today it is necessary to the provision of even our food and shelter. We think the cure will come from finding ways of channeling social forces so that the group can serve as a means of intensifying individual responsibility rather than diffusing it.

Reference Note

1. Levinger, G. Personal communication, June 1976.

References

Campbell, M. Group incentive payment schemes: The effects of lack of understanding and group size. *Occupational Psychology*, 1952, *26*, 15–21.

Cottrell, N. Social facilitation. In C. McClintock (Ed.), *Experimental social psychology*. New York: Holt, Rinehart & Winston, 1972.

Dashiell, J. F. Experimental studies of the influence of social situations on the behavior of individual human adults. In C. Murchison (Ed.), *A handbook of social psychology*. Worcester, Mass.: Clark University Press, 1935.

Davis, J. H. *Group performance*. Reading, Mass.: Addison-Wesley, 1969.

Ingham, A. G., Levinger, G., Graves, J., & Peckham, V. The Ringelmann effect: Studies of group size and group performance. *Journal of Experimental Social Psychology*, 1974, *10*, 371–384.

Köhler, O. Ueber den Gruppenwirkungsgrad der menschlichen Körperarbeit und die Bedingung optimaler Kollektivkroftreaktion. *Industrielle Psychotechnik*, 1927, *4*, 209–226.

Latané, B. *A theory of social impact*. St. Louis, Mo.: Psychonomic Society, 1973.

Latané, B., & Darley, J. M. *The unresponsive bystander: Why doesn't he help?* New York: Appleton-Century-Crofts, 1970.

Leon, D. *The kibbutz: A new way of life*. London: Pergamon Press, 1969.

Marriott, R. Size of working group and output. *Occupational Psychology*, 1949, *23*, 47–57.

Martens, R., & Landers, D. M. Coaction effects on a muscular endurance task. *Research Quarterly*, 1969, *40*, 733–737.

Moede, W. Die Richtlinien der Leistungs-Psychologie. *Industrielle Psychotechnik*, 1927, *4*, 193–207.

Olson, M. *The logic of collective action: Public goods and the theory of groups*. Cambridge, Mass.: Harvard University Press, 1965.

Petty, R., Harkins, S., Williams, K., & Latané, B. The effects of group size on cognitive effort and evaluation. *Personality and Social Psychology Bulletin*, 1977, *3*, 579–582.

Seligman, M. *Helplessness*. San Francisco: Freeman, 1975.

Seta, J. J., Paulus, P. B., & Schkade, J. K. Effects of group size and proximity under cooperative and competitive conditions. *Journal of Personality and Social Psychology*, 1976, *34*, 47–53.

Smith, H. *The Russians*. New York: Ballantine Books, 1976.

Steiner, I. D. Models for inferring relationships between group size and potential group productivity. *Behavioral Science*, 1966, *11*, 273–283.

Steiner, I. D. *Group process and productivity*. New York: Academic Press, 1972.

Stevens, S. S. On the psychological law. *Psychological Review*, 1957, *64*, 153–181.

Stevens, S. S. A metric for the social consensus. *Science*, 1966, *151*, 530–541.

Stevens, S. S. *Psychophysics: Introduction to its perceptual, neural and social prospects*. New York: Wiley, 1975.

Wädekin, K. *The private sector in Soviet agriculture*. Los Angeles: University of California Press, 1973.

Wicker, A. N. Size of church membership and members support of church behavior settings. *Journal of Personality and Social Psychology*, 1969, *13*, 278–288.

Zajonc, R. B. Social facilitation. *Science*, 1965, *149*, 269–274.

Zajonc, R. B. *Social psychology: An experimental approach*. Belmont, Calif.: Brooks/Cole, 1966.

Received June 23, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 6, 833-846

Template Matching: A Proposal for Probing the Ecological Validity of Experimental Settings in Social Psychology

Daryl J. Bem
Cornell University

Charles G. Lord
Stanford University

The question of ecological validity—Do people behave in real life as they behave in our laboratories?—is central to experimental social psychology. The template-matching technique is introduced as a way to explore such validity. First, each laboratory behavior is characterized by a *template*, a personality description (Q sort) of the hypothetical ideal person most likely to display that behavior. These templates characterize subjects' potential behaviors inside the laboratory. Next, Q-sort personality descriptions of the subjects are obtained from their peers; these characterize the subjects' behaviors outside the laboratory. Evidence for the ecological validity of the laboratory setting is then sought in the match between these two sets of data. Using this technique, support was found for the ecological validity of the mixed-motive game, a widely used laboratory paradigm whose validity is often questioned. The heuristic use of Q-sort data to draw clinical portraits of the subjects is also illustrated.

Do people behave in real life as they behave in our experimental laboratories? This question, the question of ecological validity, is central to experimental social psychology, and the possibility that the answer might be *no* is central to the field's recent crisis of confidence (e.g., McGuire, 1973; Ring, 1967). Not surprisingly, the strongest doubts are provoked by the field's more dramatic experiments: Is there really a plausible analogue between a subject's obediently delivering shocks to a fellow subject in Milgram's laboratory and an Eichmann's obediently delivering Jews to their deaths in Nazi Germany (Baumrind, 1964; Milgram, 1974; Orne & Holland, 1968)? Does a Zimbardo prison with college students playing the roles of guards and prisoners really capture any of the functional properties of actual incarceration (Banuazizi & Movahedi, 1975; Haney, Banks, & Zimbardo, 1973)? But even the field's more mundane experimental paradigms must bear the same burden of proof. Do the cooperative and competitive choices of subjects in the Prisoner's Dilemma game, for example, have any parallels with the cooperative and competitive choices of individuals in the real world—let alone the foreign policy choices of nation states? The concern is not new: Even well before the current crisis, Carl Hovland was agonizing over the fact that laboratory and field studies of persuasive communication produced widely divergent results (Hovland, 1959).

Investigators have usually attempted to establish the external validity of their laboratory procedures in two basic ways. The first strategy, often proposed but rarely performed, is to conduct conceptually parallel laboratory experiments and field studies in tandem, in order to confirm the assumed

The authors are grateful to Spencer Kagan and his colleague George Knight for providing template weights. They also wish to thank their own colleagues Craig Anderson, Marjorie Horton, Dennis Jennings, David Maxfield, and Ken Monteiro for providing template weights. For assistance in conducting the study, they would like to thank Greg Eger, Pat Flinn, and Rod Hubbard.

All Q-sort materials used in this research are available from Consulting Psychologists Press, 577 College Avenue, Palo Alto, California 94306.

Requests for reprints should be sent to Daryl J. Bem, Department of Psychology, Uris Hall, Cornell University, Ithaca, New York 14853.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3706-0833$00.75

isomorphism directly. The second strategy is to demonstrate the construct validity of the laboratory-assessed concepts, to show that the behavior relates to other variables as the theory says it should. For example, one's faith in the ring-toss game as a situation that elicits achievement motivation is bolstered by findings that individuals high on the motive stand a moderate distance from the peg as the theory of achievement motivation implies (McClelland, 1958). Similarly, one's confidence in a paradigm's external validity is enhanced if individual difference variables relate to the laboratory behaviors in theoretically compelling ways.

The Template-Matching Technique

In this article, we propose a novel extension of the individual-differences approach to assessing the degree to which subjects' behaviors in the laboratory match their behaviors in everyday life. It is called the template-matching technique, a general method introduced by Bem and Funder (1978) for characterizing situations in the language of personality description. Applied to the problem of ecological validity, the technique consists of two basic steps. First, each conceptually distinct behavior in the laboratory setting under investigation is characterized by a *template*, a personality description of the hypothetical ideal person who is most likely to display the designated behavior in that situation. These templates serve to characterize the potential behaviors of subjects inside the laboratory. Second, personality descriptions of the subjects are obtained from close acquaintances; these descriptions characterize the subjects' behaviors outside the laboratory. The ecological validity of the laboratory setting, then, is sought in the match between these two sets of data. Specifically, the model predicts that the description of the individual's personality will match most closely the template that characterizes the behavior he or she actually displays in the laboratory. A complete discussion of the procedure is provided by Bem and Funder (1978), who have applied it to the delay-of-gratification situation and the forced-compliance experiment, demonstrating how it can be used to pinpoint sources of cross-situational inconsistency and to test rival theories of psychological phenomena.

The Mixed-Motive Game

To illustrate the utility of the template-matching technique for addressing the issue of ecological validity, we have chosen to apply it to one of the most frequently used—and frequently criticized—experimental paradigms in social psychology, the mixed-motive game (e.g., the Prisoner's Dilemma game). It will be recalled that in this procedure, two subjects are required simultaneously to make a series of decisions independently of each other and with no communication between them. The payoff to each subject is contingent on both of their choices. Thus in the Prisoner's Dilemma game, the most familiar of the mixed-motive games, when both subjects select the *cooperative* response, their joint payoff is high; when both select the *competitive* response, their joint payoff is low; when one subject chooses the cooperative response and the other the competitive response, the latter (the *defector*) earns points or money at the expense of the cooperator. The mixed-motive game is a particularly appropriate test case here, not only because it has been extensively used, but also because the debate between its proponents (e.g., Kelley et al., 1970) and its critics (e.g., Knox & Douglas, 1971; Nemeth, 1972; Pruitt, 1967) nicely epitomizes the field's more general concern with the issue of ecological validity.

Most of the studies using the mixed-motive game are concerned not with its external validity but with the internal structure of the situation, the ways in which a subject's playing strategy varies as a function of the payoff matrix, communication variables, and preprogrammed features of the opponent's strategy. (See, for example, reviews by Nemeth, 1972, and Oskamp, 1971. See also Kuhlman & Marshello, 1975.) Such studies yield an internally consistent picture of cooperation and competition within the paradigm but provide very little in the way of independent evidence for its ecological validity.

There appear to be no studies that directly compare a laboratory experiment and a field

situation, and most of the studies that do bear upon the issue of external validity are those that explore group or individual differences. For example, army cadets are more competitive in mixed-motive games than Berkeley students (Pilisuk, Kiritz, & Clampitt, 1971); individuals high on the need for power are more competitive than those low on this motive (Terhune, 1968); isolationists are more competitive than internationalists (McClintock, Harrison, Strand, & Gallo, 1963); high authoritarians are more competitive than low authoritarians (Wrightsman, 1966); externals are more competitive than internals (Tiller, 1970). And, in a study that goes beyond the simple cooperative–competitive dichotomy by utilizing the *decomposed* versions of mixed-motive games introduced by Messick and McClintock (1968), it was found that subjects who systematically attempted to maximize the joint gain of both players (called J's) scored higher on the Affiliation and Intraception scales but lower on the Aggression scale of the Edwards Personal Preference Schedule (Edwards, 1959) than did subjects who attempted to maximize their own gain (O's) or subjects who attempted to maximize their gain relative to the other person (R's) (Bennett & Carbonari, 1976).

In the study reported here, we also used the decomposed mixed-motive games. Specifically, we constructed templates of three hypothetical persons: the ideal J, the ideal O, and the ideal R. Each subject in the actual experiment was then classified on the basis of his or her responses as a J, an O, or an R, and it was predicted that the subject's similarity to the corresponding template would be higher than his or her similarity to the templates characterizing the two alternative strategies.

Method

The Mixed-Motive Game and Its Decomposition

In a mixed-motive game, two subjects are isolated from one another and required simultaneously to select from Alternatives A and B in the payoff matrix shown in Table 1. This matrix is an example of the Prisoner's Dilemma game in which joint cooperation (A responses) earns the players the maximum joint payoff, joint competition (B responses) yields minimum joint payoff, and *defection* to the competitive response (B) earns the defector maximum payoff when the other player selects the cooperative response (A).

Table 1
Payoff Matrix for the Prisoner's Dilemma Game

Other person's choice	Your choice: A		Your choice: B	
A	You earn Other earns	100 100	You earn Other earns	110 70
B	You earn Other earns	70 110	You earn Other earns	80 80

Every mixed-motive game can also be decomposed into a situation in which the subject makes his or her own choices without direct simultaneous reference to the other subject's behavior. For example, the full-matrix mixed-motive game decomposes into the following decision alternatives for such a subject:

	Your choice: A	Your choice: B
You earn	60	70
Other earns	40	10

Here the subject chooses between A and B while being told that the other subject is making the same choice concurrently. Note that if the subject chooses Alternative A, he or she assigns 60 points to the self and 40 points to the other subject; if he or she chooses Alternative B, 70 points are assigned to the self and 10 points to the other subject. The game is isomorphic to the full-matrix version because if both subjects select Alternative A, each earns 100 points; if both select B, each earns 80 points; and if one selects A and the other selects B, the former earns 70 points and the latter earns 110 points. Thus the two games are formally equivalent, but the decomposed game is played without trial-by-trial feedback, and it focuses attention on each player's individual strategy, unconfounded by two-person interactions. Note, for example, that a subject who wishes to maximize the joint outcome of both players will choose Alternative A, whereas a subject who wishes to maximize his or her own payoff will select Alternative B. The elegance of this feature is best illustrated by the Triple Dominance game in which subjects choose from among three alternatives:

	Your choice: A	Your choice: B	Your choice: C
You earn	40	50	40
Other earns	40	20	0

Note that each alternative satisfies one, and only one, of the three strategies (J, O, or R) discussed earlier. The subject who selects Alternative A assigns 40 points to the self and 40 points to the other person, thus maximizing the *joint* return for both players (80 points versus 70 or 40 points for the other alternatives); the subject who chooses Alternative B assigns 50 points to the self, thereby maximizing his or her *own* return; and the subject who chooses Alternative C assigns 40 points to the self and 0 points to the other person, maximizing his or her return *relative* to the other person (40 points versus 0 or 30 points).

The Prisoner's Dilemma game and the Triple Dominance game are but two of the possible mixed-motive games that can be used. (See, for example, Kuhlman & Marshello, 1975.)

Constructing the Templates

To operationalize the template concept, Bem and Funder (1978) employ the Q-sort technique, utilizing a modification of the California Q set of items devised by Block (1961/1978). The Q set consists of 100 descriptive personality statements (e.g., "behaves in a giving way toward others"), which are sorted by an assessor into nine categories ranging from the least to the most characteristic of the person being described; each item thus receives a score from 1 to 9. Two Q sorts can be compared with one another by simply computing a Pearson product-moment correlation across the items, thus expressing directly and quantitatively the degree of similarity between the two profiles. Similarly, if one has constructed a Q sort of some hypothetical ideal personality (our template), one can correlate the Q sorts of actual individuals with this idealized sort in order to assess their similarities to the ideal type.[1]

As noted above, the Q sort of an individual is obtained by having an assessor actually sort the 100 items into nine categories. Constructing a template, however, requires a different method for assigning numerical values to the items because only a small number of items will be relevant to any particular template. Accordingly, Bem and Funder (1978) constructed templates by "adjusting" the composite Q sort of the entire subject sample according to the following formula:

$$Q_i = M_i + w_i\sigma_i$$

where Q_i is the adjusted or template value of the ith item, M_i and σ_i are the mean and standard deviation, respectively, of the ith item for the subject sample as a whole, and w_i is a weighting factor that reflects the relevance of the ith item to the criterion behavior being characterized.

Suppose, for example, that we wish to construct a template of the person who pursues the J strategy, the strategy that maximizes the joint return of both players. The Q item "is physically attractive" is conceptually irrelevant to this behavior, and accordingly, it is assigned a relevance weight (w_i) of zero. The adjusted value of this item in the template, then, simply becomes the mean of the item in the subject sample. But now consider the item "behaves in a giving way toward others," which would seem relevant to the behavior implied by the J strategy. If we assign this a weight of, say, +2, then the template value of this item will be adjusted upward from the sample mean by two standard deviations, and—the model predicts—the resulting template will be more similar to the Q sorts of individuals who pursue the J strategy than to the Q sorts of individuals pursuing the alternative strategies. A complete discussion of this template-construction algorithm will be found in Bem and Funder (1978).

In the present study, we constructed the templates by recruiting five graduate students in psychology to serve as judges. They were supplied with a written description of the three strategies along with the illustrative Triple Dominance game shown above and the complete list of Q-sort items. Each judge went through the list and gave an integer rating to each item, ranging from −2 ("very uncharacteristic of a person following this strategy") through 0 ("not relevant; neither characteristic nor uncharacteristic of the person following this strategy") to +2 ("very characteristic of the person following this strategy"). For example, the item "behaves in a giving way toward others" received a mean rating of +1.6 for the J strategy, and the item "is power oriented; values power in self and others" received a mean rating of +1.2 for the R strategy, the strategy in which subjects attempt to maximize their earnings relative to the other person. The judges were explicitly directed not to play personality theorist, not to speculate about remote personality correlates of the behavior, but rather to judge only whether an item might or might not characterize the behavior involved in the strategy itself. The judges completed this procedure for each of the three strategies. The mean rating given each item by the five judges then became the relevance weight (w_i) for that item in the template-generating algorithm described previously. Separate templates were constructed for each sex by using the same set of weights (w_is) but the separate means and standard deviations (M_is and σ_is) for each sex. This procedure, then, yielded a set of templates for the

[1] Even though this index of similarity is computed like a product-moment correlation, it cannot be treated like one statistically. For example, there is no reason to expect the correlation between any two Q sorts selected at random to have an expected value of zero. Moreover, the magnitude of the correlation is a function of factors other than the component of template-specific similarity in which we are interested. (See Bem & Funder, 1978.) To avoid any confusion on this issue, we routinely convert these similarity scores to standardized T scores with means of 50 and standard deviations of 10.

situation, each characterizing one of the three "pure" strategies that could be pursued.

For purposes of comparison, a second set of templates was constructed by enlisting the aid of Spencer Kagan, an investigator who has worked extensively with the decomposed mixed-motive games and who, in fact, distinguishes not 3, but 16 distinct strategies in his own work (Kagan, 1972, 1977). Moreover, because Kagan is trained as a clinician, we asked him to supply a set of weights (w_is) for the three strategies by drawing as heavily as he wished on his accumulated experience and clinical intuitions, instructions quite different from the conservative and restrictive guidelines we imposed upon our graduate-student judges.

The Experiment

Forty-eight male and 44 female Stanford University students who had previously provided self Q sorts in an unrelated setting were recruited to participate in a 1-hr "decision-making" experiment. For their participation, they received experimental credit points in introductory psychology as well as the money they earned during the experiment. In the experimental session, four subjects of the same sex were seated in separate experimental booths and told that they would be anonymously paired with each of the other subjects in turn and asked to make a series of decisions. In fact, the subjects did not actually play against one another; rather, they played simultaneously against a preprogrammed sequence of decisions throughout the experiment. They then made choices on 24 decomposed mixed-motive games with no feedback concerning the "other subject's" decisions. Each payoff matrix was projected on a screen which could be seen by all subjects from their separate booths. They then played 48 full-matrix mixed-motive games (allegedly 16 with each of the other three subjects in turn) with the usual trial-by-trial feedback about the joint payoff. After the session, they were debriefed and paid $2, the amount they would have earned with optimal play and maximally cooperative opponents.

During the following academic quarter, we contacted 10 subjects who had systematically adopted the J strategy, 10 subjects who had adopted the O strategy, and the entire sample of 10 subjects who had adopted the R strategy during the experiment. Each group contained 5 males and 5 females. All these subjects gave us permission to obtain confidential Q sort descriptions of them from one of their roommates. The roommates came to our laboratory and were paid $2 to complete the Q sorts; they were not told about the subjects' participation in our experiment the previous quarter. (Since students in introductory psychology participate in several studies, it is unlikely that these roommates knew about the experiment or its relation to the information they were providing.) Finally, the experimenters in the several phases of the study were always blind with respect to all other information about the subjects and the study.

Results and Discussion

The first step in the analysis of the data was to classify each subject as a J, an O, or an R, according to the criterion that he or she must have selected 50% or more of the choices compatible with that strategy on each of the four types of decomposed games. Under this criterion, we classified 36% of our subjects as J's, 39% as O's, and 11% as R's, leaving 14% unclassified. This grouping contains relatively fewer R's and unclassified subjects than reported by Kuhlman and Marshello (1975), who used the same classification criterion.

We next confirmed that behavior on the decomposed games does, in fact, predict behavior on the more familiar full-matrix games employed in most previous studies. For example, those who adopted the strategy of trying to maximize the joint return of both themselves and the other person in the decomposed games (J's) selected the cooperative response on full-matrix games 71% of the time; those who adopted the strategy of maximizing their own return (O's) selected the cooperative response 57% of the time; and those who adopted the strategy of trying to maximize their return relative to the other subject (R's) selected the cooperative response 31% of the time. This result is highly significant, $F = 20.47$, $p < .001$, with no overlap between the 95% confidence intervals for the three groups. Clearly, then, we can generalize from the decomposed games to the full-matrix games, a fact that has not previously been established using a variety of full-matrix mixed-motive games (cf. Kuhlman & Marshello, 1975).

Assessing Ecological Validity

As we noted earlier, the roommates' Q sorts of our subjects characterize their behaviors outside the laboratory, and the templates characterize their behaviors within the laboratory. Accordingly, the ecological validity of the laboratory situation is sought in the match between these two sets of data. Specifically, we predict that each subject's roommate Q sort will be more similar to the template characterizing his or her adopted

strategy in the laboratory than to the templates characterizing the two alternative strategies.

To test this hypothesis, the subjects' roommate Q sorts were correlated with each of the three sex-appropriate templates in turn; and for each template, these correlations were then converted to standardized T scores with means of 50 and standard deviations of 10 so that all templates would have the same baseline, that is, would equally characterize the average or composite subject. (Also see Footnote 1.) All analyses were then conducted on these standardized template-similarity scores.

The results show that our subjects' roommate Q sorts are, in fact, more similar to the templates characterizing their actual playing strategies than they are to the alternative templates (52.2 vs. 48.9), $t(29) = 2.33$, $p < .025$, one-tailed. Moreover, this effect is consistent across subjects: 22 of the 30 difference scores were positive, $z = 2.37$, $p = .009$, by a one-tailed sign test. The templates generated by our clinical expert, Spencer Kagan, were similarly successful (52.5 vs. 48.8), $t(29) = 2.63$, $p < .01$, one-tailed, with 21 of the 30 difference scores positive, $z = 2.01$, $p = .022$, one-tailed. The template-matching technique has thus provided one kind of support for the ecological validity of mixed-motive games as analogues to real-world situations. As seen by their peers, our subjects behave outside the laboratory as we saw them behave within it. The relationship between this kind of evidence and the full assessment of ecological validity will be discussed later.[2]

Clinical Portraits of the Subjects

Although it is the template-matching techniques that provide the formal mechanism for testing hypotheses about situations, it is the Q-sort information that has the most heuristic potential for revealing the functional properties of the setting and probing the meaning of the behaviors observed (Bem & Funder, 1978). In particular, Q sorts draw clinical portraits of the subjects and identify those attributes that discriminate one type of subject from another. The following descriptions are based upon the data presented in detail in the Appendix (Table A1). On each Q item, each subject type is contrasted with its same-sex complement (e.g., male J's vs. male O's and R's combined), using the error term from the one-way analysis of variance on the item. This is done separately for the roommate Q sorts and the self-sorts of the subjects. Within each subject type, the item is also tested to see if the roommates and the subjects disagree significantly.

Again, it is the heuristic value of this analysis that we wish to emphasize; the statistical tests should be used only as rough guidelines. First, we conducted a separate analysis on each item; some of the significant contrasts may have emerged by chance. Second, the contrasts comparing each subject type against its complement are not orthogonal; if a group is significantly higher than its complement on an item, one or both of the other types are likely to show up, nonindependently, as significantly lower than their complements on that same item. And finally, even though a subject type might be significantly higher than its complement on a particular item, this does not necessarily mean that the item is descriptive of the type in absolute terms. (The absolute standing of an item can be inferred roughly from Table A1, which lists the items in rank order for the sample as a whole, based on the roommates' sorts. Thus the top items are those that are most descriptive of the subject sample; the bottom items are those least descriptive of the sample.)

With these caveats in mind, we can venture into the clinical thicket of Table A1. Since the complete data base is provided, the reader is free to agree or disagree with our interpretations and to venture more or less boldly than we shall do.

The J's. The J strategy is the cooperative or even altruistic strategy that maximizes the players' joint return as well as the other player's return. Relative to the complement,

---

[2] No significant results are obtained if the self-sorts rather than the roommates' sorts are used in the template-matching analysis. We shall see the reason for this in what follows.

the male J is described by his roommate as thinking in unusual ways, being repressive, not arousing liking and acceptance in people nor having social poise, not being gregarious, and not being particularly interesting. Relative to the complement, the male J describes himself as being introspective, concerned with philosophical problems, and, feeling a lack of personal meaning in life, and he agrees with his roommate that he thinks and associates ideas in unusual ways. He also sees himself as having relatively little social poise and as being less masculine and less power oriented than the complement males consider themselves to be. He doesn't think he is as repressive as his roommate does, however. These various findings are reminiscent of the Bennett and Carbonari (1976) finding that J's score higher on the Intraception scale but lower on the Aggression scale of the Edwards Personal Preference Schedule than the complement groups do and lower on the Heterosexual scale than the R's do.

The female J looks a bit more rigid than the male J. She is described by her roommate as being concerned with philosophical problems, but she is also seen as moralistic, fastidious, fussy about minor things, and having persistent, preoccupying thoughts. She describes herself as being basically distrustful of people in general, but this is only in contrast to the complement groups' self-ratings; the item still falls well into the uncharacteristic end of her self Q sort. In general, her roommate sees her less positively than she sees herself, rating her significantly higher on such items as "delays gratification unnecessarily," "concerned with own bodily functions," "thin-skinned; sensitive to criticism," "would be disorganized under stress," and "has persistent, preoccupying thoughts." Her roommate rates her significantly lower on the items "is cheerful," "enjoys aesthetic impressions," and "behaves in an assertive fashion."

In general, the J's turned out to be more withdrawn and less charming than the warm, outgoing, generous people our judges had anticipated. Their generosity in the mixed-motive games seems to spring more from an ethical or moral position—and a rather rigid one among the women—than from a spontaneous love of people. Thus, the J templates, while still characterizing the J's better than the O's or the R's, tended to err on the side of being too positive about this group. Indeed, the item "behaves in a giving way toward others," the assumed sine qua non of the J strategy, does not discriminate the J's from the other groups.

The O's. The O strategy attempts to maximize the individual's own earnings irrespective of the other player's outcomes. This was the modal strategy, being adopted by 39% of the subject sample. It also turned out to be the normative strategy in the sense that it conveys the least amount of individuating information about those who adopt it. Although there are several discriminating items in the Q sorts of the O's, they do not form any obvious clusters, nor do any of them seem particularly related to the O strategy as such. Rather, it appears that the J's and the R's define the distinctive groups in the study, with the O's constituting the residual group. The templates were least successful at capturing this group, tending to describe them as much more selfish and self-centered than they turned out to be. In fact, the O's would appear to be the roommates of choice, particularly the women, who are described as cheerful, verbally fluent, aesthetically sensitive, interpersonally consistent, and not particularly thin-skinned or fussy.

The R's. The R strategy is the most competitive in that it manages to maximize the difference between the individual's return and the other player's return, as well as to minimize the absolute return for the other player. This was the most deviant strategy in our Stanford sample of 92 subjects, and our subsample of 30 on whom we have roommate sorts contains the entire set of 5 men and 5 women who adopted it. They are deviant in their personal attributes as well.

Consider the males first. Being competitive is, of course, sex-role appropriate for men in our culture, and the male R seems to epitomize the stereotypic extraverted American male. He is assertive, masculine, socially poised, conservative, conventional, and somewhat condescending. He is not introspective, guilt ridden, anxious, or plagued by doubts

about the meaning of life. He describes himself as dependable, straightforward, and candid. (Bennett & Carbonari, 1976, found the R's to be higher on Aggression and Heterosexuality than the J's were, but not significantly different from the O's.) Such a picture may not be deviant in the larger culture, but the male R's are clearly in the minority among Stanford males, and this may explain why our study uncovered fewer R's than reported by other investigators.

The female R's are the most interesting group of all. Here is a group of women who pursue a strategy quite at variance with the sex-role stereotype of the female, and yet they describe themselves as quintessentially feminine: feminine, physically attractive, socially poised, arousing liking and acceptance in others, not condescending, and not distrustful. Their roommates do not agree, finding them instead to be aloof from people, power oriented, negativistic, tending to undermine, obstruct or sabotage, not sympathetic, considerate, or giving, and not relating to everyone in the same way. Compared to the female R's self-image, her roommate sees her as hostile, distrustful, critical, histrionic, not cheerful, not gregarious, and as having little or no insight into her own motives and behavior. The female R's and their roommates disagree significantly on 25 of the 100 Q-sort items, representing 30% of all the roommate–self discrepancies observed among the six groups. This is a proportion almost twice that expected by chance, $z = 3.22$, $p < .002$, two-tailed. All in all, the female R's present a rather chilling portrait. It is also clear why the self-sorts did not function well in the template-matching procedure. In the laboratory, we saw the women their roommates saw.

Q Sorts, Template Matching, and Ecological Validity

In our view, the major payoff from the present investigation resides in the data-rich Table A1—however unwieldy and statistically unaesthetic it may be. We believe that investigators ought to collect Q-sort data routinely even when they have no interest in individual differences per se, particularly when a new laboratory paradigm is first being developed. By examining the Q items that correlate with the behavior under investigation, one can learn an enormous amount about the subjects and the experimental setting simultaneously. (See Bem & Funder, 1978.) Although we advocate collecting Q-sort data from peers as well as from the subjects themselves, the more easily obtained self-sorts would probably suffice for many purposes. Even the ranking of Q-sort items that characterizes a subject sample as a whole provides valuable information. For example, when an experiment fails to replicate, suspicions often center upon subject-sample differences, but the sparse demographic information usually available on subject samples provides little help in assessing this possibility. In contrast, composite Q sorts of the subject samples involved would provide instant leads about the possible sources of different results patterns across seemingly identical investigations. In the present study, for example, it is clear from the composite Q sort of the sample as a whole (summarized by Table A1) that we should expect to see few R's in our mixed-motive games; Stanford students, it appears, are a rather saintly group. In the "context of discovery," there is no match for Q-sort data.[3]

The template-matching technique itself extends the heuristic utility of Q-sort data to the "context of justification" as well. In the present study, we constructed our templates on the basis of observers' ratings of the laboratory setting under investigation. We then demonstrated support for the ecological validity of mixed-motive games by showing that these templates matched peer Q-sort descriptions of the subjects' real-life behaviors outside the laboratory. Bem and Funder (1978) have demonstrated that the templates can also be constructed post hoc on the basis of the data and then cross-validated; they used

[3] The descriptive use of Q-sort data did not originate with us; it has been employed in research programs by Block and his colleagues for many years (e.g., Block, 1977; Block & Petersen, 1955). Gough (1960, 1965) makes similar use of the Adjective Check List. In our view this technique has neither received the attention it has deserved nor has its heuristic value been fully appreciated.

this variation of the technique to probe the functional properties of a delay-of-gratification situation. In a second study, they showed how templates can also be derived from formal theories, employing this variation to test rival explanations of the forced-compliance experiment against one another. The template-matching technique is thus a very general method that enables investigators to display their virtuosity at characterizing situations, and to do so in a formally legitimate way, that is, within the context of justification.

Even so, however, it is important in the present study to recognize just what successful template matching does and does not imply about the ecological validity of an experimental setting. As we have noted, it does demonstrate that the subjects are behaving in the setting as they behave in the world outside the laboratory. But this is a weak form of ecological validity. As an anonymous reviewer of this article correctly noted, the concept of ecological validity requires that the *relationships* between situational variables and the behavior in the setting replicate the *relationships* between situational variables and the behavior outside the laboratory. For example, researchers who use mixed-motive games are typically interested in how variables such as the payoff matrices, communication opportunities, and features of the opponent's playing strategies affect the subject's choices. For the game to have ecological validity in the strong sense, it must be demonstrated that these kinds of variables affect behavior in similar ways in the game's real-world counterparts. Template matching does not speak to this issue. Similarly, Hovland's (1959) lament about the divergence between laboratory and field studies of persuasion was that source and message variables did not seem to show similar effects or interactions in the two spheres. Hovland would not have been much comforted to learn that a person who is easily persuaded in the laboratory is also easily persuaded in the field.

Successful template matching, then, would appear to be a necessary but not sufficient condition for establishing the full ecological validity of a laboratory setting. Nevertheless, if the template-matching technique—or some equivalent procedure—were to be employed more widely, then we might have more confidence that people are behaving in our laboratories as they behave in real life.

References

Banuazizi, A., & Movahedi, S. Interpersonal dynamics in a simulated prison: A methodological analysis. *American Psychologist*, 1975, *30*, 152–160.

Baumrind, D. Some thoughts on ethics of research: After reading Milgram's "Behavioral study of obedience." *American Psychologist*, 1964, *19*, 421–423.

Bem, D. J., & Funder, D. C. Predicting more of the people more of the time: Assessing the personality of situations. *Psychological Review*, 1978, *85*, 485–501.

Bennett, R. P., & Carbonari, J. P. Personality patterns related to own-, joint-, and relative-gain maximizing behaviors. *Journal of Personality and Social Psychology*, 1976, *34*, 1127–1134.

Block, J. *The Q-Sort method in personality assessment and psychiatric research*. Palo Alto, Calif.: Consulting Psychologists Press, 1978. (Originally published, 1961.)

Block, J. Advancing the psychology of personality: Paradigmatic shift or improving the quality of research? In D. Magnusson & N. S. Endler (Eds.), *Personality at the crossroads: Current issues in interactional psychology*. Hillsdale, N.J.: Erlbaum, 1977.

Block, J., & Petersen, P. Some personality correlates of confidence, caution, and speed in a decision situation. *Journal of Abnormal and Social Psychology*, 1955, *5*, 34–41.

Edwards, A. L. *Edwards Personal Preference Schedule manual*. New York: Psychological Corporation, 1959.

Gough, H. G. The adjective check list as a personality assessment research technique. *Psychological Reports*, 1960, *6*, 107–122.

Gough, H. G. Conceptual analysis of psychological test scores and other diagnostic variables. *Journal of Abnormal Psychology*, 1965, *70*, 294–302.

Haney, C., Banks, B., & Zimbardo, P. Interpersonal dynamics in a simulated prison. *International Journal of Criminology and Penology*, 1973, *1*, 69–97.

Hovland, C. I. Reconciling conflicting results derived from experimental and survey studies of attitude change. *American Psychologist*, 1959, *14*, 8–17.

Kagan, S. Social motives and behaviors of Mexican-American and Anglo-American children. In J. L. Martinez (Ed.), *Chicano psychology*. New York: Academic Press, 1977.

Kagan, S., & Madsen, M. Rivalry in Anglo-American and Mexican children of two ages. *Journal of Personality and Social Psychology*, 1972, *24*, 214–220.

Kelley, H. H., Shure, G. H., Deutsch, M., Faucheux, C., Lanzetta, J. T., Moscovici, S., Nuttin, J. M., Rabbie, J. M., & Thibaut, J. W. A comparative experimental study of negotiation behavior. *Journal of Personality and Social Psychology*, 1970, *16*, 411–438.

Knox, R. E., & Douglas, R. L. Trivial incentives, marginal comprehension, and dubious generalizations from Prisoner's Dilemma studies. *Journal of Personality and Social Psychology*, 1971, *20*, 160–165.

Kuhlman, D. M., & Marshello, A. F. J. Individual differences in game motivation as moderators of preprogrammed strategy effects in Prisoner's Dilemma. *Journal of Personality and Social Psychology*, 1975, *32*, 922–931.

McClelland, D. C. Risk-taking in children with high and low need for achievement. In J. W. Atkinson (Ed.), *Motives in fantasy, action and society*. Princeton, N.J.: Van Nostrand, 1958.

McClintock, C. G., Harrison, A. A., Strand, S., & Gallo, P. Internationalism–isolationism, strategy of the other player, and two-person game behavior. *Journal of Abnormal and Social Psychology*, 1963, *67*, 631–636.

McGuire, W. J. The yin and yang of progress in social psychology: Seven koan. *Journal of Personality and Social Psychology*, 1973, *26*, 446–456.

Messick, D. M., & McClintock, C. G. Motivational basis of choice in experimental games. *Journal of Experimental Social Psychology*, 1968, *4*, 1–25.

Milgram, S. *Obedience to authority*. New York: Harper & Row, 1974.

Nemeth, C. A critical analysis of research utilizing the Prisoner's Dilemma paradigm for the study of bargaining. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6). New York: Academic Press, 1972.

Orne, M. T., & Holland, C. H. On the ecological validity of laboratory deceptions. *International Journal of Psychiatry*, 1968, *6*, 282–293.

Oskamp, S. Effects of programmed strategies on cooperation in the prisoner's dilemma and other mixed motive games. *Journal of Conflict Resolution*, 1971, *15*, 225–259.

Pilisuk, M., Kiritz, S., & Clampitt, S. Undoing deadlocks of distrust: Hip Berkeley students and the R.O.T.C. *Journal of Conflict Resolution*, 1971, *15*, 81–95.

Pruitt, D. G. Reward structure and cooperation: The decomposed Prisoner's Dilemma game. *Journal of Personality and Social Psychology*, 1967, *7*, 21–27.

Ring, K. Experimental social psychology: Some sober questions about some frivolous values. *Journal of Experimental Social Psychology*, 1967, *3*, 113–123.

Terhune, K. W. Motives, situation, and interpersonal conflict within Prisoner's Dilemma. *Journal of Personality and Social Psychology Monograph*, 1968, *8*(3, Pt. 2).

Tiller, E. H. Cooperative behavior in a Prisoner's Dilemma game as a function of selected reinforcement schedules and personality variables (Doctoral dissertation, University of Tennessee, 1970). *Dissertation Abstracts International*, 1970, *31*, 5646B. (University Microfilms No. 71–7685)

Wrightsman, L. S. Personality and attitudinal correlates of trusting and trustworthy behaviors in a two-person game. *Journal of Personality and Social Psychology*, 1966, *4*, 328–332.

Appendix

Table A1
Q-Item Differences Between Each Mixed-Motive Game Subgroup and its Same-Sex Complement and Between Self and Roommate Ratings Within Each Subgroup

Item	J subgroup						O subgroup						R subgroup					
	Male			Female			Male			Female			Male			Female		
	S	R	D	S	R	D	S	R	D	S	R	D	S	R	D	S	R	D
Has high degree of intellectual capacity					+												−	(S)
Values own independence and autonomy																		
Responds to humor																		
Interested in members of opposite sex							(−)			(−)	−							
Arouses liking and acceptance in people		−														(+)	(−)	
Genuinely dependable and responsible							(−)						+		(S)			
Verbally fluent; can express ideas well					(−)						+	(R)						
Is productive; gets things done			(R)															
Is calm, relaxed in manner										+								
Enjoys aesthetic impressions						S					+				R			S
Behaves in ethically consistent manner																		
Physically attractive; good-looking										(−)		R						
Is cheerful						S					+							S
Has social poise; appears socially at ease	−	−												+	(R)	(+)		
Behaves in masculine/feminine manner	−									(−)		(R)	+			+		
Has a wide range of interests			S					(+)										
Genuinely values intellectual matters																	−	S
Concerned with own bodily functions				−		R	−									(+)		
Introspective	+		S						S				−					
Concerned with own adequacy as a person			(S)															
Sympathetic, considerate															(S)		−	(S)
Has warmth; capacity for close relationships																		(S)
Enjoys sensuous experiences										(+)		(S)						
Has high aspiration level for self																		
Appears straightforward, candid							−						(+)					
Compares self to others																		(R)

(table continued)

Table A1 (*continued*)

Item	J subgroup						O subgroup						R subgroup					
	Male			Female			Male			Female			Male			Female		
	S	R	D	S	R	D	S	R	D	S	R	D	S	R	D	S	R	D
Regards self as physically attractive		(−)																
Is an interesting, arresting person		(−)														(+)		
Has a clear-cut, consistent personality			(R)							−						(+)		
Is personally charming										−								
Behaves in a giving way toward others																	(−)	(S)
Aware of impression she/he makes on others																		
Has fluctuating moods							−											
Sees to the heart of important problems															R			
Unaware of self-concern; satisfied with self									(R)			R						
Behaves in an assertive fashion						S		−						+	R			
Has insight into own motives and behavior																		S
Favors conservative values														(+)	(R)			
Initiates humor								−	S									
Prides self on being "objective," rational															(S)			
Emphasizes being with others; gregarious		−															(−)	S
Uncomfortable with uncertainty, complexities																		
Good at imaginative play, pretending, humor															(R)			
Seeks reassurance from others																		
Is facially and/or gesturally expressive	−						(+)		(S)									
Evaluates motivations of others															(S)			
Is self-indulgent									(R)									(R)
Sensitive to anything resembling a demand		(+)	(R)		(−)													(R)
Critical, skeptical, not easily impressed				−	(+)	(R)												S
Has persistent, preoccupying thoughts					−												+	R
Keeps people at a distance																		

Table A1 (*continued*)

Item	J subgroup						O subgroup						R subgroup					
	Male			Female			Male			Female			Male			Female		
	S	R	D	S	R	D	S	R	D	S	R	D	S	R	D	S	R	D
Socially perceptive of interpersonal cues															S			(S)
Judges self and others in conventional terms											(−)	S		(+)			(+)	
Relates to everyone in same way			(R)							(+)	+						−	
Is a talkative individual		(−)								+		S		+				
Is protective of those close to him/her																		
Delays gratification unnecessarily				(−)		R												
Tends to delay or avoid action							(+)	(−)	S									
Emphasizes nonverbal communication														−				
Concerned with philosophical problems	+				+												−	S
Is fastidious; fussy about minor things					+						−							
Has rapid personal tempo; acts quickly																		
Is turned to for advice and reassurance		(−)	S															S
Is basically anxious			(R)											−				
Emotionally bland; has flattened affect												(R)		(−)				
Engages in personal fantasy, daydreams			S						(S)		−	S		(−)				
Tends to proffer advice			(S)								(+)						−	(S)
Tensions find outlet in bodily symptoms												(S)	(+)					
Projects own feelings/motives onto others															(R)			R
Repressive tendencies		+	R															
Interprets situations in complicated ways	(+)																(+)	
Power oriented; values power	−																	
Tends to be self-defensive																		
Thin-skinned; sensitive to criticism						(R)					−		(−)					
Feels a lack of personal meaning in life	+					(R)							−		(R)			
Thinks and associates ideas in unusual ways	+	+					−	−		(+)								S
Is moralistic						+		(+)				(S)						

(*table continued*)

Table A1 (*continued*)

Item	J subgroup						O subgroup						R subgroup					
	Male			Female			Male			Female			Male			Female		
	S	R	D	S	R	D	S	R	D	S	R	D	S	R	D	S	R	D
Has a readiness to feel guilt													(−)	(−)				
Gives up in face of frustration and adversity																		
Unpredictable in behavior and attitudes							−	−					+	+				
Arouses nurturant feelings in others											(+)						(−)	
Has hostility towards others															(R)			R
Self-dramatizing; histrionic											(−)	(S)						(R)
Would be disorganized under stress						R												
Eroticizes situations		+			(−)			(−)									+	
Submissive; accepts domination comfortably							(+)				(+)	R						
Overreacts to minor frustrations; irritable											(−)							
Tries to see what she/he can get away with																		R
Vulnerable to real or fancied threat			S														(+)	
Basically distrustful of people in general				+												−		R
Unable to delay gratification					(−)						(+)							
Is self-defeating																		
Expresses hostile feelings directly																		
Negativistic; tends to undermine, obstruct												R					+	R
Tends to be rebellious and nonconforming		(+)								(+)								
Extrapunitive; transfers or projects blame															(S)			
Creates and exploits dependence in people																		
Feels cheated and victimized by life																		
Is condescending in relations with others		(−)										(S)		+		(−)		
Guileful, deceitful, manipulative, opportunistic											(−)							

Note. S = self-sorts; R = roommate sorts; D = discrepancy between S and R; + = item is more characteristic of the subject type than of its same-sex complement; − = item is less characteristic of the subject type than of its same-sex complement. Under D column heading, S = item rated higher on the subject types' self-sorts than on their roommates' sorts of them; R = item rated higher on the subject types' roommates' sorts of them than on their own self-sorts.

Parentheses indicate significance at <.10; roman type indicates significance at <.05; boldface indicates significance at <.01.

Received June 28, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 6, 847-857

Factors Affecting the Evaluation of Improvement: The Role of Normative Standards and Allocator Resources

André deCarufel
Faculty of Administration
University of Ottawa, Ottawa, Canada

In a simulated industrial setting, subjects performed a clerical task, believing that their pay was being determined by a (fictitious) peer allocator. After being treated inequitably, subjects were able to request a (fictitious) third party to review these allocations. The third party established one of two "normative" standards to govern a second series of allocations: either "fairness from now on" or "fairness now, plus compensation for the initial discrimination." Subjects subsequently received increased outcomes that matched one or the other of these standards. Concurrently, the resources available to the allocator either remained constant, increased, or decreased. The results indicated dissatisfaction despite the improvement in outcome level when the increase was felt to be inadequate relative to a normative standard and to the "social" standard formed by the allocator's own outcomes. Related hypotheses based on a relative deprivation interpretation of discontent with improvement are also presented.

Numerous social agencies and programs have been set up with the intention of improving the outcomes accruing to disadvantaged parties in our society. A particularly fascinating observation is that the recipients in these programs often become discontented with the aid and alienated from their benefactors. This pattern has been reported in a wide variety of settings, involving foreign aid recipients (Gergen & Gergen, 1971), disabled persons (Ladieu, Hanfman, & Dembo, 1947), and welfare clients (Briar, 1966). Similar observations have been reported in connection with the occurrence of riot and revolution during times of increasing prosperity (e.g., Davies, 1962; Gurr, 1970; Tocqueville, 1856). Clearly these negative responses by recipients are at variance with those envisioned by well-intentioned aid donors and social reformers. It is also clear that the improvement does not always lead to discontent, but the prevalence of these "paradoxical" responses suggests that a more thorough examination of the experience of receiving improved outcomes is in order.

A major theoretical approach to the evaluation of improvement by disadvantaged parties has its roots in the concept of relative deprivation (Crosby, 1976; Stouffer et al., 1949), which suggests that outcomes are evaluated not in objective terms but rather in relation to a subjective standard of deserving. An individual who receives less than he or she feels is deserved is hypothesized to experience a sense of injustice, which leads to cognitive and behavioral efforts to restore justice. There are at least two ways for discontent to arise in the specific case of outcome improvement (deCarufel & Schopler, 1979). If the standard of deservingness rises and falls with

This article is adapted from a doctoral dissertation submitted to the faculty of the University of North Carolina at Chapel Hill. The author wishes to thank his committee chairman, John Thibaut, and the members of his committee, John Schopler, Vaida Thompson, Andrea Sedlak, and Bernadette Gray-Little, for their advice and encouragement. Preparation of this article was facilitated by Canada Council Doctoral Fellowship W76 3918. The author wishes to thank John Thibaut, Wayne Harrison, and Chester A. Insko for their helpful comments on an earlier draft of this manuscript.

Inquiries concerning this article and requests for reprints should be sent to André deCarufel, Faculty of Administration, University of Ottawa, 135 Wilbrod Street, Ottawa, Ontario, Canada K1N 9B5.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3706-0847$00.75

experienced outcomes, then the improvement itself may provoke feelings of entitlement. To the extent that the actual improvement lags behind these "rising expectations," discontent results. Alternatively, the parties involved may negotiate, formally or informally, a "contract" to govern subsequent allocations. Discontent arises in this case to the extent that the improvement does not match the terms of the contract.

Research and Rationale

Despite a large relative deprivation literature, few experimental studies have examined "dynamic," or changing, outcome distributions, particularly in the case of improvement over time. Two recent studies (deCarufel & Schopler, 1979; Folger, 1977) examined a simulated work setting in which subjects in the role of a worker were initially unfairly treated by a (fictitious) allocator. Folger found discontent with a subsequent improvement when it followed an expression of allocation desires by the subject ("voice") and the improvement itself failed to restore overall equality of outcomes with the allocator. Folger suggested that the voice–improvement sequence was interpreted by subjects as an acknowledgment that they deserved more pay and created the expectation that fair allocations would now follow. These hopes were, of course, dashed by the subsequent allocations.

deCarufel and Schopler examined a number of alternative interpretations based on the type of "voicing" available to subjects and also included a broader range of outcome improvements provided by the allocator within the context of overall inequality of outcomes. The results indicated increased satisfaction with the improvement when it was introduced arbitrarily by the allocator or was in response to a threat from the worker. When the "voice" took the form of an appeal to principles of fairness, an improvement providing the worker and allocator with equal outcomes in a second series of allocations led to high satisfaction. However, an objectively *larger* improvement, which gave the worker more than the allocator in the second half as a kind of "compensation" for the initial unfairness, led to no more satisfaction than no improvement at all! deCarufel and Schopler suggested that the size of the improvement may have implied different standards of fairness and set the stage for discontent. Receiving equal outcomes may have suggested a "proximal" standard of "fairness from now on." This in fact was matched, leading to high satisfaction. The compensation, however, may have implied a "distal" standard of full overall equality, which was *not* matched, leading to dissatisfaction.

Although this evidence is suggestive of the operation of rising expectations, there remain several points of ambiguity. The first is that the improvement has been interpreted not only as an increase in the subject's outcome level but also as an implicit legitimizing by the allocator of a standard of fairness for the remaining allocations. It would seem desirable to clarify the dynamics of the voice–improvement conditions with independent manipulations of the improvement and the standard. Second, the role of social comparison with the allocator's outcomes has been largely neglected in the interpretation of subjects' evaluation of the improvement. Subjects may have felt that the allocator did not meet the standard in order to maintain his/her own outcomes at a high level. Thus, a manipulation of the resources available to the allocator with which to provide various levels of improvement would be of some interest. Finally, there exists an alternative to deCarufel and Schopler's interpretation of their legitimacy conditions. This interpretation suggests that subjects may have perceived the allocator as having been overly generous in the compensation conditions. The dissatisfaction may therefore have been due to a sense of "guilt" rather than to discontent. This hypothesis, which casts doubt on the rising expectations account, also deserves experimental attention.

Hypothesis 1. If the magnitude of the improvement fails to raise the outcome level of the individual to that specified by the legitimate standard, the individual will feel dissatisfied despite the increase in his/her absolute outcome level.

This hypothesis is a direct test of the process assumed to have mediated the discontent

in the Folger (1977) and the deCarufel and Schopler (1979) studies. This hypothesis is consistent with the relative deprivation literature and with a fundamental proposition of equity theory that individuals become distressed when they find themselves participating in an inequitable relationship (Walster, Berscheid, & Walster, 1976). Discontent should be most clearly evident here in those conditions where the distal standard of full equality is legitimized but an improvement matching "only" the proximal standard is provided by the allocator. If the proximal and distal standards are referred to as the low and high standards, respectively, and the improvements matching these standards are called the small and large improvements, respectively, discontent should be evident in the high-standard–small-improvement conditions. There are two relevant comparisons. The first is the high-standard–large-improvement conditions. In this case, the distal standard is legitimized in both sets of conditions, but the small improvement does not match the standard, whereas the large one does. The statistical comparison is the simple main effect of improvement within the high-standard conditions. The second comparison is the low-standard–small-improvement conditions. Here, the magnitude of the improvement is small in both cases but it is compared to either the proximal or the distal standard. In the first case the standard is matched, whereas in the second it is not. The statistical comparison is the simple effect of standard within the small-improvement conditions. The hypothesis will be accepted only if *both* simple effects comparisons are significant.

Recall that there exists an alternative interpretation of the deCarufel and Schopler (1979) legitimacy conditions, based on the subject's experiencing guilt over the excessive magnitude of the improvement in the overcompensation cell. It has been suggested by Adams (1965) and by Walster et al. (1976) that oversufficient as well as insufficient rewards may cause uneasiness in the recipient. Several studies (e.g., Austin & Walster, 1974; Gergen, Ellsworth, Maslach, & Seipel, 1975) have in fact demonstrated a preference for equitably distributed rewards to those that were inequitable in either direction. This possibility can be examined in the present design by comparing the low-standard–large-improvement conditions, where the legitimate standard is exceeded by the improvement, with the low-standard–small-improvement conditions, where the low standard is matched, and with the high-standard–large-improvement conditions, where the high standard is matched by the improvement. The statistical comparisons are the simple main effect of improvement within the low-standard conditions and the simple main effect of standard within the large-improvement conditions, respectively. This suggestion about the possible operation of guilt is being offered in the spirit of the devil's advocate, however, since the thrust of this article has been to argue for a rising-expectations interpretation. Nevertheless, if *both* statistical comparisons are significant, the operation of guilt over the magnitude of the improvements will be included in the theoretical model.

Hypothesis 2. When the legitimate standard of fair allocation is matched by the improvement, satisfaction should be equally high regardless of the absolute level of the standard itself or the magnitude of the improvement required to match it.

This hypothesis is a direct implication of the central tenet of relative deprivation theory. The theory states that dissatisfaction with outcome depends on the existence of a discrepancy between experienced outcomes and the standard used to evaluate them rather than on the absolute outcome level itself. It follows that eliminating the discrepancy, by raising the outcomes or by lowering the standards, should also eliminate the discontent. It should also be true that if two individuals have different standards and both experience improvements that may differ in magnitude but in each case are sufficient to reduce the outcome–standard discrepancy to zero, then each should experience equal satisfaction. This hypothesis may be tested by comparing the satisfaction of subjects in the two sets of conditions where the normative standard is matched by the improvement: the low-standard–small-improvement and the high-standard–large-improvement conditions. In this

case one would predict that the statistical contrast should be *nonsignificant*, indicating no difference in satisfaction levels. This null hypothesis prediction may seem rather unusual, but in this case it is a legitimate derivation, and within the context of the other predicted significant effects, it provides a stringent test of the theory.

Hypothesis 3. The individual's satisfaction with a given level of improvement should be independently affected by the level of the allocator's resources. In general, satisfaction should be lower when the allocator's resources increase than when they remain constant or diminish.

This hypothesis suggests that subjects will use the allocator's outcomes as one standard against which to evaluate their own outcomes. Social comparison (Festinger, 1954) is a process that is central to the experience of relative deprivation. The outcomes received by others are a powerful determinant of what we decide we ourselves deserve. Crosby (1976) made it her first precondition (seeing others possess x), and numerous studies (e.g., Weick & Nesset, 1968) have demonstrated explicitly the role of social comparison as a contributor to equity judgments. Further, the literature in the area of riot participation is also consistent with the primacy of social comparison (e.g., Pettigrew, 1967; Sears & McConahay, 1973). In this paradigm, the greater the resources of the allocator, the more will be left for him/her after providing the improvement. In other words, the greater the resources, the less the improvement narrows the gap between subject's and the allocator's outcomes created by the initial inequity. This would contribute to the overall degree of inequity between the participants and should lead to increasing dissatisfaction. The predicted statistical effect is a main effect for the resource factor.

Hypothesis 4. Satisfaction with the magnitude of the improvement should be affected by the level of the allocator's resources.

This hypothesis suggests that satisfaction with the magnitude of the improvement will depend on the perceived *ability* of the allocator to provide adequate compensation and upon his/her apparent *effort* to allocate fairly. The larger the allocator's resources, the larger is the improvement that could be provided. Should only a small rather than a large increase be provided, illegitimate motives such as the allocator's greed may be suggested. This hypothesis predicts an Improvement × Resource interaction.

Finally, although it has not been specifically predicted by any of the separate hypotheses, it has been implied that all three independent variables should exert powerful effects on subjects' satisfaction with the allocations. Thus, an overall Standard × Improvement × Resources interaction might be expected to form a context within which to examine the specific predicted effects. This interaction should indicate the most marked discontent in the high-standard–small-improvement cells, especially where the allocator's resources increase. Here, the subject would experience deprivation relative to *both* standards (normative and social) simultaneously.

Method

Participants

The participants were 108 male and female undergraduates at the University of North Carolina at Chapel Hill. All participated to fulfill a course requirement. Subjects were run in groups of three, but since interaction among them was minimal and confined to a brief period before the introduction of any of the manipulations, the individual was used as the unit of analysis. Nine subjects were assigned to each of 12 experimental conditions.

Procedure

Subjects reported in groups of three for an experiment called "Organizational Behavior" and were immediately ushered into separate rooms, which were linked to the experimenter's control room via an intercom system and a set of signal lights. The experimenter explained that the focus of the experiment was on the behavior of individuals in organizations. Subjects were told that the organization they would be participating in would have three roles, worker, allocator, and government.

At this point subjects were assigned to roles by the experimenter, using a series of signal lights. All subjects discovered that they were the worker in the organization and assumed that the other two roles were being played by the other subjects. In fact, these other roles were played by the experimenter. The experimenter then visited each subject and gave him/her a "worker's folder" containing experimental materials.

The experimenter returned to the control room and resumed the instructions, which specified the duties of the worker and allocator roles. The worker was told that his/her duties would be to perform a clerical task that consisted of transcribing sets of numbers from a computer printout onto a work sheet. The number of items transcribed was to be used as a measure of worker productivity. Subjects were also told that they would be paid by the allocator over a series of eight pay periods, each of which would last 1 minute. The allocator was to receive a sum of money from the experimenter at each pay period and would then divide up that sum between himself/herself and the worker. This money was to be exchanged for tickets in a $25 lottery to be held at the end of the semester. Subjects were asked to keep a record of the allocations on the "payment record card" in their folder. The allocator was also told that later she/he would check the worker's performance.

In order to make the experimental situation more plausible and to set the stage for the introduction of the independent variables, subjects were told that the study was concerned with the degree of contact among organization members and the market conditions in which the organization was to operate. The contact aspect was used to account for the subjects being in separate rooms and to make it plausible that the allocator would be dividing up the pay without knowing how the worker was performing and would check the worker's performance only at the end of the final pay period. The "market conditions" were described simply as the amount of money made available by the experimenter to the allocator at each pay period. This explanation set up the resource manipulation in such a way as to allow the allocator's available pay to change without making him/her appear responsible for the change (in equity theory terms, to keep the allocator's "inputs" equal in all conditions, while allowing his/her outcomes to vary). At this point, the experimenter visited each subject to answer any questions about the procedure.

The experimenter then announced that the first series of pay periods was about to begin, and the subjects were instructed to begin their tasks. At 1-minute intervals, the experimenter interrupted the workers to inform them of the amount of money available to the allocator at that pay period and how the sum had been divided between them. *All* subjects were told at each of the first four pay periods that the allocator had had 24¢ available and had kept 16¢ for himself/herself and had given 8¢ to the worker. Thus, the degree of initial inequity was identical in *all* conditions.[1] At the conclusion of Pay Period 4, subjects totaled up on their payment record cards the allocations to that point.

Introduction of the Independent Variables

When everyone had finished, the experimenter continued. He indicated that the role of the government would be to mediate any possible dispute between the allocator and worker. Both were asked to fill out a "payment review request card." This would enable them to request a government review of the allocations of the first half and to express allocation desires for the second half. The government would then review these appeals and offer a fair standard for the allocations in the second half. The standard was set by a third party to make the experimental procedures more credible, especially in those cases in which the standard would not be met. This arrangement also avoided the possibility that the allocator would be perceived as having changed his/her mind about what was fair in the interval between the appeal and the second series of pay periods. It was emphasized, though, that the allocator could still allocate freely, so that the subject would perceive him/her to be *choosing* to match or not to match the standard in the second half.

When the payment review request cards had been filled out, the experimenter collected them, ostensibly delivered them to the government, and after a short interval returned them to the subjects. The experimenter returned to the control booth and said that the government had examined the appeals and had suggested that the worker should receive either 12¢ during the second half (the proximal standard) or 16¢ during the second half (the distal standard). Each of these standards was described as the "fairest way to allocate in the second half, according to a neutral third party." Subjects were asked to write down the standard at the top of their payment record cards.

The second half of the pay periods resumed at this

---

[1] This is in contrast to both the Folger (1977) and deCarufel and Schopler (1979) studies. The former confounded initial inequity with both his equity and improvement manipulations, whereas the latter confounded it with their improvement manipulation. The present study does contain a confounding of the size of the improvement with the amount received by the allocator in the various resource conditions, however (e.g., in the diminished-resource conditions: 12¢ and 6¢ vs. 16¢ and 2¢ for the worker and allocator, respectively). This type of confounding is inevitable. To remove it would have required a confound of size of improvement with either initial inequity or with allocator resources. The significance of this confounding has been discussed in the text in reference to the "gaps" between worker and allocator outcomes, especially in connection with Hypotheses 3 and 4. In the discussion of Hypothesis 1, this gap may provide an alternative to the normative-standard–outcome-discrepancy interpretation of the simple effect of the improvement within the high-standard conditions. However, it does not account for the parallel results obtained on the simple effects test of standard within the small-improvement conditions. This being the case, parsimony would dictate that the interpretation based on discrepancies from the normative standard would be the preferred one.

point. During these four pay periods, the worker received either 12¢ or 16¢ per pay period, and the amount available to the allocator to allocate on each pay period either remained the same (24¢), increased (to 36¢), or decreased to (18¢).

Dependent Variables

The main dependent variables were assessed on the "job satisfaction questionnaire," which was administered directly after the final pay period. Five measures were selected to tap various aspects of the evaluation of the improvement and were analyzed together as a multivariate set. These measures were (a) satisfaction with pay ("How satisfied were you with your pay during the second series of pay periods?"); (b) allocator fairness ("How fairly was the allocator dividing up the pay during the second series of pay periods?"); (c) satisfaction with pay change ("How satisfied were you with your second-half pay in relation to your first-half pay?"); (d) satisfaction with standard ("How satisfied were you with your second-half pay in relation to the suggestion made by the government regarding second-half allocation?"); (e) satisfaction with allocator's pay ("How satisfied were you with your second-half pay in relation to what the allocator received in the second half?"). All of these variables were measured on 9-point scales (1 = not at all, 9 = very much).

In addition, measures were taken of subjects' satisfaction with, and perceptions of fairness of, the *total* allocations, and subjects' work performance (number of items transcribed per pay period) was tallied. Finally, several additional measures were included to test the hypothesized mediation of the predicted effects. These were a magnitude estimation measure of the size of the improvement in relation to the government standard and some "intent" measures to assess subjects' perceptions of the allocator's motives.

When subjects had completed this form and an "information questionnaire," which assessed the impact of the manipulations, the experimenter brought them into a common room and fully debriefed them.

Results and Discussion

Effectiveness of the Manipulations

The allocations during the first four pay periods were designed to create in the subject the experience of having received an inequitably small share of the available pay. In fact, 99 of the 108 subjects requested a government review of these allocations, and *all* of these subjects indicated that the reason for their appeal was that they felt they had received too little pay in the first half. The 9 subjects not requesting a review were distributed across eight cells of the design. The improvement manipulation was checked with an item from the "information questionnaire," which was administered after the final pay period ("How much of a change was there in your pay from the first to the second series of pay periods?"—1 = very little, 9 = very much). The means for the small- and large-improvement conditions were 5.8 and 7.6, respectively, $F(1, 96) = 43.85$, $p < .001$. The standard manipulation was also checked with an item from the information questionnaire ("How much of a change in your pay did the government suggest?"—1 = very little, 9 = very much). The means for the low- and high-standard conditions were 6.1 and 7.9, respectively, $F(1, 96) = 48.25$, $p < .001$. Finally, the resource manipulation was checked by asking subjects to indicate how much pay had been available to the allocator during the second half. Counting 18¢, 24¢, and 36¢ as correct for the diminished, constant, and increased resource conditions, respectively, 105 of the 108 subjects (97%) responded correctly.

Overall Impact of the Independent Variables

Two measures asked subjects to consider the magnitude of the improvement in relation to the government standard. These were the magnitude estimation measure ("How did the change in pay you received from the allocator compare with the amount suggested as appropriate by the government?"—1 = too little, 9 = too much), and the satisfaction-with-standard measure from the evaluation of the improvement set of measures. Both of these showed a Standard × Improvement × Resource interaction, $F(2, 96) = 6.48$, $p < .01$, and $F(2, 96) = 5.39$, $p < .01$, respectively. Two other measures were marginally significant for this effect: how satisfied subjects were with their *total* pay, $F(2, 96) = 2.99$, $p < .06$, and the extent to which they perceived that the allocator was "trying" to allocate fairly during the second half, $F(2, 96) = 2.65$, $p < .08$. The means for these items were in accord with expectations. Satisfaction was lower and subjects felt that the improvement compared unfavorably with the standard in the high-standard–small-improvement condi-

tions, especially where the allocator's resources increased. The triple interaction was also significant on the behavioral measure of number of items transcribed during the second series of pay periods, $F(2, 96) = 4.04$, $p < .05$.[2]

Tests of the Hypotheses [3]

Hypothesis 1. This was the major "discontent" hypothesis, which predicted that subjects would be more dissatisfied with the increase in their outcome level if the improvement failed to meet the normative standard than if it did. Before proceeding to the simple effects tests, it should be noted that the overall Standard × Improvement interaction on the items in the improvement set was significant, multivariate $F(5, 92) = 5.78$, $p < .001$.

The simple effect of improvement within the high-standard conditions was significant, multivariate $F(5, 92) = 8.08$, $p < .001$. The F values for the individual improvement measures organized by test of hypothesis are presented in Table 1. The means for these items are presented in Table 2. Further evidence comes from two auxiliary measures, the extent to which subjects perceived that the allocator "went far enough" in his/her efforts to be fair and the "trying" measure discussed above. Both of these were significant for the simple effect of improvement within the high-standard conditions, $F(1, 96) = 21.34$, $p < .001$, and $F(1, 96) = 13.38$, $p < .001$, respectively. Thus when the distal standard had been legitimized, subjects felt more dissatisfied with the small improvement that failed to match it than they did with the large improvement that did. They also felt that the allocator was not trying as hard to be fair.

The simple effect of standard within the small-improvement conditions was also significant, multivariate $F(5, 92) = 6.83$, $p < .001$. The two intention measures were supportive as well: "went far enough," $F(1, 96) = 7.30$, $p < .01$, and "trying," ($p < .10$). This effect suggests that subjects were more satisfied with the small improvement when it was evaluated against the proximal rather than against the distal standard.

These data of course do not prove that the arousal, legitimation, and subsequent violation of normative standards were responsible for the Folger (1977) and deCarufel and Schopler (1979) results. They do show, however, that if the conditions they speculated about are experimentally created by separating the procedure of setting the normative standard from the actual delivery of the improvement, discontent is evident in those cases where the improvement is evaluated as inadequate. This was true whether discontent was measured using a single standard and two levels of improvement that either did or did not match it, or with a single level of improvement being evaluated against either a low or a high standard.

Recall that the guilt interpretation of deCarufel and Schopler's (1979) study was also evaluated. The simple effect of improvement within the low-standard conditions was significant, multivariate $F(5, 92) = 2.71$, $p <$

---

[2] There had been no behavioral differences during the first half when the allocators were equal across conditions. The pattern was somewhat different from that described above, but still contained strong links to the theoretical framework. The resource effect (Hypothesis 3, below) was evident, but only under the large-improvement conditions. In fact, the Improvement × Resource interaction was also significant, $F(2, 96) = 3.83$, $p < .025$. In the small-improvement conditions, productivity varied by standard. In the low-standard conditions, productivity was highest in the constant resource condition, whereas it was *lowest* in the constant resource cell under the high standard. Two things need to be said about these productivity effects. The first is that neither Folger (1977) nor deCarufel and Schopler (1979) found them in similar paradigms. Perhaps this was because in those studies, the standards of fairness were largely implicit as subjects arrived at an interpretation of the voice–improvement sequences. In the present study, the standards were explicit and allowed subjects to be aware of the unfairness throughout the second half and, thus, to alter their work output accordingly. Second, it is interesting that the hypothesized pattern of means was most clearly evident in the constant resource conditions. Here the normative standard was most clearly relevant to the subsequent allocations because there was no change in the allocator's resources to suggest the appropriateness of any other standard. This is consistent with the speculations offered above that productivity effects would be most likely under conditions of clear-cut violation or no violation of the standard.

[3] There were no effects for sex of subject, so the data for males and females were combined.

Table 1
F Values for Measures in the Improvement Set, by Hypothesis

Measure	Hypothesis 1: Simple effects tests on Standard × Improvement interaction		Hypothesis 2: Comparison of low-standard–small-improvement conditions with high-standard–large-improvement conditions			Hypothesis 3:	Hypothesis 4:
	Improvement within high standard	Standard within small improvement	Diminished resource	Constant resource	Increased resource	Resource main effect	Improvement × Resource interaction
Satisfaction with pay	6.04*	1.51	<1	<1	2.14	81.72***	1.88
Allocator fairness	3.32	1.64	1.65	<1	3.49	36.93***	6.51**
Satisfaction with pay change	6.69*	2.88	<1	1.59	5.67*	57.16***	9.08***
Satisfaction with standard	34.28***	30.18***	<1	<1	2.26	50.45***	5.36**
Satisfaction with allocator's outcome	<1	<1	<1	<1	1.09	38.57***	4.41**

* $p < .05$.
** $p < .01$.
*** $p < .001$.

.05. Inspection of the means for the items in the improvement set indicates, however, that the direction of the difference is the *opposite* of a guilt effect! Subjects evaluated the improvement *more* positively when the standard was exceeded than when it was matched. This effect, combined with the simple effect of improvement within the high-standard conditions, resulted in a significant improvement main effect, multivariate $F(5, 92) = 4.98$, $p < .001$. Clearly, there was no evidence of guilt here.

The simple effect of standard within the large-improvement conditions was *not* significant ($.10 < p < .15$) at the multivariate level, again indicating the lack of a guilt effect. The major factor in this paradigm that would seem to blunt a guilt effect is the subjects' experience of initial inequity. In studies done within the "obligation tradition" (e.g., Ger-

Table 2
Means for Measures in the Improvement Set

Measure	Low standard						High standard					
	Small improvement			Large improvement			Small improvement			Large improvement		
	Resource			Resource			Resource			Resource		
	Diminished	Constant	Increased	Diminished	Constant	Increased	Diminished	Constant	Increased	Diminished	Constant	Increased
Satisfaction with pay	8.11	8.00	2.44	8.44	7.89	4.89	7.89	6.56	2.33	8.11	8.56	3.67
Allocator fairness	7.22	7.78	2.33	6.11	6.44	5.00	6.89	6.33	2.00	6.00	8.11	4.11
Satisfaction with pay increase	8.44	8.33	3.44	8.89	7.89	6.78	8.11	7.22	2.56	8.78	7.33	5.33
Satisfaction with standard	8.56	8.78	3.89	8.00	8.00	4.56	8.00	3.00	1.78	8.00	8.56	5.22
Satisfaction with allocator's pay	6.33	7.67	1.89	5.78	7.22	4.67	7.56	6.78	1.56	6.22	6.89	3.00

Note. All of the measures in the improvement set used 9-point scales (1 = not at all, 9 = very much).

gen et al., 1975), where guilt effects are common, the subjects' need usually arises because of their bad luck or the difficulty of the task. In these cases, the aid may be viewed as *help*, which may imply the need to reciprocate at a later date. In the present situation, however, the subject is initially mistreated by the allocator. Here, the improvement is viewed as *compensation*. This interpretation of the aid by the recipient seems to have little capacity to arouse the need to repay the former tormenter (e.g., de-Carufel & Carere, Note 1).

Hypothesis 2. Hypothesis 2 examined subjects' satisfaction in those conditions where the standards were matched. Specifically, it was predicted that since in neither the low-standard–small-improvement nor in the high-standard–large-improvement condition was there a discrepancy between experienced and deserved outcomes, there should be no significant difference in satisfaction between them despite the difference in absolute outcome level. A special contrast of these conditions was run separately in each of the levels of the resource factor in anticipation of the concurrent resource main effect (Hypothesis 3, below). Thus, each contrast examined *only* the effect of matching the normative standards, holding constant the level of the allocator's resources. Consistent with the hypothesis, this contrast was *nonsignificant* at the multivariate level in each of the three resource rows ($ps > .20$). Thus, the relative deprivation hypothesis that dissatisfaction is based on a discrepancy between experienced outcomes and the standard used to evaluate them rather than on the basis of their absolute level received further support.

Hypothesis 3. Hypothesis 3 suggested that subjects would use not only the normative standards established by the government to evaluate the improvement but also the outcomes received by the allocator. It was expected that satisfaction would be lowest in those conditions where the allocator's resources increased. This prediction was confirmed. The resource main effect was significant, multivariate $F(10, 184) = 18.22$, $p < .001$. The effect was also evident on the intention items: "went far enough," $F(2, 96) = 92.79$, $p < .001$, and "trying," $F(2, 96) = 61.73$, $p < .001$, as well as satisfaction with total pay, $F(2, 96) = 39.02$, $p < .001$, and fairness of the total allocations, $F(2, 96) = 41.17$, $p < .001$. Subjects also transcribed fewer numbers onto their work sheets when the allocator's resources increased, $F(2, 96) = 3.00$, $p < .06$.

Thus, it is clear that subjects did base their perceptions of fairness and feelings of satisfaction at least in part on social comparison with the allocator's outcomes. In the increased resource conditions, the allocator was able to provide the required compensation without depriving himself/herself of a high outcome level. When the allocator continued to prosper and the gap created by the initial unfairness was not reduced, subjects felt that no real effort to be fair was being made. This dissatisfaction is reminiscent of Pettigrew's (1967) observation that blacks felt deprived relative to whites despite their economic and educational gains during the 1960s because the initial gap between blacks and whites had not been reduced and in some cases had grown wider. This perception of insufficient effort may have been reinforced in the present experiment by the fact that the increased resources were provided to the allocator by the experimenter rather than resulting from increased effort on the allocator's part.

Further evidence of the operation of social comparison processes *over and above* the effect of matching the normative standards comes from the fact that the resource main effect was significant within the low-standard–small-improvement and the high-standard–large-improvement conditions, multivariate $F(10, 184) = 7.11$, $p < .001$, and multivariate $F(10, 184) = 6.63$, $p < .001$, respectively. Thus, it seems clear that the "external" resource factor caused subjects to feel that matching *either* normative standard was an insufficient indication of the allocator's goodwill when his/her resources increased. The likely result would be a desire to renegotiate the standard, beginning anew the process of voice and (possible) improvement.

Hypothesis 4. Hypothesis 4 predicted that if the allocator were to be perceived as being

able to provide a large improvement because of increased resources, subjects would be dissatisfied if "only" a small one were granted. The predicted Improvement × Resource interaction was significant, multivariate $F(10, 184) = 3.79$, $p < .001$. The intention measures also showed the effect: "went far enough," $F(2, 96) = 2.92$, $p < .06$, and "trying," $F(2, 96) = 12.45$, $p < .001$, as did the fairness of the total allocations measure, $F(2, 96) = 4.15$, $p < .05$. It is clear that subjects took the allocator's ability to provide adequate compensation into account in their evaluation of the improvement. The small increase was seen as inadequate when the allocator's resources increased and led to lower satisfaction than did the large one.

General Discussion and Conclusions

The present study was concerned with the evaluation of the adequacy of compensation received by victims of injustice in relation to salient legitimate standards of fairness. The major theoretical perspective was that of relative deprivation. The theory was extended to predict that individuals develop expectations about what they ought to receive and that dissatisfaction may result even with *improving* outcomes if the increases do not match expectations. In fact there is evidence (Folger, 1977) that insufficient improvement may *intensify* discontent by legitimizing the victim's perceived right to higher outcomes and by bringing the former injustice into sharper relief.

Discontent with improvement was evident in two sections of the present study, in the high-standard–small-improvement conditions and in those where the allocator's pay increased. In both cases, subjects' legitimate expectations of fair allocation were violated by the allocator, apparently to his/her own benefit. In the former conditions, normative standards were legitimized by the government and not matched by the improvement; and in the latter, subjects felt spontaneously deprived relative to the allocator's own improved outcomes. This discontent was matched by perceptions of unfairness of the allocations and with the attribution that the allocator had not tried hard enough to be fair. These demonstrations are important because they serve to clarify the dynamics of the discontent in the Folger (1977) and deCarufel and Schopler (1979) studies. In those studies, the operation of normative standards was a matter of conjecture, and the role of social standards and social comparison was virtually ignored. Further, the present study found no evidence to support a guilt-based interpretation of these previous results. Thus, with the results of this experiment, the relative deprivation account of discontent with improvement in this paradigm has been considerably strengthened.

Finally, since this article opened with a discussion of discontent in "real world" settings, the focus of the discussion should be broadened briefly to show the relevance of the present research to those situations. Normative standards of allocation such as distributive justice are used in a variety of settings as a criterion of fair pay (Adams, 1965), and when they are violated, discontent may be expressed in strike action or absenteeism from the work place. Similarly, the data of Sears and McConahay (1973) indicated that Watts riot participants used normative standards such as the formal egalitarian ideology of the U.S. Constitution as a basis for their discontent with racial inequality. The normative standards used in the present study are of direct relevance to those often invoked to deal with the problems of the disadvantaged. The proximal standard is similar to statutes that prohibit injustice from that point on, whereas the distal standard represents the contention of many victims of injustice that they should receive compensation for their past misfortune. This standard finds expression in quota systems for hiring, awards of back pay lost through discrimination, and so on. Social standards are also clearly used, as Pettigrew (1967) and others have noted. As in the present study, the outcomes received by the (previous) harm-doer are often those selected by the disadvantaged as the criterion of their own deservingness.

Unlike many social psychological phenomena studied in laboratory settings, in which the task is to establish some "external validity," the literature in the area of relative

deprivation contains many striking instances of revolution, riot, and so on that have been interpreted in this framework in a post hoc fashion. What is lacking here is a sufficiently elaborated and tested *theory* that can make a priori predictions about the instigating conditions for discontent. The results of the present experiment offer cause for optimism that the rising-expectations framework under consideration here may prove fruitful in this regard.

Reference Note

1. deCarufel, A., & Carere, R. *The evaluation of outcome improvement by disadvantaged parties.* Paper presented at the annual meeting of the Canadian Psychological Association, Ottawa, Canada, June 1978.

References

Adams, J. S. Inequity in social exchange. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2). New York: Academic Press, 1965.

Austin, W., & Walster, E. Reactions to confirmations and disconfirmations of expectancies of equity and inequity. *Journal of Personality and Social Psychology,* 1974, *30,* 208–216.

Briar, S. Welfare from below: Recipients' views of the public welfare system. *California Law Review,* 1966, *54,* 370–385.

Crosby, F. A model of egoistical relative deprivation. *Psychological Review,* 1976, *83,* 85–113.

Davies, J. C. Toward a theory of revolution. *American Sociological Review,* 1962, *27,* 5–19.

deCarufel, A., & Schopler, J. Evaluation of outcome improvement resulting from threats and appeals. *Journal of Personality and Social Psychology,* 1979, *37,* 662–673.

Festinger, L. A. A theory of social comparison processes. *Human Relations,* 1954, *7,* 117–140.

Folger, R. G. Distributive and procedural justice: Combined impact of "voice" and improvement on experienced inequity. *Journal of Personality and Social Psychology,* 1977, *35,* 108–119.

Gergen, K. J., Ellsworth, P., Maslach, C., & Seipel, M. Obligation, donor resources, and reactions to aid in three cultures. *Journal of Personality and Social Psychology,* 1975, *31,* 390–400.

Gergen, K. J., & Gergen, M. M. International assistance from a psychological perspective. In G. W. Keeton & G. Schwarzenberger (Eds.), *Yearbook of world affairs.* New York: Praeger, 1971.

Gurr, T. R. *Why men rebel.* Princeton, N.J.: Princeton University Press, 1970.

Ladieu, G., Hanfman, E., & Dembo, T. Studies in adjustment to visible injuries: Evaluation of help by the injured. *Journal of Abnormal and Social Psychology,* 1947, *42,* 169–192.

Pettigrew, T. F. Social evaluation theory: Convergences and applications. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 15). Lincoln: University of Nebraska Press, 1967.

Sears, D. O., & McConahay, J. S. *The politics of violence.* Boston: Houghton Mifflin, 1973.

Stouffer, S. A., Suchman, E. A., DeVinney, L. C., Star, S. A., & Williams, R. M. *The American soldier: Adjustment during army life* (Vol. 1). Princeton, N.J.: Princeton University Press, 1949.

Tocqueville, A. de. [*The old regime and the revolution*] (J. Bonner, trans.). New York: Harper, 1856.

Walster, E., Berscheid, E., & Walster, G. W. New directions in equity research. In L. Berkowitz & E. Walster (Eds.), *Advances in Experimental Social Psychology* (Vol. 9). New York: Academic Press, 1976.

Weick, K. W., & Nesset, B. Preferences among forms of equity. *Organizational Behavior and Human Performance.* 1968, *3,* 400–416.

Received December 27, 1977 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 6, 858-864

Cardiovascular Changes During Social Competition in a Mixed-Motive Game

Lawrence F. Van Egeren
Department of Psychiatry
Michigan State University

Male and female subjects played a mixed-motive game against a male confederate under either a 20% cooperative or an 80% cooperative strategy while cardiovascular responses were computer monitored. Females had larger heart rate responses than males during play against the competitive strategy, and the opposite was true during play against the cooperative strategy. Subjects who were more competitive during the game or who scored higher on a coronary-prone (Type A) behavior scale or who reported having an action orientation toward life stress tended to have larger heart rate responses during the game than the remaining subjects. The results draw attention to the importance of covert autonomic responses for understanding overt behavioral choices in mixed-motive games and to the potential utility of this behavioral model for studying the role of psychosocial factors in psychosomatic illnesses.

Cardiovascular responses in competitive conflict situations are potentially interesting for two reasons: (a) Competitiveness (struggles for dominance, control) is a putative etiologic factor in cardiovascular disorders (Friedman & Rosenman, 1974; Henry & Ely, 1976), and (b) in human evolution, conflicts over resources were usually settled by vigorous action requiring cardiovascular arousal (Tiger & Fox, 1971); people today may carry important vestigial, now metabolically excessive, competition-released cardiovascular response tendencies. What the *form* of cardiovascular responses during mixed-motive games might be is not clear, because the nature of these games as a stressor is not clear. Cardiovascular stress studies by Obrist (e.g., Obrist et al., 1978) indicate that the type of response depends on whether stress is controllable (i.e., whether some active response is available to deal with it). When it is controllable, the heart pumps more vigorously, presumably in preparation for the action, and systolic blood pressure rises. When it is not controllable, the diastolic blood pressure rises, apparently as the result of constriction of peripheral blood vessels.

Two mixed-motive game experiments involving cardiovascular variables have been reported. In one, after an anger-induction experience, subjects played a game against a confederate of the experimenters, the anger instigator, under either predictable or unpredictable conditions (Van Egeren, Abelson, & Thornton, 1978). When the confederate's behavior was unpredictable and subjects had little control over outcomes, blood volume pulses were transmitted faster from the heart to a finger, and diastolic blood pressure decreased less during the game than when the confederate was predictable and subjects had greater control. These results seem to agree with the Obrist et al. (1978) findings for non-

I wish to thank Lawrence Sniderman and Dozier Thornton for their help in this experiment. Mr. Sniderman designed and implemented computer software for this research.

The project was supported by National Heart, Lung, and Blood Institute Grant 1 RO1 HL21319-01 to the author and by National Research Service Grant RR05656-09 awarded by the Biomedical Research Support Grant Programs, National Institute of Health, to Michigan State University.

Address reprint requests to Lawrence F. Van Egeren, Department of Psychiatry, East Fee Hall, Michigan State University, East Lansing, Michigan 48824.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3706-0858$00.75

social stressors. In another experiment, Blascovich, Nash, and Ginsburg (1978) related resting heart rates prior to a mixed-motive game to game behavior. Since the games were zero-sum, it was possible to divide the subjects into winners and losers. Male winners (i.e., the more competitive males) had faster pregame heart rates than male losers, 82 versus 74 beats per minute. Females had the opposite pattern, 82 versus 83 for winners versus losers, but this difference was nonsignificant.

Three individual differences in game response were explored in the present experiment: action orientation towards stress, a behavior pattern called Type A, and sex membership. The Type A behavior pattern, described by two cardiologists (Friedman & Rosenman, 1974), includes intense competitiveness, high achievement drive, impatience, and intense facial and vocal mannerisms. The more unhurried and relaxed person is referred to as Type B. In predictive studies, Type A's have had from two to five times the incidence of ischemic heart disease as Type B's (Rosenman, 1974; Rosenman et al., 1966). Sex differences in game behavior are often inconsistent (Davis, Laughlin, & Komorita 1976), but males seem to be more concerned with winning than females (Komorita, 1965). Because the present experiment was conducted before the appearance of the Blascovich et al. (1978) study, no prediction of an interaction between sex membership and competitiveness in heart rate was made. It was simply felt that because of male–female differences in aggression training, coded in part in sex role stereotypes surrounding aggression (Frodi, Macaulay, & Thome, 1977), social competition has a different emotional significance for men and women and that this would influence cardiovascular responses during a mixed-motive game. Specifically, it was predicted that when subjects were paired with a competitive opponent, cardiovascular arousal would be greater for males than females, greater for Type A's than Type B's, and greater for action-oriented subjects than passive-oriented subjects. The opponent for all subjects was a male confederate of the experimenter. Subjects appeared to play the game against the confederate but actually played against a laboratory computer that was programmed to be either very competitive or very cooperative. An electrocardiogram and blood volume pulses in a finger were computer monitored during the game.

Subject's Choice	Confederate's Choice: Cooperate	Confederate's Choice: Compete
Cooperate	1, 1	1, 3
Compete	3, 1	-3, -3

Figure 1. Matrix representation of payoffs. (The first number in each cell is the payoff in points to the subject. The second number is the payoff to the opponent [confederate].)

Method

Subjects

The subjects were 16 male and 16 female Michigan State University students enrolled in introductory psychology courses. Subjects of each sex were randomly assigned to treatment conditions.

Payoff Matrix

All subjects played against a first-year male medical student (confederate), using the payoff matrix shown in Figure 1.

The payoff structure was designed to make it easy to cooperate, tempting to compete, and more palatable to give in to the opposing player's competitiveness than is true in the Prisoner's Dilemma game. The major force of the game lies in the direction of a struggle for the "compete" response, but the asymmetric outcome is not as "unfair" to the loser as it is in Prisoner's Dilemma. Subjects were paid 2¢ per point.

Experimental Setting

The experiment involved the use of a computer-controlled system for the study of two-person interactions described in greater detail elsewhere (Van Egeren et al., 1978. A laboratory computer (DEC PDP 11/40 system) presented information to subjects, paced the game, collected behavioral and physiological responses, and reduced the data after each subject was run. An equipment room and a

subject room were connected by one-way mirrors and an intercom system. The subject and confederate sat side by side, 2 feet (61 cm) apart, separated by a curtain that dropped from the ceiling to the seat of the chair. The two players interacted by pressing buttons on a response panel placed in front of them. The computer "read" these button presses, decoded them, and indicated the players' moves (behaviors) on a television receiver placed so that both players could see it. Electrophysiological signals were recorded continuously from the subject and sampled intermittently at critical times by the computer (described below).

Procedure

After brief orienting instructions, electrodes and transducers were applied to the subject and the confederate. Players were handed typed instructions, taken from Rapoport and Chammah (1965) with slight modifications. A short rest was followed by a 10-trial practice (Prisoner's Dilemma) game to familiarize players with the procedures. A second brief rest was followed by new instructions and 45 plays of the game using the payoffs shown in Figure 1. The new instructions indicated that points were now worth 2¢ each. Following the game, the subject filled out two questionnaires (described below) and was debriefed and paid.

A play was initiated by display of the message "MAKE DECISION" on the TV screen, followed by a message "MAKE RESPONSE," which terminated after both subjects responded. Players responded by pressing either a green (cooperate) or red (compete) button. After the subject responded there was 3–5-sec delay before the outcome of the play was displayed for 10 sec. The display included the play number, players' choices, players' payoffs, and cumulative gains or losses. The interplay interval was random over the range of 10–20 sec.

The confederate pressed buttons during the games, but the button presses were not read by the computer. The computer followed a preprogrammed strategy to play all subjects 50% cooperative in a fixed play sequence during the practice game. The computer played 80% cooperative in a *cooperative strategy* condition and 20% cooperative in a *competitive strategy* condition during the game that followed the practice game. Eight males and eight females were assigned to each strategy condition.

Physiological Measures

Physiological responses were computer sampled every third play during the game, starting with Play Number 3. Three 6-sec samples were taken on each play sampled: (a) a prestimulus sample (PRE) prior to onset of the message "MAKE DECISION," (b) an anticipatory response sample (ANT) during the message "MAKE DECISION," and (c) an outcome sample (OUT) during display of the outcome message. Samples b and c were initiated 1 sec after onset of the messages.

1. *Electrocardiogram.* A single lead (V_5) was recorded. The output of a Grass 7P3 cardiotachometer was sampled by the computer at 10 Hz and converted to heart rate in beats per minute (bpm).

2. *Digital blood volume pulse.* Relative changes in blood volume pulse were recorded using a transmissive photoplethysmograph transducer placed on the first phalanx of the left-hand little finger. The output of a driver amplifier coupled to a Grass 7P1 preamplifier was computer sampled at 20 Hz. A computer routine for the analysis of blood volume pulses computed the trough and peak of the pulses associated with each heart beat. The term "pulse amplitude" refers to peak-minus-trough values. Mean digital pulse amplitude (DPA—mean DPA during ANT and OUT samples divided by mean DPA during PRE sample) will be reported.

3. *Respiration.* Tracings of respiratory excursions were recorded in order to detect the interference of cardiovascular activity caused by respiratory maneuvers. The tracing itself was not scored.

Questionnaires

Two questionnaires were administered at the end of the experiment: a student version of the Jenkins Activity Survey for Health Prediction (JAS) and the Habit Survey of Nervous Tension (HSNT). The JAS is a set of items sampling major components of the Type A behavior pattern (competitiveness, impatience, time urgency, etc.) developed by Jenkins (1965). In a validation study of 2,960 men, the JAS agreed with A–B classifications based on a criterion interview conducted 3 to 5 years earlier in 73% of the cases (Jenkins, Zyzanski, & Rosenman, 1974). A version of the JAS adapted for college students by Krantz, Glass, and Snyder (1974) was administered. The HSNT is a checklist of cognitive, behavioral, emotional, and autonomic responses to stressors developed by Thomas and Ross (1963). A score indicating the degree of action orientation towards stress was derived by scoring the following five items in the direction indicated: depressed feelings (−), increased activity (+), decreased activity (−), increased urge to sleep (−), and urge to be by oneself and get away from it all (−).

Results

The mean percentages of competitive responses (red button presses) of males and females playing against cooperative and competitive programmed strategies were 62% and 64% (males) and 57% and 62% (females), respectively. The percentages were computed over 45 plays of the game. An

analysis of variance of percentages of competitive responses indicated that neither main effect (sex and strategy) nor the interaction was statistically significant at a 5% alpha level.

Physiological responses during the prestimulus (PRE) period were chosen as reference points for evaluating the effects of stimuli during the anticipatory and outcome periods surrounding a behavioral choice. A two-way (Sex × Strategy) analysis of variance of PRE period heart rates, averaged over 15 trials, was computed. The sex and strategy main effects were not significant at a 5% alpha level, but the interaction was significant, $F(1, 28) = 6.01$, $p < .02$. Females playing against the competitive preprogrammed strategy had larger PRE heart rates (84 bpm) than females playing against the cooperative strategy (78 bpm); the opposite was true for males, for whom the heart rates were 70 and 81 bpm, respectively.

Two-way analyses of covariance of maximum heart rate, minimum heart rate, and heart rate range (maximum minus minimum) during the anticipatory and outcome periods were computed, with mean PRE heart rate the covariate. Measures were averaged over trials before the analysis, and tests were two-tailed. Again, the main effects were not significant, but there was a significant Sex × Strategy interaction for the maximum heart rate, $F(1, 27) = 7.50$, $p < .02$, and heart rate range, $F(1, 27) = 6.94$, $p < .02$, during the anticipatory periods, and for the minimum heart rate, $F(1, 27) = 5.43$, $p < .03$, and heart rate range, $F(1, 27) = 4.34$, $p < .05$, during the outcome period. The adjusted heart rates, displayed in Figure 2, indicate that females who played against the competitive strategy accelerated more prior to making a behavioral choice and decelerated more when the results were displayed than females who played against the cooperative strategy; the opposite was true for males. A significant Sex × Strategy interaction appeared in within-trial heart rate responses as well as between-trial heart rate levels, even when the former were adjusted for group differences in the latter.

The method for measuring blood volume pulses used in this experiment does not yield valid estimates of absolute blood volume. The prestimulus period measures were therefore not analyzed, since only relative changes from an unstimulated to a stimulated state are physiologically meaningful. Two-way analyses of variance of percentage changes in pulse amplitude from the prestimulus period to the anticipatory and outcome periods were computed. Neither the main effects nor the interaction was significant.

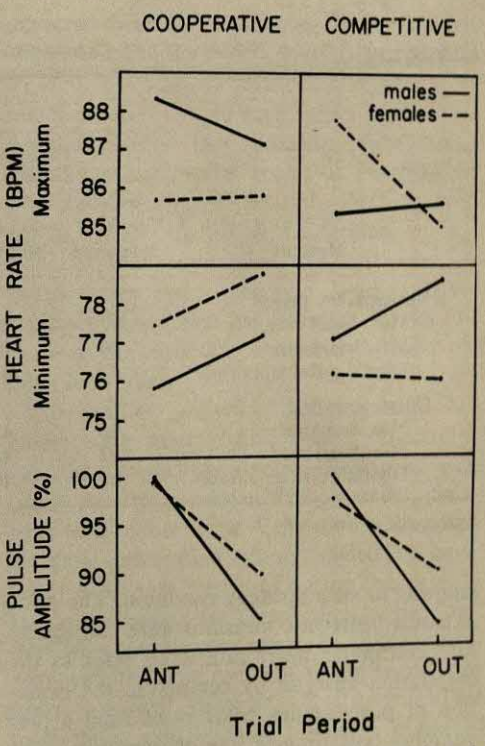


Figure 2. Comparison of cardiovascular activity of males and females during anticipatory (ANT) and outcome (OUT) periods surrounding behavioral choices while playing against either a cooperative or competitive programmed strategy. (Each data point is based on 16 subjects. The heart rates displayed have been adjusted for group differences in prestimulus heart rate level. BPM = beats per minute.)

Three behavioral measures (percentage of competitive responses, Type A behavior scores, and action-oriented coping scores) and six cardiovascular measures (maximum heart rate, minimum heart rate, and digital pulse amplitude during the anticipatory and outcome periods) were intercorrelated across

Table 1
Correlations Between Behavioral and Cardiovascular Measures

Measure	Strategy condition: Competitive			Strategy condition: Cooperative		
	% Competitive responses	Type A behavior	Action-oriented coping	% Competitive responses	Type A behavior	Action-oriented coping
Anticipatory period						
Max. heart rate	.00	−.05	−.15	.22	−.05	−.53*
Min. heart rate	.00	.07	.09	−.55*	−.07	.43*
Digital pulse amplitude	.22	.45*	.24	−.21	.16	−.04
Outcome period						
Max. heart rate	.20	.47*	.61*	.43*	−.17	−.15
Min. heart rate	.22	.45*	.59*	−.20	−.13	.53*
Digital pulse amplitude	.24	.40	−.58*	−.28	−.04	.13

* $p < .05$.

subjects in each strategy condition. The poststimulus heart rate measures were "adjusted" for prestimulus heart rate levels prior to the correlation analysis by computing the deviation of poststimulus heart rates from a prestimulus–poststimulus line of regression. The results appearing in Table 1 may be summarized as follows. First, the more the subject competed during play against the cooperative program the greater his/her heart rate acceleration when the outcomes were displayed. Second, the higher the subject's Type A questionnaire score, the larger his/her heart rate response when outcomes were displayed during play against a competitive strategy (i.e., the higher the subject's maximum and minimum heart rate following an adjustment for the subject's mean prestimulus heart rate). Third, subjects who reported having an action orientation towards stressful life events also had larger maximum and minimum heart rate responses when outcomes were displayed during play against the competitive strategy. The three behavioral measures themselves were not significantly correlated; the largest intercorrelation was .11.

Discussion

Subjects in this experiment may be classified as either "winners" or "losers" on the basis of the strategy played against them. With two exceptions, subjects who played against the cooperative strategy won the game (gained more points than the confederate), and subjects who played against the competitive strategy lost the game. Then, the unexpected finding of a Sex × Outcome interaction in heart rate agrees with the Blascovich et al. (1978) results. For both a pregame period (Blascovich et al.) and pretrial periods during the game (present experiment), winning was associated with faster pumping of blood and losing with slower pumping in males; the opposite was true for females. Furthermore, the interaction continued *within trials,* male winners having larger heart rate accelerations prior to making behavioral choices and greater slowing of the heart when the outcomes of plays were displayed than losers; and again the opposite was true for females. Males seemed to find play against the cooperative program, during which they won, more "stimulating" (eliciting a larger maximum heart rate response and a greater range of response) than play against the competitive program, during which they lost; and the opposite was true for females. How might these rather surprising results be explained?

Males and females may have approached the one-on-one competitive game differently. Males may have focused more on winning (Komorita, 1965) and females more on the chances of losing (e.g., winning may have

been more exciting to males and losing more fear provoking to females). Frodi et al. (1977) comment on sex differences in affective response to aggressive cues in their recent extensive review as follows: "Some evidence suggests that men and women react differently to external aggressive cues and provocation. . . . What may be anger-provoking for men may be anxiety-provoking for women" (p. 634).

Sex differences in dominance–subordination relations is one of the foci of the current sociobiology debate (e.g., Tiger & Fox, 1971; Wilson, 1978). Viewed this way, heart rate responses were greatest when the males were dominant (winning) and the females were subordinate (losing) in this experiment. Whether men and women are culturally or genetically conditioned to respond in such ways is one of the issues in the debate. Blascovich et al. (1978) mentioned that men and women are socialized differently for competition and have different degrees of familiarity with competitive contests. Sociobiologists point out that the hunting-gathering way of life, in which men hunted wild animals and women gathered wild plants was a human universal for more than a million years (Lee & DeVore, 1976). Surely this major adaptation over such a long period of time, with its sharp sexual division of labor, exerted different selection pressures on males and females. Thus, it is not surprising that men and women attach a different socioemotional significance to modern equivalents of the prehistoric hunt such as sporting contests (and mixed-motive games?), or so the argument goes (see Tiger & Fox, 1971).

An explanation of the results based on sex role stereotypes and conformity to social norms is also worth considering. In Western cultures males are expected to be aggressive and females to be nonaggressive. It may have been more difficult (conflictual, distressing) for the males interacting with a friendly confederate, and the females interacting with an unfriendly confederate, to perform sex role-expected behaviors than was true of their same-sex counterparts. Thus, the subjects who interacted with a confederate whose behavior was noncomplementary to their own sex role-expected behavior had larger heart rate responses than did subjects who interacted with a "complementary" confederate, regardless of sex of subject. Others have shown in modified Asch situations that pressures encouraging deviations from social norms or expectations produce physiological arousal (Back & Bogdonoff, 1963; Costell & Leiderman, 1968). Regardless of explanation, the difference between males and females in heart rate response in a competitive situation in the absence of a significant behavioral difference is quite interesting.

The nonsex individual differences in behavior were most clearly related to cardiovascular changes during the display of outcomes against a competitive opponent. Subjects with higher Type A questionnaire scores or action-orientation scores tended to have larger heart rate responses during this period. It was as though, for these subjects, the "action" was not yet over when the outcome was displayed against a competitive opponent; the heart was still preparing for action. The larger heart rate response toward a competitive opponent of subjects with high Type A questionnaire scores is in agreement with Type A behavior theory (Friedman & Rosenman, 1974), whereas the lack of correlation between the questionnaire and the actual competitiveness of subjects during the game is not. A recent experiment in our laboratory showed that subjects taken from the *extremes* of the questionnaire distribution do differ in competitiveness in a mixed-motive game (Van Egeren, 1979).

Generalization of the present results for sex is limited by the fact that both males and females played against a *male* confederate. Subjects in the Blascovich et al. (1978) experiment, however, were paired male–male and female–female and also exhibited a Sex × Outcome interaction in heart rate. It is worth noting that the difference between the female winners and losers in heart rate response was larger in the present experiment than in the Blascovich et al. experiment. Thus, the effect of females playing against an opposite-sex opponent appeared to be to accentuate the interaction in heart rate reported in both experiments.

Conclusions

First, men and women differed in cardiac response to someone else's aggressiveness, women responding more strongly to a competitive opponent and men more to a cooperative opponent. This suggests that competitive mixed-motive games have a different sociophysiological significance for men and women.

Second, cardiovascular events during mixed-motive games may help clarify motivational aspects of these games and aid in testing hypotheses about social and interpersonal contributions to heart disease.

References

Back, K., & Bogdonoff, M. An interpretation of experimental conformity through physiological measures. *Behavioral Science*, 1963, *8*, 34.

Blascovich, J., Nash, R. F., & Ginsburg, G. P. Heart rate and competitive decision making. *Personality and Social Psychology Bulletin*, 1978, *4*, 115–118.

Costell, R. M., & Leiderman, P. H. Psychophysiological concomitants of social stress: The effects of conformity pressure. *Psychosomatic Medicine*, 1968, *30*, 298–310.

Davis, J. H., Laughlin, P. R., & Komorita, S. S. The social psychology of small groups: Cooperative and mixed-motive interaction. *Annual Review of Psychology*, 1976, *27*, 501–541.

Friedman, M., & Rosenman, R. H. *Type A behavior and your heart*. New York: Knopf, 1974.

Frodi, A., Macaulay, J., & Thome, P. R. Are women always less aggressive than men? A review of the experimental literature. *Psychological Bulletin*, 1977, *84*, 634–660.

Henry, J., & Ely, D. Biologic correlates of psychosomatic illness. In R. Grenell & S. Galay (Eds.), *Biological foundations of psychiatry*. New York: Raven Press, 1976.

Jenkins, C. D. *The Jenkins Activity Survey for Health Prediction*. Chapel Hill, N.C.: C. D. Jenkins, 1965.

Jenkins, C. D., Zyzanski, S. J., & Rosenman, R. H. Toward a redefinition of the coronary-prone behavior pattern: Inferences from test items which predict future disease. *Psychosomatic Medicine*, 1974, *36*, 464–465. (Abstract)

Komorita, S. S. Cooperative choice in a Prisoner's Dilemma game. *Journal of Personality and Social Psychology*, 1965, *2*, 741–745.

Krantz, D., Glass, D., & Snyder, M. Helplessness stress level and the coronary-prone behavior pattern. *Journal of Experimental Social Psychology*, 1974, *10*, 284–300.

Lee, R. B., & DeVore, I. *Kalahari hunter-gatherers*. Cambridge, Mass.: Harvard University Press, 1976.

Obrist, P., Gaebelein, C., Teller, E., Langer, A., Light, K., Grignolo, A., & McCubbin, J. The relationship among heart rate, carotid dp/dt, and blood pressure in humans as a function of type of stress. *Psychophysiology*, 1978, *15*, 102–115.

Rapoport, A., & Chammah, A. M. *Prisoner's Dilemma*. Ann Arbor: University of Michigan Press, 1965.

Rosenman, R. H. The role of behavior patterns and neurogenic factors in the pathogenesis of coronary heart disease. In R. S. Elliot (Ed.), *Stress and the heart*. Mount Kisco, N.Y.: Futura, 1974.

Rosenman, R. H., Friedman, M., Straus, R., Wurm, M., Jenkins, C. D., & Messinger, H. Coronary heart disease in the Western Collaborative Group study. *Journal of American Medical Association*, 1966, *195*, 130–136.

Thomas, C. B., & Ross, D. C. Observations on some possible precursors of essential hypertension and coronary artery disease: VIII. Relationship of cholesterol level to certain habit patterns under stress. *Bulletin of John Hopkins Hospital*, 1963, *113*, 225–238.

Tiger, L., & Fox, R. *The imperial animal*. New York: Dell, 1971.

Van Egeren, L. F. Social interactions, communications, and the coronary-prone behavior pattern. A psychophysiological study. *Psychosomatic Medicine*, 1979, *41*, 2–18.

Van Egeren, L. F., Abelson, J. L., & Thornton, D. W. Cardiovascular consequences of expressing anger in mutually dependent relationships. *Journal of Psychosomatic Research*, 1978, *22*, 537–548.

Wilson, E. O. *On human nature*. Cambridge, Mass.: Harvard University Press, 1978.

Received January 3, 1978

Journal of Personality and Social Psychology
1979, Vol. 37, No. 6, 865–878

Vocal Interruptions in Dyadic Communication as a Function of Speech and Social Anxiety

Michael Natale
Franklin and Marshall College

Elliot Entin
Ohio University

Joseph Jaffe
New York State Psychiatric Institute

Interruptions have been defined as a breach of the "turn-taking" contract in interpersonal communication. The relation between a speaker's personality and his or her propensity to interrupt was examined in 30-min unstructured conversations for 36 dyads (12 male, 12 female, and 12 mixed sex). The following predictions were made: (a) Interruptive behavior is inversely related to speech anxiety and positively related to confidence as a speaker; (b) interruptive behavior is inversely related to social anxiety (avoidance–distress; fear of negative evaluation). A stepwise multiple regression analysis was performed, controlling for the systematic effects of sex, the conversational partner's personality and amount of speech, and the speaker's use of back-channel responses. These hypotheses were confirmed for rate of total interruptions and rate of successful interruptions, for percentage of successful interruptions, and for mean duration of interruptions; the results withstood cross-validation analysis.

Ervin-Tripp (1964, 1969) and Hymes (1972) have proposed a taxonomy of the "situational determinants" of speech choice, with situational determinants being primarily defined as either endogenous or exogenous. Many studies (see Giles & Powesland, 1975, for a review) have shown speech modifications to be affected by topic, general context, formality–informality, and so forth, thus demonstrating the influence of external factors. However, the influence of internal variables, particularly personality states, on the use of certain patterns of speech has not been researched to any great extent. The few studies that have directly examined the relation of personality to conversational style have used form, as opposed to content, in constructing measures of vocal behavior. Markel, Bein, Campbell, and Shaw (1976) have shown amount of speech to be positively related to extroversion. In a series of studies, Natale (1975a, 1975b, 1976b) has demonstrated that the magnitude of engaged "matching" (modeling) for response latency, pause duration, and vocal intensity is directly related to a person's need for social approval (Social Desirability scale, Crowne & Marlowe, 1960). The general purpose of the present investigation was to further explore the relation of personality characteristics to speech behavior in an unconstrained and freewheeling conversation.

To pursue this research goal, some salient aspect of conversational behavior should be examined. Investigators have suggested that alternation of speakers is an integral part of dyadic communication that is governed by rules (Jaffe & Feldstein, 1970; Schegloff, 1968; Yngve, 1970). Empirical support of the conversational "turn-taking" mechanism has been provided by Duncan (1972, 1975), who demonstrated that behaviors in several communication modalities actively served as turn-taking signals. A relevant index of speaker

The authors would like to express appreciation to Louis Driever and Eric Bredeson for their efforts in the data collection. The also wish to thank the anonymous reviewers whose comments were extremely helpful in preparing the article.

Requests for reprints should be addressed to Michael Natale, Department of Psychology, University of Pennsylvania, 3815 Walnut Street, Philadelphia, Pennsylvania 19174.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3706-0865$00.75

switching is the interruption. At first glance, an interruption might be construed as always representing a battle for the conversational floor, but some evidence runs counter to that notion: (a) Stephenson, Ayling, and Rutter (1976) have reported that visual access during conversation promotes spontaneity and is associated with increased speech interruptions (Rutter & Stephenson, 1977), and (b) Gallois and Markel (1975) have provided evidence suggesting that speech interruptions may have different psychological relevance (e.g., dominance or discomfort) during different phases of a conversation. In other words, "it would be a mistake . . . to infer that each interruption event is a miniature battle for ascendancy" (Meltzer, Morris, & Hayes, 1971, p. 392). On the other hand, it may serve heuristic purposes to consider interruptions as a breach of the turn-taking process (Schegloff, 1968). If simultaneous speech is viewed as a violation of norms regulating conversational exchange, it can be expected that anxiety associated with interpersonal communication would be inversely related to an individual's propensity to interrupt. The purpose of the present study was to test this "commonsense" hypothesis in an attempt to explore the psychological factors contributing to interruption as a salient form of speech behavior.

There have been several previous studies that attemped to relate simultaneous speech to personality, but the results are not conclusive. A recent study by Feldstein, BenDebba, and Alberti (Note 1) has shown that the frequency of initiated simultaneous speech is a stable and characteristic vocal feature of an individual. Gallois and Markel (1975) published an experiment that was purported to examine the relation of social personality to conversational style. However, the only significant finding of the study was that the frequency of successful interruptions changed over the course of spontaneous conversations, and nothing was reported concerning the personality of the interrupter. We know of only two previous studies that directly assessed the relation between personality and conversational style for simultaneous speech. Feldstein, Alberti, BenDebba, and Welkowitz (Note 2) used Cattell's Sixteen Personality Factor Questionnaire (16PF) to measure dimensions of personality and found that "persons who are relaxed, complacent, secure, and not positively dependent on the approval of others tend to initiate more simultaneous speech than those who are generally apprehensive, self-reproaching, tense, and frustrated." A recent study by Ferguson (1977) examined the relation of dominance to the propensity to interrupt and found that interruptive speech was not affected by the dominance measure. The results of Feldstein et al.'s study (Note 2) are inconclusive for two reasons: (a) The personality descriptions were based on factor scores from the 16PF, yet it has been emphasized by Bourchard and Porer (cited in Buros, 1972, p. 329) that because of weak predictive validity, personality assessment from Cattell's scales or profile interpretation is not desirable; (b) the results of Feldstein et al. were obtained from an uncross-validated regression analysis of only 24 subjects, thus obscuring the generalizability of the findings.

To sum up, Feldstein et al.'s (Note 2) data do suggest that tense and apprehensive people are less likely to initiate simultaneous speech. This tentative finding is consonant with the hypothesis of the present study: The more fear individuals have concerning their speaking role, (i.e., speech anxiety), the less likely it is that they will interrupt. This notion was tested in unconstrained conversations of previously unacquainted individuals so that "task" factors were minimized. We also "covaried out" the effects of potential moderator variables (speaking style and personality of dyadic partner) so that the relation of an individual's personality to interruption behavior could be distinctly assessed.

Method

Subjects

The subjects were 72 students (ages 18 to 20) of Ohio University enrolled in introductory psychology courses who volunteered and who received course credit for their participation. The subjects were randomly paired: 12 male dyads, 12 female dyads, and 12 mixed-sex dyads. It was also a necessary condition that the dyadic partners not be previously

acquainted with each other and that English be their native language.

Personality Measures

Self-reports of speech anxiety. Four self-report measures were selected to assess anxiety during conversation and the generalization of anxiety to other social situations.

1. An overall self-report rating of General Speech Anxiety on a scale from 1 (no anxiety) to 9 (extreme anxiety). This simple measure has been successfully used by Fremouw and Harmatz (1975), and because of its face validity we deemed it appropriate.

2. The Personal Report of Confidence as a Speaker (Paul, 1966) requires the subject to complete 30 true–false items that assess the respondent's feelings, thoughts, and behavior during the most recent speech. This scale has been shown to be both reliable and valid (Paul, 1966, p. 48).

3. The Social Avoidance and Distress scale consists of 28 items that assess an individual's generalized anxiety in many interpersonal situations. Watson and Friend (1969) demonstrated this scale to have average reliability ($r = .85$) and to predict social-avoidance behavior.

4. The Fear of Negative Evaluation scale (30 items) measures the magnitude of a subject's concern about other people's evaluations of him. Watson and Friend (1969) showed this scale to have acceptable reliability ($r = .72$) and to predict social-avoidance behavior.

All four of these measures can be used to assess speech anxiety (Fremouw & Harmatz, 1975); however, it should be noted that the speaker anxiety and confidence measures, Scales 1 and 2, evaluate anxiety directly associated with the *speech act,* whereas Scales 3 and 4 assess general anxiety associated with the *communication situation.*

Social desirability. A measure of the need for approval was obtained by using the Marlowe–Crowne scale (Crowne & Marlowe, 1960). Although not directly related to the main hypothesis of the present study, this scale was administered because prior research (Natale, 1976b) suggested that an individual's need for social desirability was related to his propensity to interrupt.

Definition of Speech Interruption Behavior

For the purposes of the present study, an interruption was defined as the occurrence of simultaneous speech and was assigned to the participant who initiated speech while not possessing the "conversational floor." It should be noted that for any one individual, possession of the conversational floor begins with the first sound that person utters alone and ends with the first unilateral sound uttered by the other participant. This definition of speech interruption has been used by other investigators (Jaffe & Feldstein, 1970; Meltzer et al., 1971).

Jaffe and Feldstein (1970) summarized a decade of interview research that clearly demonstrated the ability of an analogue-to-digital computer (Automatic Vocal Transacter Analyzer—AVTA) to measure the noise and silence pattern of recorded dialogue. Using AVTA, speech was measured at 300-msec intervals, and a separate analysis of speech parameters was provided for each subject in a dyad. The interested reader is referred to Jaffe (1970) for a sophisticated discussion of the methodology and validity of speech measures derived from the AVTA system. There is, however, an issue of measurement that deserves note. There is a possibility that brief utterances, such as "yes" or "uh huh," which the listener intersperses during the speaker's pauses and speech may be classified by the computer (AVTA) as interruptions.[1] These brief utterances (back-channel responses) are easily recognized and have been shown to function as the opposite of a speaker-switching device, in that they appear to ensure that the speaker who is holding the conversational floor continues to do so (Duncan, 1975). These vocal back-channel signals are not attempts to change the conversational floor and therefore should not be included in the analysis of interruptions (Meltzer et al., 1971, p. 395).[2] To this end, back-channel responses were scored by a trained rater. Each recorded dialogue was heard twice by this judge so that each conversationalist's back-channel responses were scored separately. Concerning the reliability of the trained judge's scoring of back-channel responses, a pilot showed the rater to have a 91% agreement with one of the experimenters (MN). It should be noted that back-channel responses were the only speech measure not obtained from AVTA.

Total speech interruptions. An interruption was defined as the occurrence of simultaneous speech and was assigned to the participant who initiated speech while not possessing the conversational floor. Possession of the conversational floor begins with the first sound a subject utters alone and ends with the first unilateral sound uttered by the conversational partner. As a measure, *total speech interruptions* does not discriminate between interruptions that result in a change of the conversational floor and those in which the interrupted speaker retains possession of conversational floor.

Successful speech interruptions. A speech interruption that results in a speaker switch is a successful interruption. This measure is assigned to the speaker who did not originally possess the conversational floor.

Percentage of successful interruptions. It should be noted that unsuccessful and successful interruptions summate to total speech interruptions. The

---

[1] The vocalizations have been variously called "listener responses" (Dittman & Llewellyn, 1967), "attention signal" (Kendon, 1967), and backchannel responses (Duncan, 1975; Yngve, 1970). In this article we employ the latter term.

[2] The authors thank Robert Krauss, Department of Psychology, Columbia University, for clarifying this issue.

proportion of total interruptions that were successful (i.e., that resulted in a speaker switch) was calculated.

Mean duration of interruption. Jaffe and Feldstein's (1970) AVTA allows for the calculation of duration of interruption in .1-sec units. The mean duration was calculated on each subject's total speech interruptions.

Total speech time. The total vocalization time of each conversationalist was calculated, via AVTA, using 300-msec recording intervals.

Apparatus

The conversation between the pair of subjects took place in one of two conditions: (a) with the conversationalists in the same room (10×18 ft., or 3×5 m), sitting 15 ft. apart in half-view of each other (from the waist up); (b) with the subjects in separate rooms but in half-view of each other, sitting 2 ft. from a full window and directly facing each other. Approximately half of the conversations were recorded in each condition.[3]

All subjects wore a headset consisting of a Kellogg earset and a Sony omnidirectional microphone that rested approximately 1 in. from the mouth of the subject. The close proximity of the microphone allowed us to maintain a low gain on the recorder, the result of this procedure being that each conversationalist's speech was clearly recorded without spillover from his dyadic partner. The conversationalists heard each other monophonically via the earset. The recordings were made on a Sony stereo tape recorder that was out of view of the subjects.

As stated, Jaffe and Feldstein's (1970) AVTA system was used to obtain all the speech measures with the exception of back-channel responses, which were scored by an independent rater. AVTA is an analogue-to-digital computer that accepts stereo recordings of a dialogue and provides the previously described speech parameters for each conversationalist.

Procedure

Prior to the conversational task, a pair of subjects was asked to fill out the personality measures described earlier (25 to 40 min); then some preliminary instructions were given to ensure proper tape-recording. The subjects were told that this was merely an experiment on interpersonal communication and that the only experimental task was to talk freely with their partner. At this point, the research assistant activated the tape recorder and left the room. Each recorded dialogue lasted exactly 30 min, whereupon the research assistant returned to the room, terminated the conversation, and explained the purpose of the study to the participants.

Statistical Analysis

The 36 dyads were randomly divided into two equal-sized samples, with the stipulation that each sample have an equal number of male, female, and mixed-sex dyads. One was called the initial sample and the other the cross-validated sample. All analyses were first conducted on the initial sample of 18 dyads.[4] In case the regression analysis (described later) capitalized upon chance, the results were validated using the cross-validated sample.

Stepwise regression. For the initial sample, a stepwise regression analysis was used to evaluate the contribution of an individual's self-reported speech anxiety and desire for social approval (predictors) to the use of the following interruption behaviors (dependent variables): total speech interruptions, successful speech interruptions, and mean duration of speech interruptions. A separate regression analysis was performed for each dependent variable.

Since the purpose of the present study was to determine the precise effect of specific personality measures on an individual's propensity to interrupt, it was necessary to eliminate the effects of the following modifying variables in the regression analysis for the initial sample. (a) It is obvious that an individual will have greater opportunity to interrupt a talkative partner and less opportunity to interrupt a taciturn partner; therefore, the effect of the partner's total speech time should be eliminated in evaluating the relation of an individual's personality to his or her propensity to interrupt. (b) As previously mentioned, certain brief interruptions (back-channel responses) are not intended by the speaker to indicate a desire to grab the conversational floor; hence, any study wishing to examine interruptions as a speaker-switching mechanism should eliminate the role of back-channel responses (Melzer et al., 1971, p. 395). (c) A *person perception* variable may also influence an individual's behavior in the interpersonal situation; therefore, the personality traits of the dyadic partner should be taken into account when assessing the individual's behavior. (d) It has been suggested that the individual's sex and the dyad's sex affect the use of interruptions (Zimmerman & West, 1975); hence, these sex variables must be taken into account.

In this light, the stepwise regression of an individual's personality measures to his or her interruptive behavior was performed using the individual's back-channel responses, the speaking time and personality measures of the conversational partner, and the individual's and the dyad's (male, female, mixed) sex as covariates.

Marsden, Kalter, and Ericson (1974) have pointed out that this analysis can easily be achieved by forc-

---

[3] The statistical technique used in this study was stepwise multiple regression. This regression procedure capitalizes on the particular characteristics of the sample; therefore, a slight variation in the recording procedures was viewed by the experimenters to provide robustness to the findings.

[4] There were 18 dyads in the initial sample, but each dyad member was treated as a subject in the analysis. Hence, $N = 36$ for the initial sample.

ing the order of the predictors. Stepwise multiple regression works by adding one variable at a time to the regression while assessing to determine whether the independent variables significantly increase the prediction of the dependent variable. If one wants to treat certain predictors as covariates, all that is necessary is to force the entry of these covariates into the multiple regression prior to the other free-floating predictors (see Cohen & Cohen, 1975, for a detailed discussion of this technique). In other words, there were two sets of independent variables used in the stepwise regression procedure: a set of covariates followed by a set of predictors of theoretical value (variates). The order of the independent variables within each set of predictors was "free," that is, selected by the computer to maximize prediction of the dependent variable. This straightforward procedure was used in the present study. To restate, the multiple regression analysis tested the relation between an individual's personality and speech-interruption behavior, while controlling for the systematic effects of sex, of the conversational partner's personality and amount of speech, and of the speaker's use of back-channel responses.

Cross-validation technique. As stated earlier, a stepwise regression was used in the initial sample. This statistical procedure adds (steps) one variable at a time to the regression for the purpose of increasing the prediction of the criterion. When the computer is free to determine the order of the predictors, it is possible that the stepwise procedure will capitalize on chance relationships. In the present study, the predictors were grouped as two sets (covariates and variates), and the computer freely determined their order *within the set.* Thus a cross-validation was not absolutely necessary but was performed for the purpose of testing the replicability of the findings.

The regression formula (beta weights) from the first sample was imposed on the independent variables of the second, cross-validation, sample. This procedure produces *predicted* values of the dependent variable in the cross-validated sample. For the second sample, a zero-order correlation between the estimated and the actual dependent variables was performed at each step of the regression and was considered the cross-validated multiple regression coefficient.

Results

Table 1 presents the initial stepwise correlations obtained for Sample 1. A separate regression analysis was performed for each dependent variable. The set of covariates was entered as predictors prior to the set of variates, but within each set the computer was free to order the predictor so as to maximize prediction. The regression of each independent variable to the dependent variable was tested for statistical significance by an F ratio (Cohen & Cohen, 1975, p. 107) that took into account the effect of other independent variables. In other words, the F test indicated whether the increment in variance accounted for was significant when a predictor was added.

The multiple correlations obtained in Sample 1 (Table 1) were cross-validated in a second sample. The cross-validation procedure (see Method) used all of the predictors (covariates and variates), but for the sake of brevity, the cross-validated coefficients are reported (see Table 2) only for the significant predictors of the first sample.

Total Interruptions

In reference to the total amount of interruptions initiated by a speaker in the dyadic situation, the total speaking time of the conversational partner was a strong positive factor ($r = .79$, 63% of the variance). Obvious logic explains this finding—the more a person talks, the more likely it is that he or she will be interrupted. Examination of Table 1 also shows another covariate to have had an effect on a speaker's use of interruptions. It was found that males interrupted more often than females (correlation increased from .79 to .83; additional variance accounted for was 7%), a finding that Zimmerman and West (1975) have also reported. There was no significant contribution of the conversational partner's personality traits to the speaker's propensity to interrupt. Of direct relevance to the hypothesis of the present study, examination of the variates in Table 1 shows the speaker's desire for social approval to be positively related to the total use of interruptions.

Examination of Table 2 shows that even after a cross-validation procedure a person's desire for social approval was still shown to be significantly related to his or her use of interruptions. The relationship of partner's speaking time and sex to rate of vocal interruptions also withstood cross-validation (see Table 2).

Successful Interruptions

This dependent measure is an expression of interruptions that result in the interrupter gaining the conversational floor. In this sense,

Table 1
Stepwise Multiple Regression Relating Personality Characteristics to Engaged Speech Interruption

Variable[a]	Multiple regression analysis		
	R[b]	R^2	R^2 increment
Total interruptions (criteria)			
Covariates			
P's speech time	+.790	.625	.625****
S's sex	−.834	.696	.071**
P's avoidance–distress	+.859	.722	.026
P's negative evaluation	+.863	.754	.022
Dyad's sex	.871	.759	.014
S's back channel	+.876	.769	.010
P's need for social approval	−.884	.781	.013
P's speech anxiety	−.885	.783	.002
P's confidence	+.885	.784	.001
Variates			
S's need for social approval	+.911	.829	.045*
S's speech anxiety	−.913	.837	.003
S's confidence	+.918	.843	.011
S's negative evaluation	−.921	.847	.004
S's avoidance–distress	−.921	.848	.001
Successful interruptions (criteria)			
Covariates			
S's back channel	+.466	.217	.217***
P's speech time	+.522	.272	.055*
P's need for social approval	+.561	.315	.043
P's confidence	−.592	.350	.035
P's speech anxiety	−.600	.360	.010
S's sex	.612	.375	.015
P's negative evaluation	+.616	.379	.004
P's avoidance–distress	−.622	.387	.008
Dyad's sex	.623	.387	.000
Variates			
S's speech anxiety	−.754	.568	.180***
S's need for social approval	+.781	.611	.043
S's negative evaluation	−.802	.643	.032
S's avoidance–distress	−.805	.649	.006
S's confidence	−.806	.649	.001
Percentage of successful interruptions (criteria)			
Covariates			
P's confidence	+.425	.179	.179*
P's speech anxiety	−.505	.255	.077
P's avoidance–distress	+.527	.277	.022
P's need for social approval	−.534	.286	.008
P's back channel	−.538	.289	.004
S's sex	.541	.292	.003
P's speech time	+.543	.295	.003
P's speech anxiety	−.545	.297	.001
Dyad's sex	.545	.297	.000

Table 1 (*continued*)

Variable[a]	Multiple regression analysis		
	R[b]	R^2	R^2 increment
Variates			
S's avoidance–distress	−.696	.484	.187**
S's need for social approval	+.712	.507	.023
S's speech anxiety	−.716	.512	.005
S's negative evaluation	−.716	.513	.001
S's confidence	+.716	.513	.000
Mean duration of interruption (criteria)			
Covariates			
P's speech time	+.645	.415	.416****
S's sex	.678	.459	.044
Dyad's sex	.700	.489	.030
P's speech anxiety	−.708	.502	.011
P's confidence	+.715	.512	.010
P's negative evaluation	+.729	.532	.020
P's need for social approval	−.730	.533	.001
S's back channel	+.731	.534	.001
S's avoidance–distress	+.731	.534	.000
Variates			
S's avoidance–distress	−.816	.666	.132***
S's need for social approval	+.845	.713	.047
S's confidence	+.848	.719	.006
S's negative evaluation	−.848	.719	.000
S's speech anxiety	−.848	.719	.000

Note. Sample 1, $n = 36$. A separate multiple regression analysis was performed for each dependent variable (criterion). The ability of each independent variable to make a unique nonzero contribution to R^2 was tested by examining the semipartial correlation (sr) of each predictor to the criteria as follows:

$$= \frac{sr^2/1}{(1 - R^2)/N - k - 1}$$

where N = number of subjects (36) and k = number of independent variables; $df = 1\ (N - k - 1)$ (Cohen & Cohen, 1975, p. 107).

[a] P denotes partner and S denotes speaker.

[b] + or − indicates the direction of the relationship of each predictor to the dependent variable.

* $p < .05$.
** $p < .025$.
*** $p < .01$.
**** $p < .001$.

a multiple regression with this dependent measure provides a more stringent test of the hypothesized relations between personality and interruptive behavior. Looking at Table 1, one sees that a person's use of back-channel responses was significantly related to his or her overall rate of successful interruptions ($r = .46$, 21% of the variance). It was also the case that the conversational partner's amount of speech was positively related to a speaker's use of successful interruptions (multiple R increased from .47 to .52). The personality characteristics of the conversational partner were not significantly related to the magnitude of successful interruptions used by a speaker.

Over and above the effects of covariates on the incidence of successful interruptions, the hypothesized effect of the speaker's personality was demonstrated. Table 1 shows a speaker's general speech anxiety to be inversely related to his or her rate of successful

interruptions; the increase in the accounted for variance was rather large (18%). Table 2 shows this finding to be generalizable under cross-validation procedures. The positive relation of a speaker's back-channel responses to the frequency of successful interruptions used by the same speaker was marginally significant under cross-validation procedures ($r = .29, p < .07$).

Percentage of Successful Interruptions

This measure reflects the proportion of initiated interruptions that result in a speaker switch in favor of the interrupter. Table 1 shows that a speaker's percentage of successful interruptions was positively related to his or her conversational partner's confidence as a speaker ($r = .43$, 18.7% of the variance). The speaker's general anxiety about social situations (avoidance–distress) was negatively related to his or her propensity to effect a speaker switch once he or she initiated an interruption. In other words, a person who has high social anxiety is less likely to persist in his or her interruptions. It should be noted that the speaker's social avoidance–distress accounted for a substantial amount of the variance over and above the effects of the covariates (8.75%, see Table 1). This relation of percentage of successful interruptions to speaker's avoidance–distress was also cross-validated (see Table 2).

Mean Duration of Interruption

This measure can be conceptualized as an index of a person's tenacity in effecting a speaker switch. In other words, when an individual initiates an interruption does that individual prolong or shorten the simultaneous speech situation and how does this relate to his or her personality? Before answering this question, we look at the effect of the covariates on this dependent measure. Table 1 shows the conversational partner's amount of speech to be positively related to how long an individual persists in an interruption ($r = .64$, 41% of variance, Table 1). There was a nonsignificant trend ($p \leq .08$) for males to engage in longer interruptions (Table 1). This effect of sex on interruptive behavior agrees with the findings of Zimmerman and West (1975).

Table 2
Cross-Validated Multiple Regression Correlations Between Speech Interruptions and Personality

Dependent variable	Multiple regression coefficient[a]
Total interruptions	
P's speech time	.35*
S's sex	.41**
S's need for social approval	.44**
Successful interruptions	
S's back channel	.29
S's speech anxiety	.36*
Percentage of successful interruptions	
P's confidence	.21
S's avoidance–distress	.38*
Mean duration of interruption	
P's speech time	.39*
S's avoidance–distress	.45**

Note. Sample 2, $n = 36$. The product-moment correlation between the predicted and the observed dependent variable at each step of the regression was defined as the cross-validated regression coefficient ($df = 34$, two-tailed test). Only those covariates and variates that showed significant predictive power in the initial regression (see Table 1) were used in the above analysis.

[a] P denotes partner and S denotes speaker.

* $p < .05$.

** $p < .01$.

The hypothesized effect of personality on interruption was again supported by the regression analysis. Looking at the variates in Table 1, we see that an individual's general social anxiety (avoidance–distress) was negatively related to the mean duration of the interruptions he or she initiated. The predicted variance contributed by this variable was 13.2%.

Total Speech Time

The relation of personality to speaking time was not directly relevant to the hypothesis of the present study. However, a multiple regression analysis was performed for the sake of completeness, since the information was available. Table 3 shows that the sex and the personality characteristics of the conversa-

Table 3
Stepwise Multiple Regression Relating Personality to Speaking Time and Back-Channel Responses

Variable[a]	Multiple regression analysis		
	R[b]	R^2	R^2 increment
Total speech time (criteria)			
Covariates			
Sex	.283	.080	.080
P's confidence	+.360	.129	.049
P's avoidance–distress	−.520	.270	.141*
P's need for social approval	+.530	.280	.010
S's back channel	+.533	.284	.003
P's negative evaluation	−.534	.285	.001
Dyad's sex	.535	.287	.002
P's speech anxiety	+.536	.287	.000
Variates			
S's speech anxiety	−.707	.499	.212**
S's avoidance–distress	+.717	.515	.015
S's confidence	+.745	.555	.040
S's negative evaluation	+.747	.559	.004
S's need for social approval	−.749	.561	.003
Back-channel responses (criteria)			
Covariates			
P's confidence	−.349	.122	.122
P's anxiety	−.461	.213	.091
P's avoidance–distress	−.528	.278	.065
P's need for social approval	−.552	.304	.027
Dyad's sex	.576	.332	.028
P's negative evaluation	−.583	.339	.007
Sex	.584	.341	.002
Variates			
S's negative evaluation	+.697	.490	.144**
S's confidence	+.707	.500	.013
S's speech anxiety	+.723	.506	.015
S's avoidance–distress	+.735	.522	.022
S's need for social approval	+.736	.522	.000

Note. Sample 1, $n = 36$. R^2 increments were tested for each predictor by assessing the significance of each semipartial coefficient (Cohen & Cohen, 1975, p. 107). See note to Table 1 for further details.
[a] P denotes partner and S denotes speaker.
[b] + or − indicates the direction of the relationship of each predictor to the dependent variable.
* $p < .025$.
** $p < .01$.

tional partner were poorly related to the total amount of speech time of a speaker. Looking at the variates in Table 3 it is apparent that the subjects' speech anxiety was inversely related to their speech time.

Back-Channel Responses

Although research has suggested that certain brief interjections are "listener responses" (Duncan, 1975), their relation to the personality characteristics of the listener or the speaker has not been examined. Data from the present study allowed us to address this question. Table 3 shows a person's use of back-channel interruptions to be related uniquely to the personality of the individual who is emitting the listener response. The conversational partner's personality was not significantly related to the use of back-channel responses by an individual (see Table 3). A speaker's use of back-channel responses was found to be positively related to his or her fear of negative evaluation; 14% of the variance was accounted for by this variable (Table 3). Table 4 shows this finding to withstand cross-validation.

Discussion

Specific Findings

In evaluating the meaningfulness of the results, two questions should be kept in mind. (a) When a predictor shows a significant relation to the dependent variable, what is the amount of variance accounted for? In other words, even if the addition of a variable to the multiple regression increases R^2 significantly, the increase may be quite small. (b) Does the significance of the predictors sustain cross-validation procedures? With these points in mind, a discussion follows.

Concerning the total rate of vocal interruption that a person displays (see Table 1), the amount of speaking time of a conversational partner is a strong determinant. The more people talk, the greater their chance of being interrupted. This very basic finding indicates the limits of a psychological explanation of speech interruptions. In other words, the present study shows that roughly two thirds of all vocal interruptions can be explained in terms of physical opportunity as opposed to personality factors. Within this limit, a significant but small proportion of total interruptions was predicted by the sex of the speaker; males interrupted more often than females. This effect of sex on interruptive behavior verifies the earlier findings of Zimmerman and West (1975), which showed males to interrupt more often than females.

Personality characteristics of the conversational partner did not significantly contribute to the total use of interruptions. This finding, together with the positive relation of a speaker's desire for social approval to vocal interruption behavior (see Table 1), may indicate the inappropriateness of conceptualiz-

Table 4
Cross-Validated Multiple Correlations Between Total Speech Time Back-Channel Responses and Personality

Dependent variable	Multiple regression coefficient[a]
Total speech time	
P's avoidance–distress	.38*
S's speech anxiety	.43**
Back-channel responses	
S's negative evaluation	.51**

Note. $n = 36$. The product-moment correlation between the predicted and the observed dependent variable at each step of the regression was defined as the cross-validated regression coefficient ($df = 35$, two-tailed test). Only those covariates and variates that showed significant predictive power in the initial regression (see Table 3) were used in the above analysis.

[a] P denotes partner and S denotes speaker.

* $p < .05$.

** $p < .01$.

ing speech interruptions as invariably representing a contest for the conversational floor. In the introduction we noted that speech interruptions correlate positively with intimacy (e.g., Rutter & Stephenson, 1977). This interpretation of speech interruptions may explain why a person who has a high need for social approval tends to interrupt more often. Interruptions, rather than representing a violation of speaker-switching roles, may serve to express joint enthusiasm and the need for an individual to achieve this level of communication (a measure of social desirability).

However, simultaneous speech may represent conversational floor grabbing in the situations where successful interruptions were examined. Support for this notion comes from the multiple regression analysis (Table 1) that used successful interruptions as the dependent variable: An individual's speech anxiety had a sizable negative relation to the number of successful interruptions. In this light, partial support for the original hypothesis can be stated as follows: Vocal interruptions represent an "offense" against the turn-taking rules and as such are negatively related to an individual's speech anxiety, but only in the case of successful interruptions. This finding adds validity to the notion that certain brief interjections (i.e., back-channel responses) are emitted for the sole purpose of reinforcing the listening role. The data in Tables 3 and 4 show fear of negative evaluation to be significantly related to the use of back-channel responses. Specifically, a person's fear of negative evaluation was positively related to his or her propensity to utter brief interruptions (i.e., back-channel responses). The positive social function of back-channel responses is apparently strong enough to answer to a person's need to be positively evaluated. The positive relationship between a person's back-channel responses and his or her use of successful interruptions is not readily interpretable, thus indicating the need of future research to clarify the psychological role of back-channel responses.

The percentage of successful interruptions can be taken as a measure of a person's intensity in forcing his or her way into a conversation. After initiating an interruption, does an individual continue until he or she takes over the conversational floor? Interestingly, the percentage measure of interruptions yielded a solid relation to the partner's personality (see Table 1). The more confident the partner felt about speaking, the higher the proportion of successful interruptions by the other subject (approximately 18% of the predicted variance was accounted for by the partner's speech confidence). One possible interpretation of this finding is that the partner communicates a positive attitude toward interpersonal communication and thus somehow gives an interrupter more confidence to attempt speaker switches. Conversely, if we perceive a person to be highly sensitive about conversing, we might avoid trying tenaciously to grab the conversational floor. Future research will have to clarify the role of person perception variables in the use of simultaneous speech.

The intensity measure (percentage successful) of interruptions was strongly related to a speaker's social anxiety. The more anxiety (avoidance–distress) a person has concerning social interaction, the smaller his or her proportion of successful interruptions. This finding offers direct support of the general hypothesis of the present study: The use of speech interruptions was negatively related to interpersonal anxiety.

The strong relation between partners' speaking time and interruption duration (see Table 1) is self-explanatory. The duration of the interruption was negatively related to social anxiety (avoidance–distress), thus replicating a similar relation between social anxiety and the percentage of successful interruptions. The findings presented in Tables 1 and 2 can be summed up as follows:

1. A significant negative relation between interruption and a speaker's interpersonal anxiety was demonstrated for all interruption measures with the exception of total interruptions. This finding indicates that analysis of speech interruptions necessitates the delineation of the types of simultaneous speech.

2. The fact that a person's desire for social approval had a modest but significant positive relation to interruptive behavior indicates that speech interruptions may not always

represent contests for the conversational floor. An interruption can also express a positive state of excitement and the like.

3. The findings that males engaged in vocal interruption more than females did fully supports the earlier research of Zimmerman and West (1975).

4. The present findings indicate that the partner's speech time was strongly related to a subject's interruptive rate. The more people talk, the more likely they will be interrupted.

It was also found (Tables 3 and 4) that a person's speech anxiety was negatively related to total speaking time. To the extent that speech anxiety determines speaking time of an individual, it also affects the interruption rate of the conversational partner. The personalities of the dyadic partners interact and influence each other's behavior, but it was the speaker's, as opposed to the partner's, personality that was the strongest predictor of vocal interruptions.

General Theoretical Relevance

Of general theoretical interest is the relation between verbal behavior and personality. Clinical psychology has traditionally emphasized the relation of verbal content to the presence of psychopathology (e.g., thought disorder, delusions, schizophrenia, subjective unhappiness, depression). Over the past decade, researchers have tried to develop objective scoring systems for verbal behavior in an attempt to overcome some of the problems of unreliability (e.g., spectral analysis of voice for anxiety, Smith, 1974; interaction chronography, Jaffe & Feldstein, 1970, and Matarazzo & Weins, 1972). Although automation of vocal measures has vastly improved reliability, the relation of these speech measures to variables of psychological interest has not been shown to be strong. This may be due to the fact that formal aspects of speech may represent an endogenous (neurolinguistic) characteristic of communication that is to a great degree invariant. Evidence for this notion comes from Jaffe and Feldstein (1970), who have shown that speech timing of monologues and dialogues can be accounted for by a Markovian (stochastic) process.[5]

If speech timing of dyadic interaction is partially predicted by a "random walk" model then it follows that psychological factors might only partially affect conversational timing. This may account for the fact that only a minor, albeit significant, proportion of speech interruptions was affected by the personality measures used in the present study. The stochastic nature of dialogic speech timing may also account for Ferguson's (1977) meager empirical support of the long-held clinical notion that interruptions represent dominance in family interactions.

However, there is some evidence that suggests that the Markovian nature of dialogic speech timing (Jaffe & Feldstein, 1970) breaks down in the course of a conversation. Cobb (1973) showed that there were long-term periodicities in the speech timing of individuals who knew each other. The presence of speech rhythms obviously violates a fundamental assumption of Jaffe and Feldstein's stochastic model; the transitional probabilities of the vocal parameters cannot be assumed to be constant. It appears that the timing of speech, and therefore interruptions, roughly obeys stochastic principles which are nonetheless modified by external variables (personality, task orientations, etc.). This was also found to be the case for gaze behavior. Natale (1976a) found the timing of gaze behaviors generally to obey Markovian principles, but it was also observed that prolonged looks could not be accounted for by the stochastic model.

In summary, it appears that although the timing of gaze and of speech (including speech interruptions) generally obeys Markovian principles, these behaviors still are susceptible to influence. The present study examined only speech interruption behavior in free conversations. It may be that a variety of situational variables are more likely to elicit vocal interruptions of a psychological nature. Feldstein and Welkowitz (1978, p. 341) have

[5] Gaze timing, often hypothesized to reflect cognitive (Rutter & Stephenson, 1977) and affective state (Natale, 1977), has also been shown to obey stochastic principles (Natale, 1976a; Rutter, Stephenson, Lazzerini, Ayling, & White, 1977).

stated that an AVTA (Jaffe & Feldstein, 1970) analysis of free conversations shows speech interruptions (successful and nonsuccessful) generally to have durations of approximately .4 sec. (The present study obtained the same finding.) Feldstein and Welkowitz have aptly noted that Meltzer et al. (1971) reported instances of simultaneous speech as long as 6 sec and suggest that the difference in findings may be a function of task. Meltzer et al. recorded speech interruptions during problem-solving sessions, whereas Feldstein and Welkowitz (1978) and the present study examined naturalistic conversations. The relevant point is that casual social conversation and task-oriented dialogue may elicit different patterns of speech interruption reflecting different relation to personality. It may be that interpersonal anxiety accounts for a larger proportion of interruption in interviews, therapy, and the like. Future research must examine these questions.

Another theoretical issue relevant to the present findings is the extent to which the outcome of simultaneous speech is determined by cognitive variables. Morris (1971) and Meltzer et al. (1971) have provided sound evidence for the role of vocal amplitude in determining interruption outcome. These investigators have suggested that the briefness of simultaneous speech indicates the small role of semantic content and hence the importance of physicalistic cues (vocal amplitude) in interruption outcome. But Feldstein and Welkowitz have stated that "amplitude may function as a mediating factor that allows speakers to respond quickly to content, or it may be so highly related to content . . . that its role in determining outcome is artifactual" (1978, p. 353). The relation of interruption outcome to cognitive processes is not ruled out and does indeed receive indirect support from the present study. Successful interruptions were negatively related to speech anxiety and social anxiety, but the accounted for variance was moderate. Also, duration of interruption, which averaged a mere .45 sec, was significantly related to personality. The fact that interruption behavior is subject to individual differences (Notes 1 and 2), and is related in some degree to personality, does suggest that the role of cognitive processes in determining interruption behavior cannot be ruled out.

Finally, the issue of whether an interruption represents a fight for the conversational floor was not directly examined in the present study. It will be recalled that for the sake of examining the relation of simultaneous speech to personality, we assumed that vocal interruptions represent a contest for speaking. However, there is evidence that suggests that interruptions represent spontaneity and positive states (Stephenson et al., 1976). This positive relation between interpersonal communication and interruptions could account for the moderate negative relation between interpersonal anxiety and the initiation of simultaneous speech. It could be that vocal interruptions are either positive (emotional) or negative (competitive) according to situational determinants. Future research should examine the relation of cognitive, mood, task, and personality variables to formal aspects of dialogue so that the magic of "good" and "bad" conversations might be discovered.

Reference Notes

1. Feldstein, S., BenDebba, M., & Alberti, L. *Distributional characteristics of simultaneous speech in conversation.* Paper presented at the meeting of the Acoustical Society of America, New York, April 1974.
2. Feldstein, S., Alberti, L., BenDebba, M., & Welkowitz, J. *Personality and simultaneous speech.* Paper presented at the annual meeting of the American Psychological Association, New Orleans, August 1974.

References

Buros, O. K. (Ed.) *The seventh mental measurements yearbook* (Vol. 1). Highland Parks, N.J.: Gryphon Press, 1972.

Cobb, L. *Time series analysis of conversations.* Unpublished doctoral dissertation, Cornell University, 1973.

Cohen, J., & Cohen, P. *Applied multiple regression: Correlational analysis for the behavioral sciences.* New York: Wiley, 1975.

Crowne, D. P., & Marlowe, D. A new scale of social desirability independent of psychopathology. *Journal of Consulting and Clinical Psychology,* 1960, *24,* 349–354.

Dittman, A. T., & Llewellyn, L. G. The phonemic clause as a unit of speech decoding. *Journal of Personality and Social Psychology,* 1967, *6,* 341–348.

Duncan, S. Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology*, 1972, *23*, 283–292.

Duncan, S. On the structure of speaker–auditor interaction during speaking turns. *Language in Society*, 1975, *2*, 161–180.

Ervin-Tripp, S. M. An analysis of the interaction of language, topic and listener. *American Anthropologist*, 1964, *66*(supplement), 86–102.

Ervin-Tripp, S. M. Sociolinguistics. *Advances in Experimental Social Psychology*, 1969, *4*, 91–165.

Feldstein, S., & Welkowitz, J. An analysis of conversational turn-taking: In defense of an objective approach. In A. W. Siegman & S. Feldstein (Eds.), *Nonverbal behavior and communication*. Hillsdale, N.J.: Erlbaum, 1978.

Ferguson, N. Simultaneous speech, interruptions and dominance. *British Journal of Social and Clinical Psychology*, 1977, *16*, 295–302.

Fremouw, W. J., & Harmatz, M. G. A helper model for behavioral treatment of speech anxiety. *Journal of Consulting and Clinical Psychology*, 1975, *43*, 652–660.

Gallois, C. G., & Markel, N. N. Turn-taking: Social personality and conversational style. *Journal of Personality and Social Psychology*, 1975, *31*, 1134–1140.

Giles, H., & Powesland, P. F. *Speech style and social evaluation*. New York: Academic Press, 1975.

Hymes, D. Models of the interaction of language and social setting. In J. J. Gumperez, & G. D. Hymes (Eds.), *Directions in sociolinguistics: The ethnography of communication*. New York: Holt, Rinehart & Winston, 1972.

Jaffe, J. Linked probabilistic finite automata: A model for the temporal interaction of speakers. *Mathematical Biosciences*, 1970, *7*, 191–204.

Jaffe, J., & Feldstein, S. *Rhythms of dialogue*. New York: Academic Press, 1970.

Kendon, A. Some functions of gaze direction in social interaction. *Acta Psychologica*, 1967, *26*, 22–63.

Markel, N. N., Bein, M. F., Campbell, W. W., & Shaw, M. E. The relationship between self-rating of expressed inclusion and speaking time. *Language and Speech*, 1976, *19*, 117–120.

Marsden, G., Kalter, N., & Ericson, W. A. Response productivity: A methodological problem in content analysis studies in psychotherapy. *Journal of Consulting and Clinical Psychology*, 1974, *42*, 224–230.

Matarazzo, J. P., & Wiens, A. N. *The interview: Research on its anatomy and structure*. New York: Aldine & Atherton, 1972.

Meltzer, L., Morris, W., & Hayes, D. Interruption outcomes and vocal amplitude. Explorations in social psychophysics. *Journal of Personality and Social Psychology*, 1971, *18*, 382–402.

Morris, W. N. Manipulated amplitude and interruption outcomes. *Journal of Personality and Social Psychology*, 1971, *20*, 319–331.

Natale, M. Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personal and Social Psychology*, 1975, *32*, 790–804. (a)

Natale, M. Social desirability as related to convergence of temporal speech patterns. *Perceptual and Motor Skills*, 1975, *40*, 827–830. (b)

Natale, M. A Markovian model of adult gaze behavior. *Journal of Psycholinguistic Research*, 1976, *5*, 53–63. (a)

Natale, M. Need for social approval as related to speech interruption in dyadic communication. *Perceptual and Motor Skills*, 1976, *42*, 455–458. (b)

Natale, M. Induction of mood states and their effect on gaze behavior. *Journal of Consulting and Clinical Psychology*, 1977, *45*, 960.

Paul, G. *Insight versus desensitization in psychotherapy*. Stanford, Calif.: Stanford University Press, 1966.

Rutter, D. R., & Stephenson, G. M. The role of visual communication in synchronizing conversation. *European Journal of Social Psychology*, 1977, *7*, 29–37.

Rutter, D. R., Stephenson, G. M., Lazzerini, A. J., Ayling, K., & White, J. Eye-contact: A chance product of individual looking. *British Journal of Social and Clinical Psychology*, 1977, *16*, 191–192.

Schegloff, E. A. Sequencing in conversational openings. *American Anthropologist*, 1968, *70*, 1075–1095.

Smith, G. A. The measurement of anxiety: A new method by voice analysis. *Medical Science*, 1974, *2*, 1707.

Stephenson, D. R., Ayling, K., & Rutter, D. R. The role of visual communication in social exchange. *British Journal of Social and Clinical Psychology*, 1976, *15*, 113–120.

Watson, C. K., & Friend, R. Measurement of social evaluation anxiety. *Journal of Consulting and Clinical Psychology*, 1969, *33*, 448–457.

Yngve, V. H. On getting a word in edgewise. In *Papers from the sixth regional meeting of the Chicago Linguistics Society*. Chicago: Chicago Linguistics Society, 1970.

Zimmerman, D. H., & West, C. Sex roles, interruptions and silences in conversation. In B. Thorne & N. Henley (Eds.), *Language and sex: Difference and dominance*. Rowley, Mass.: Newbury House, 1975.

Received April 17, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 6, 879-896

Multidimensional Input in Equity Theory

Arthur J. Farkas and Norman H. Anderson
University of California, San Diego

Multiple dimensions of input, which are typical of equity judgments in practical situations, may be handled in two ways. The standard assumption of *input integration* implies that the input cues are summed for each individual to yield a single unitary value of input for each separate person; the ratio model of equity division then applies to these unitary input values. Input integration thus transforms the multidimensional stimulus field into a one-dimensional value of input that mediates the equity division. The alternative assumption of *equity integration* implies a reverse order of processing: First, the ratio model is applied to make an implicit equity division along each separate input dimension; the resulting equity ratios are then averaged to yield a final judgment. The difference in processing order implies different patterns of interaction that thus serve to test the two assumptions. The results of the present research agreed very well with the hypothesized rule of equity integration when the input dimensions were dissimilar. When the input dimensions were similar, the results agreed moderately well with the hypothesized rule of input integration, although there were small, worrisome discrepancies. Overall, the results provide some promising support for a cognitive algebra of equity.

It is widely assumed that equity should obtain when a person's *outcome* is proportional to *input*. But how can this simple rule apply when input is multidimensional? In many social situations a person's deservingness depends not only on concrete accomplishment but also on how hard the person tried. How can a one-dimensional outcome be proportional to a two-dimensional input? How can proportionality be maintained when the person tried hard but accomplished very little? Or accomplished a great deal with little effort?

This problem of multidimensional input is fundamental in equity theory. Adams (1965) has emphasized how numerous factors (seniority, age, educational and family background, for example) can all affect the perceived value of input. Similarly, a person might consider his or her intentions, efforts, or even need, as part of input, over and above actual performance. In social practice, therefore, input will generally be multidimensional. Any theory of equity must solve the problem of multidimensional input.

An obvious solution is to assume a one-dimensional mediator. All relevant stimulus information about the person would be integrated to obtain a single value of deservingness. The person's outcome would then be taken as proportional to this deservingness value. Under this hypothesis of *input integration*, the multidimensionality resides in the stimulus field but the effective mediator is itself one-dimensional.

This hypothesis of input integration has been taken more or less for granted. Adams (1965) and Walster, Berscheid, and Walster (1976) have assumed an input summation rule, while Anderson (1976a) and Leventhal (1976) have suggested an averaging rule.

This work was supported by National Science Foundation Grants GS-36918 and BNS 75-21235. Support was also received from grants from the National Institute of Mental Health to the Center for Human Information Processing, University of California, San Diego. Parts of Experiment 1 were presented at the Western Psychological Association meetings, San Francisco, April 1974.

Requests for reprints should be sent to Arthur J. Farkas, Code 307, Navy Personnel Research and Development Center, San Diego, California 92152, or to Norman H. Anderson, Psychology C-009, University of California, La Jolla, California 92093.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3706-0879$00.75

Only one experiment (Anderson, 1976a) has provided an experimental test, however, and these results raised doubt about the validity of any rule of input integration.

The assumption of input integration is stringent. It requires that all the diverse determinants of input be reduced to a common currency. That might be hard to do with qualitatively distinct dimensions, such as actual accomplishment and effort, for example, or productivity and seniority. To integrate such different dimensions into a unitary, one-dimensional mediator may be difficult.

There may be an easier way to handle multidimensional input. The first step would be an implicit equity division for each separate dimension of input. That involves comparing the several persons on only one dimension at a time. Further, these implicit divisions are themselves comparable across different dimensions, because they all represent relative deservingness values. Hence, they should be easily integrated to obtain an overall division.

This scheme may be called *equity integration* (Anderson, 1976a), because it is the implicit equity divisions that are integrated. Both steps in equity integration involve quantities that are naturally comparable. Thus, equity integration offers a psychologically attractive way to handle multidimensional input.

The present experiment was designed to test between the two hypotheses of input integration and equity integration. This turns out to be fairly straightforward by means of cognitive algebra. However, it requires an explicit statement of the two hypotheses. For simplicity, this statement will be restricted to the equity model employed in information integration theory (Anderson, 1976a; Anderson & Farkas, 1975; Farkas, 1977; Farkas & Anderson, Note 1; Anderson, Note 2). Other equity models are taken up in the discussion.

In this model, the state of equity for two persons, A and B, is defined by the equality of two proportionate ratios:

$$I_A/(I_A + I_B) = O_A/(O_A + O_B), \quad (1)$$

where I_A and I_B are the inputs for Persons A and B, and O_A and O_B are their corresponding outcomes.

With a fixed amount T to divide between A and B, Equation 1 becomes

$$O_A = [I_A/(I_A + I_B)]T. \quad (2)$$

Thus, A's fair share is just A's proportionate input. This model, however, assumes that I_A and I_B are one-dimensional. The theoretical problem concerns multidimensional input.

In the present experiments, information was given about two input dimensions: effort (E), or how hard the person tried, and performance (P), or how much the person actually accomplished. The rule of *input integration* could then be written as an adding-type model:

$$I_A = P_A + E_A, \quad I_B = P_B + E_B.$$

Weight parameters, w_P and w_E, are omitted for convenience, since they are effectively absorbed in the units of the scales of P and E. Substitution of these values into Equation 2 yields the model for input summation:

$$O_A = [(P_A + E_A)/(P_A + E_A + P_B + E_B)]T. \quad (3)$$

Under *equity integration*, separate equity calculations are made for performance and for effort, and these two equity ratios are then integrated. Under an averaging rule,

$$O_A = w_P[P_A/(P_A + P_B)]T + w_E[E_A/(E_A + E_B)]T. \quad (4)$$

Here, w_P and w_E are the relative weights associated with the two input dimensions, with $w_P + w_E = 1$.

It is simple to test between input summation and equity summation because they predict quite different data patterns. For example, Equation 4 implies that P_A and E_A combine additively because they are separated by a plus sign. The corresponding two-way factorial graph should thus exhibit a pattern of parallelism.

In contrast, Equation 3 implies that P_A and E_A combine nonadditively because they are separated by a division sign as well as by a plus sign. In this particular case, the corresponding two-way factorial graph should exhibit the pattern of a diverging fan of curves.

This two-way $P_A \times E_A$ graph thus provides

a simple sign to diagnose which, if either, of the two integration rules underlies the overt judgment. Other two-way interactions may be analyzed in the same way. This method of diagnosis requires some care, because the theoretical deviations from parallelism may be too small to detect under certain conditions. With due care, however, these two-way data patterns can be used to delineate the structure of the underlying integration processes.

Method

The subjects received information about the inputs of two persons, A and B, who had worked together on a common task. This input information specified effort (how hard A and B had tried) and performance (how much A and B had actually accomplished). The subject's job was to divide a fixed reward fairly between A and B on the basis of their inputs. Judgments were made on a 200-mm graphic scale labeled "Person A" and "Person B" at the ends.

Experiment 1

In Parts 1 and 2, the five levels of performance and effort were specified as follows: extremely above average, above average, average, below average, and extremely below average. The input information for each judgment was typed on an index card, and the deck of cards for each design was presented in a different shuffled order for each subject. Preliminary practice was given for each design. Subjects were 20 student volunteers from introductory psychology classes who received extra class credit for their services. They were run individually, served in all four designs, and made one judgment to every stimulus combination.

Part 1. Two 5×5 designs were used, the first of which specified the performances of A and B and the second of which specified the efforts of A and B. The first design was presented twice, with the first presentation serving as added practice, and the second design was presented once.

Part 2. One 5×5 design was used, in which both the performance of A and the effort of A were varied. Subjects were told that B had worked at the task but that they would not be informed of what B had done.

Part 3. One 3^4 design was used. Both performance and effort were specified for A and B. Except for a difference in the levels (namely, above average, average, and below average), the procedure was the same in this part as in Parts 1 and 2.

Experiment 2

This experiment had two main purposes. The first was to replicate the main design of Experiment 1 (Part 3) with more extreme levels of the stimuli. The second was to attempt to induce a strategy of input summation by making the input information homogeneous. By presenting two pieces of performance information about each person, it was thought that subjects would integrate this homogeneous information into a preliminary judgment of deservingness for each person and then make an equity division based on these two deservingness values.

Each factor had three levels in all designs, namely, very much above average, average, and very much below average. Subjects were 32 new subjects from the same pool as in Experiment 1. General procedure was similar to that of Experiment 1.

Part 1. Two 3^4 designs were used. In one, performance and effort were specified for A and for B, as in Part 3 of Experiment 1, except for the change in the phrases used to define the levels. In the other, two performances from separate and independent days were specified for A and for B. For each design, four examples were used as part of the instructions. These were followed by four end anchors and then 18 representative trials from the regular design. The order of presentation of these two designs was balanced across subjects.

Part 2. Two 3^2 designs were used. In one, performance was specified for A and for B. In the other, performance was specified on 2 separate and independent days for A, and B's performance over the 2 days combined was specified as average.

Results

The main goal in the data analysis was to test the two alternative rules, input integration and equity integration. These two rules predict quite different data patterns. These data patterns, especially the two-way factorial plots, may be used to diagnose which, if either, model underlies the observed judgments. Accordingly, the main exposition will be based on the visual appearance of the factorial graphs from the four-factor designs. Technical aspects of the model analyses are considered in more detail in a separate section.

Effort and Performance: Experiment 1

Theoretical interest centers on the two-way factorial graphs in the top layer of Figure 1. There are six graphs in this top layer, one for each two-way interaction in the main design. The two on the left have the shape of a slanted barrel; the remaining four exhibit parallelism. These six patterns of data must be related to the theoretical predictions.

Equity integration. The hypothesis of

equity integration provides a complete account of these data. All four panels on the right are predicted to be parallel by the reasoning given in connection with Equation 4 in the introduction. To illustrate, consider the third graph in the row. This shows that the fair share for A is an inverse function of B's performance (listed on the horizontal axis) and a direct function of A's effort (listed as curve parameter). From Equation 4, B's performance (P_B) and A's effort (E_A) combine additively, since they are separated by a plus sign. Accordingly, the theoretically expected shape is one of parallelism, and that is verified in the graph. Exactly the same reasoning applies to the three graphs farther to the right because each represents effort for one factor, and performance for the other.

The slanted barrel shapes of the two graphs at the left also agree with equity integration. To illustrate, consider the leftmost graph. It shows that the fair share for A is an inverse function of B's performance (listed on the horizontal axis) and a direct function of A's performance (listed as curve parameter). The barrel shape reflects the basic equity ratio, $P_A/(P_A + P_B)$, in Equation 4 (e.g., Anderson & Farkas, 1975; see also below). In the same way, the barrel shape of the second left graph reflects the basic equity ratio, $E_A/(E_A + E_B)$.

All six two-factor graphs thus agree with predictions from the hypothesis of equity integration. This hypothesis also predicts that the higher-order interactions are zero, a prediction that was largely supported, as noted below.

Input integration. The hypothesis of input integration disagrees with four of the two-way graphs in the top layer of Figure 1. It predicts that the two leftmost graphs should be very nearly parallel, whereas they actually show a barrel shape. It also predicts that the two rightmost graphs should form converging or diverging fans, whereas they are nearly parallel. Only the near parallelism of the two middle graphs agrees with prediction. These data, therefore, are sharply inconsistent with the hypothesis of input integration.

It needs mention that Equation 3 for input integration in principle predicts nonparallelism for all six graphs in Figure 1. In certain cases, however, predicted deviations from parallelism may be negligible, as for the four leftmost graphs of Figure 1. And although not of present concern, parallelism can also obtain when only two levels of each variable are employed (Anderson & Butzin, 1978). Un-

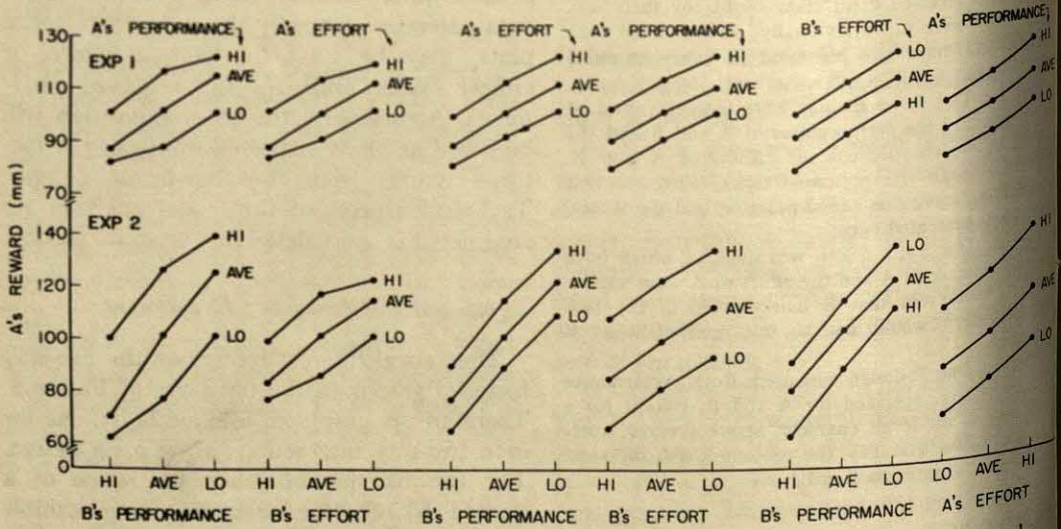


Figure 1. Mean reward allotted to Person A as a function of several stimulus variables. (Top and bottom layers show graphs of the six two-way interactions from corresponding designs of Experiments 1 and 2, respectively. HI = high; AVE = average; LO = low.)

fortunately, the algebraic structure of Equation 3 is not very informative about the sizes of the interactions. Accordingly, it appears necessary to use numerical analysis, as illustrated below in the section on model analysis.

Goodness of fit. Formal tests of the model may be obtained directly from the analysis of variance. Because of its theoretical importance, the complete analysis is presented in Table 1. Of primary interest are the F ratios for the two-way interactions: These F ratios test for deviations from parallelism for the two-factor graphs in the top layer of Figure 1. The Fs for the two leftmost graphs are 32.68 and 18.07, $p < 10^{-9}$. The sizes of these F ratios reflect the high reliability of the data, which allows ready detection of what are rather small deviations from parallelism. The other four graphs, in left-to-right order, correspond to nonsignificant Fs of 1.47, .67, .19, and 2.24, respectively. These tests thus confirm the above graphical analysis and support the hypothesis of equity integration.

Equity integration also implies that the higher-order interactions should be nonsignificant. This implication follows from Equation 4, because any set of three factors contains two that are separated only by a plus sign. In fact, none of the three-factor interactions in Table 1 is significant. The four-factor F of 1.83 is technically significant, but it is small and did not seem meaningful. Overall, therefore, the data provide strong support for equity integration and clear evidence against input integration.

Effort and Performance: Experiment 2

Further support for equity integration is given by Experiment 2, which included a nearly exact replication of the main design of Experiment 1 considered just above. The two-way graphs are plotted in the bottom layer of Figure 1. Each of these graphs shows essentially the same pattern as the corresponding graph from Experiment 1 in the top

Table 1
Summary Analyses of Variance for Four-Way Designs

Source	Performance × Effort: Experiment 1: F	Performance × Effort: Experiment 2: F	Source	Performance × Performance: Experiment 2: F
A: Performance-A	149.58	298.85	A: Performance-A1	669.78
B: Performance-B	173.13	314.79	B: Performance-B1	509.04
C: Effort-A	111.57	71.33	C: Performance-A2	583.17
D: Effort-B	99.52	72.54	D: Performance-B2	535.16
M error	472.02	1667.70	*M* error	517.40
AB	32.68	40.93	AB	7.98
AC	2.24	1.07	AC	6.61
AD	.67	.24	AD	6.13
BC	1.47	.50	BC	12.04
BD	.19	3.84	BD	17.96
CD	18.07	23.12	CD	11.70
M error	73.44	180.05	*M* error	104.55
ABC	1.55	6.09	ABC	10.47
ABD	.37	5.10	ABD	6.46
ACD	1.55	2.22	ACD	11.69
BCD	.67	3.07	BCD	6.94
M error	72.39	167.03	*M* error	98.20
ABCD	1.83	5.91	ABCD	18.75
Error	65.91	153.09	Error	111.60

Note. Two-, three-, and four-way interactions have 4, 8, and 16 *df*, respectively. Each interaction has its own error term, with 19 and 30 times as many *df* in Experiments 1 and 2, respectively. Critical values of F at the .05 level are 2.51, 2.00, and 1.66, on 4/76, 8/152, and 16/304 *df*, respectively, and 2.45, 1.96, and 1.65, on 4/120, 8/240, and 16/480 *df*, respectively.

layer. The two leftmost graphs show the predicted barrel shape, and the four rightmost graphs show parallelism. Just as in Experiment 1, therefore, these data provide clear evidence against the hypothesis of input integration and strong support for the hypothesis of equity integration.

There is one difference between the two experiments in Figure 1. The graphs show greater vertical spread in Experiment 2 than in Experiment 1. Also, the performance–effort asymmetry is considerably larger, as can be seen by comparing the two leftmost graphs in each layer. These effects arise from the changes in the definition of the stimulus levels and do not require any qualification of the theoretical interpretation.

However, one complication arose in the statistical analysis for Experiment 2. Although the two-way interactions were much the same as in Experiment 1, Table 1 shows that the three-factor and four-factor interactions are all substantial, contrary to the equity integration model. At first it was thought that these ineractions represented a mild degree of input integration. However, more detailed analysis showed that the interactions predicted by input integration were much too small.

Similar patterns of interaction were also obtained in collateral work on children's judgments of equity (Anderson & Butzin, 1978). These data suggested the presence of a configural element within the basic strategy of equity integration. When A and B had equal performance (or effort), subjects appeared to place greater weight on effort (performance). Thus, a dimension of difference appeared to influence the judgment more than a dimension of no difference. When this configural assumption was incorporated into the model, it provided a remarkably good account of the data, as is shown in the more detailed model analysis below.

Performance and Performance: Experiment 2

The rule of input integration was expected to apply when the input consisted of two pieces of performance information about each person. Since these two pieces of input information are similar, it should be easy to integrate them to obtain a single value of deservingness for each person. These deservingness values would then constitute the effective input values for the equity division between the two persons.

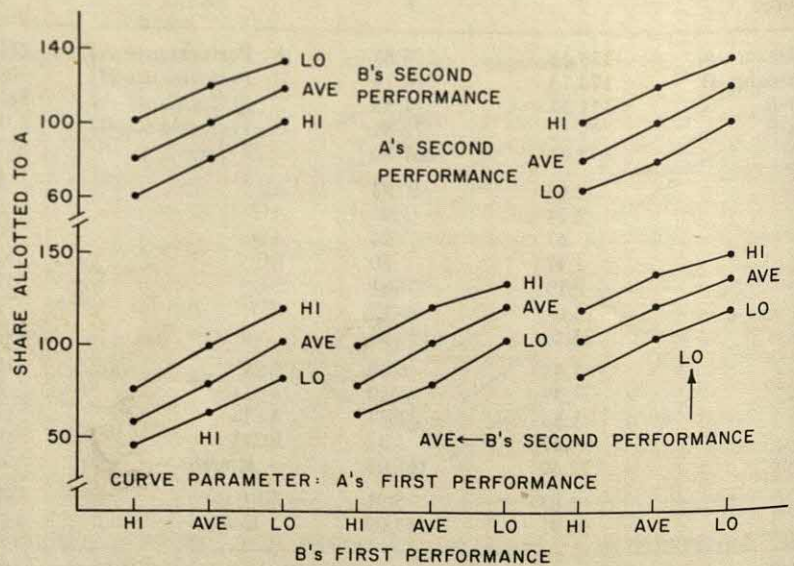


Figure 2. Mean reward allotted to Person A as a function of several stimulus variables, Experiment 2, Part 1. (Upper graphs show two representative two-way interactions; lower graphs show representative three-way interaction. HI = high; AVE = average; LO = low.)

The model of input integration provided an approximate account of the data, but there were small, worrisome discrepancies. Figure 2 gives an overview of the data. The graph at the upper right is representative of the four two-way interactions that involved one piece of information about each person. Theoretically, these graphs should be almost parallel. However, all four showed a slight barreling, and these deviations from parallelism were significant, as shown in Table 1.

A more serious difficulty for input integration is the pattern in the two-way graph in the upper left of Figure 2. This graph shows a mild convergence to the right, whereas the model predicts a moderate divergence. The arithmetic rationale for the divergence follows directly from the fact that both terms in this interaction, being performance levels of B, appear in the denominator of the model. The larger the value of one term, therefore, the smaller the predicted effects of the other. This discrepancy is reliable, since it also appeared for A. The given arithmetic rationale suggests that this discrepancy cannot be rectified by replacing the sum $P_A + E_A$ in Equation 3 by a more general input integration function, $f(P_A,E_A)$. These data therefore present a serious problem for the hypothesis of input integration.

A representative three-way interaction is shown in the lower layer of Figure 2. The left and right panels show mild divergence and convergence, respectively, whereas the center graph exhibits the barrel shape. The three panels are not geometrically congruent, and that incongruence produced a significant three-way interaction. The other three-way interactions had a similar pattern, a consequence of the symmetry in the design. Such a shift in pattern across the three panels, from divergence to barrel shape to convergence, is predicted by input summation, and it was initially thought that these three-way interactions provided good support for the model. However, more detailed analyses showed that the predicted interactions had substantially smaller magnitude than the observed interactions. Accordingly, these interaction patterns also present a difficulty for the hypothesis of input integration.

It deserves mention that the interaction patterns in Figure 2 provide a strong test of the input integration model of Equation 3, because that model has only two free parameters in the present experimental design. All four performance factors in the design are equivalent by symmetry, and the value of the middle level may arbitrarily be set equal to unity. Thus, the only two free parameters are the values of the high and low levels of performance. Since these values are essentially determined by the main effects, the interaction patterns provide clear tests of the model. It is possible, of course, that some configural assumption similar to that considered above would resolve the discrepancies. However, several attempts in this direction met with little success. Even though the discrepancies are small, therefore, they raise a serious question about the validity of the hypothesis of input integration.

A natural question is whether the rule of equity integration, which did well in the other main design, might also apply to the data in Figure 2. That this is not the case may be seen by comparing the two-way graphs between Figures 1 and 2. In Figure 2, all four graphs that included one piece of information about each person had the same slight barrel shape shown in the upper right. In Figure 1, however, the four graphs that included one piece of information about each person are the four leftmost graphs in either layer. These do not all have the same shape. Further, none of them is overly similar to the upper-right graph in Figure 2. From this comparison alone, it is clear that the rule of equity integration does not account for these data.

It is worth noting that a simple linear, adding–subtracting model would describe the data rather well, for the deviations from parallelism in Figure 2 are visibly small. For practical applications, at least, an adding-type model could provide a good approximation. Of course, that does not mean that the underlying integration process is basically additive (Anderson & Shanteau, 1977), although the possibility that the subjects tended to simplify the task in this manner deserves consideration (Shanteau & Anderson, 1972).

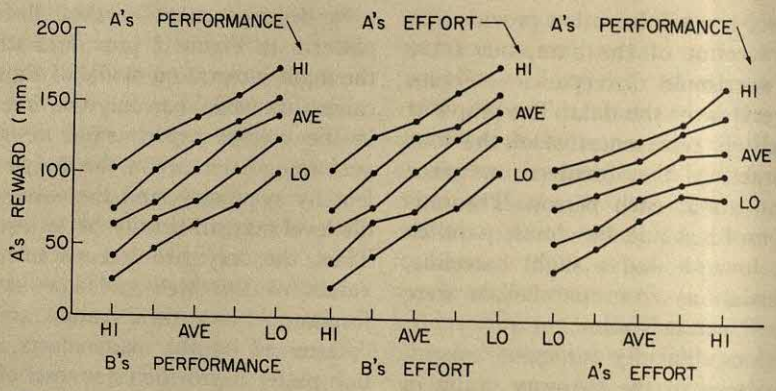


Figure 3. Mean reward allotted to Person A as a function of several stimulus variables. (Each panel represents one of the two-way designs of Experiment 1. HI = high; AVE = average; LO = low.)

Other Tests of the Equity Ratio Model

Both experiments contained designs in which the equity judgment was based on two pieces of information. These data do not bear on the main question of input integration, but they are relevant to the basic ratio model of equity theory given by Equations 1 and 2.

The results from the three designs of Experiment 1 are given in Figure 3. These graphs show little or no sign of the slanted barrel shape that is predicted by the basic ratio model. Indeed, each panel appears to be roughly parallel, as though an adding-type integration obtained. The deviations from parallelism are statistically significant in each case, $F(16, 304) = 2.72$, 4.04, and 3.99 in the three respective graphs. However, no pattern is apparent in these deviations, and they certainly do not follow the theoretical barrel shape.

A different picture was found in Experiment 2, as can be seen in Figure 4. Here both graphs show a clear barrel shape, in agreement with the equity model. The deviations from parallelism are highly significant, $F(4, 124) = 17.20$ and 24.86 in the two respective graphs. The basic model of Equation 2 gave a good fit to these data, as is shown below.

The five graphs of Figures 3 and 4 give a mixed picture on the equity ratio model. The data of Experiment 2 support the ratio model,

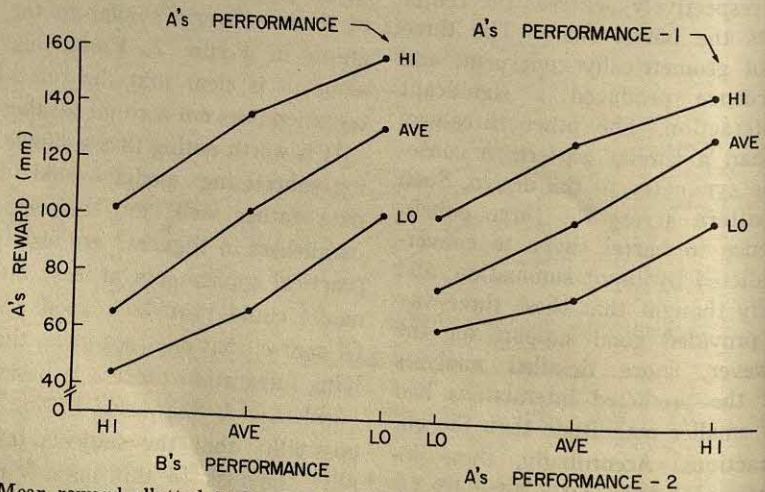


Figure 4. Mean reward allotted to Person A. (Each panel represents one of the two-way designs of Experiment 2. HI = high; AVE = average; LO = low.)

whereas those of Experiment 1 appear to support an adding-type formulation. Of course, the ratio model was not expected to hold for the case of heterogeneous information, namely, the Performance-A × Effort-A design in the rightmost panel of Figure 3. Similar results had been obtained by Anderson (1976a, Figure 3), and those results had led to the hypothesis of equity integration that was studied in the present experiments. However, the other two graphs of Figure 3 represent homogeneous stimulus information, and the apparent failure of the ratio model in these two cases is more disturbing.

One possible source of this difference in results lies in the different ordering of the two-way and the four-way designs in the two experiments. The two-way designs came first in Experiment 1, last in Experiment 2. It is possible, therefore, that the extended practice at integrating four pieces of information produced a tendency to use a ratio model that

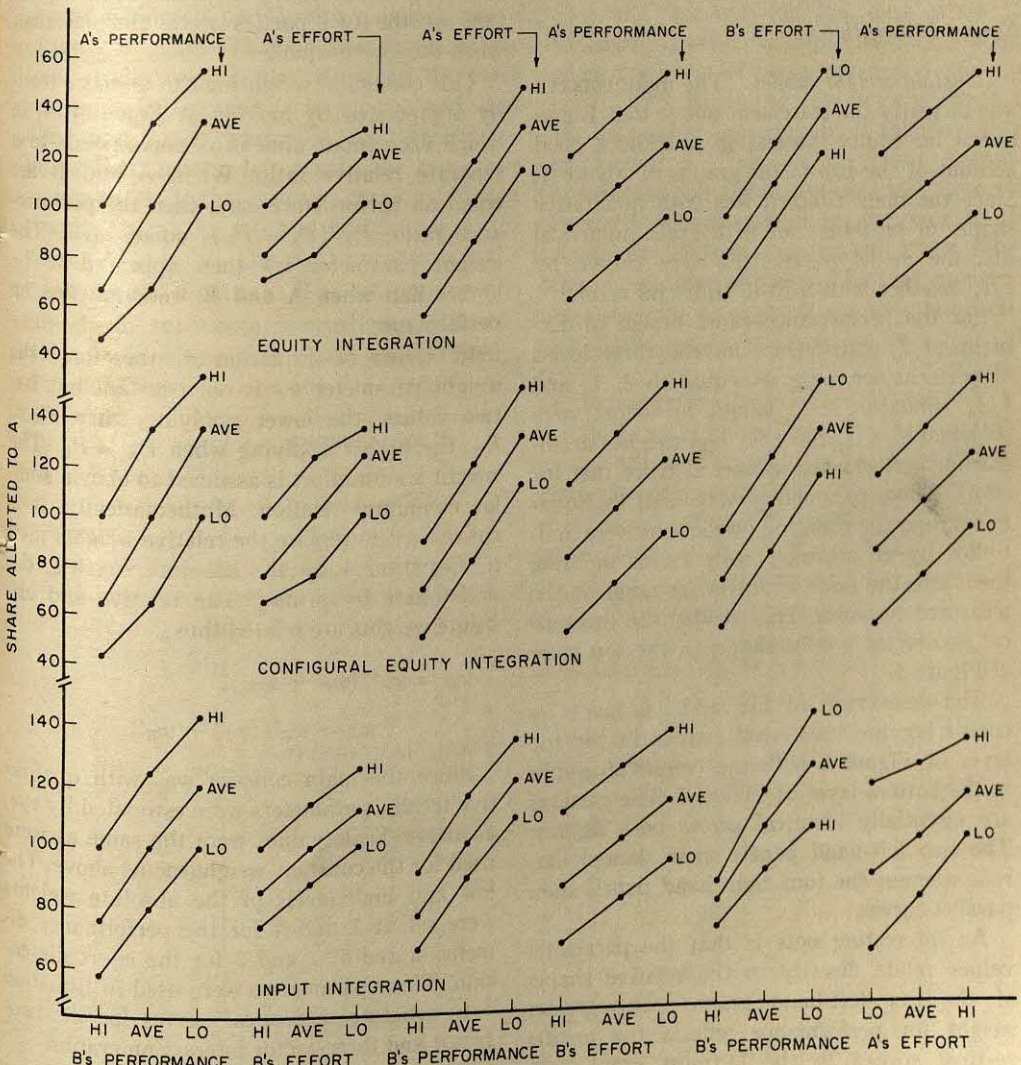


Figure 5. Theoretical predictions of two-way factorial graphs for the performance–effort design, Experiment 2. Top layer: Predictions from constant-weight equity integration. Middle layer: Predictions from configural-weight equity integration. Bottom layer: Predictions from input integration. (HI = high; AVE = average; LO = low.)

carried over to the later two-way designs in Experiment 2. When these two-way designs came first, as in Experiment 1, a simpler adding-type strategy may have prevailed.

Model Analyses

This section presents the more detailed analyses for the models that have been discussed above. For simplicity, only the data of Experiment 2 will be explicitly considered.

Equity Integration

Constant-weight model. The main concern was to verify the statement above that Equation 4 for equity integration provides a good account of the two-factor graphs of Figure 1. Since the main concern was with qualitative shapes or patterns, not with exact numerical fits, the model parameters were chosen by eye, together with a little trial and error.

For the performance–effort design of Experiment 2, scale values for the three levels of performance were set equal to 3, 1, and 1/3; similarly, scale values for effort were set equal to 4, 1, and 1/4. The weight parameter for performance was set at twice that for effort. These parameters were used in Equation 4, and the resulting predictions were multiplied by an arbitrary scale factor to bring them into the same approximate range as the measured response. This yielded the theoretical two-factor graphs shown in the top layer of Figure 5.

The assessment of the model is made by comparing the theoretical graphs in the top layer of Figure 5 with the empirical graphs in the bottom layer of Figure 1. The patterns are essentially identical across both figures: The two left-hand panels show slanted barrels, whereas the four right-hand panels show parallel curves.

An interesting note is that the parameter values relate directly to the relative shapes of the two left-hand graphs. The greater weight for performance reflects the greater vertical spread in the leftmost graph; the greater range of scale values for effort reflects the greater proportionate barreling of the next-to-left graph. Extreme levels of effort thus appear to have greater valence but less influence than extreme levels of performance.

Configural-weight model. Although Equation 4 for equity integration provided a good account of the two-factor graphs, it failed to account for the three-factor interactions that appeared in Experiment 2. Inspection of the data suggested that these three-factor interactions resulted from configural weighting. Each input dimension seemed to have greater influence on the equity division when the inputs of the two parties were unequal than when they were equal.

This configural weighting can be more readily appreciated by looking at Equation 4, in which each input dimension corresponds to a separate relative ratio. When A and B are equal on performance, say, then the performance ratio, $P_A/(P_A + P_B)$, equals 1/2. The weight parameter w_P then appeared to be lower than when A and B were unequal on performance.

In terms of Equation 4, therefore, the weight parameter w_P is not constant but has two values, the lower applying when $P_A = P_B$, the higher applying when $P_A \neq P_B$. The weight w_E for effort is assumed to obey a similar configural relation. Mathematically, it is convenient to replace the relative weights used in Equation 4 by the absolute weights, denoted here by primes. The relative and absolute weights are related thus:

$$w_P = w_P'/(w_P' + w_E'),$$

$$w_E = w_E'/(w_P' + w_E').$$

Since the main concern was with qualitative trends, parameters were estimated by eye, as above. Scale values were the same as were used for the constant-weight model above. The low and high levels of the absolute weights were set at 2 and 4 for the performance dimension and at 1 and 2 for the effort dimension. These parameters were used in Equation 4 to generate theoretical curves for the two-factor and three-factor interaction graphs.

The predicted two-way graphs are shown in the middle layer of Figure 5. Each two-way graph has the same shape as the corresponding graph from the constant-weight model

just above. Thus the configural-weight and constant-weight models yield virtually identical two-factor patterns. This pattern similarity comes about because the performance and effort dimensions are represented by separate ratio terms in Equation 4.

The main purpose of the configural model was to account for the three-factor interactions. This it does in remarkable detail. One of the empirical interactions is shown as the three 3×3 graphs in the upper layer of Figure 6. All three graphs show the same basic shape: a slanted barrel with a middle curve of greater slope. That is the shape of the basic equity ratio model. However, the three graphs differ somewhat in shape, and that difference is the source of the three-factor interaction. In the left graph, the top curve is bowed, but the bottom curve is almost straight; also, the top and bottom curves diverge to the right. The right graph shows the complementary pattern: The bottom curve is bowed, the top curve is nearly straight, and the two curves converge to the right. Finally, the center graph shows the pure form of the basic equity ratio model.

The theoretical predictions from the configural-weight model are shown in the lower layer of Figure 6. These have virtually the identical pattern as the data. The other three-factor interactions all had similar shape and were accounted for in the same manner. Thus, the configural-weight model for equity integration provides a detailed account of the main patterns of the data.

It is worth adding that the configural model does not account for the data when the weighting is reversed to use the larger weight when the levels of performance and effort are equal. This weighting reverses the shape of the leftmost predicted graph in the lower layer of Figure 6. The bottom curve becomes bowed, the top curve straight, and these two curves converge instead of diverging to the right; the rightmost predicted graph shows the complementary changes. The configural weighting is thus quite sensitive to specific pattern features in the data.

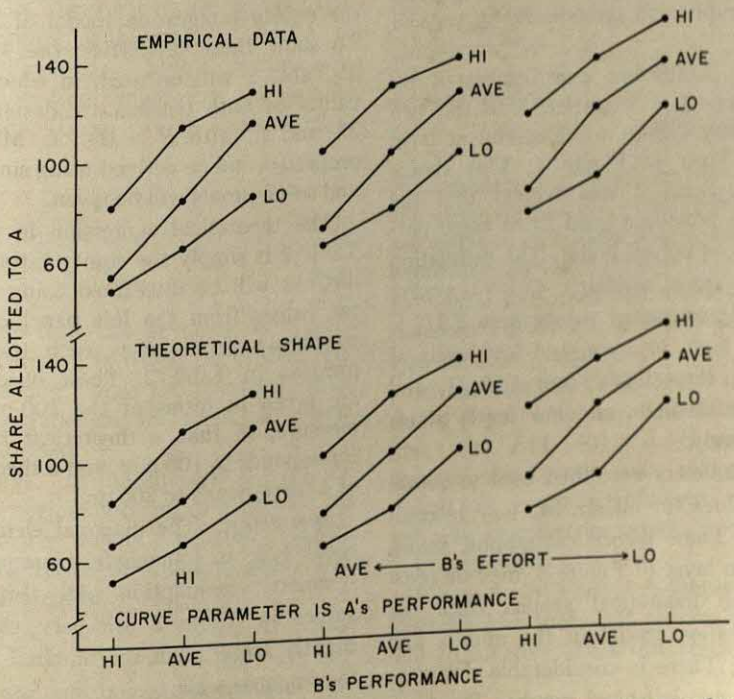


Figure 6. Comparison of observed three-way interaction with predictions from configural-weight model of equity integration. (HI = high; AVE = average; LO = low.)

Input Integration

One remaining question in the model analysis is whether the hypothesis of input integration can account for the data from the main designs that included information about both performance and effort. The model for equity integration gave a good account of these data, but perhaps the model for input integration could do the same.

To obtain a definite answer to this question, an exact fit of the model was made in the following way. Scale values for the medium levels of performance and effort were both set equal to unity; this is no restriction, since it merely sets the scale units. The weights are not identifiable parameters in this model, since they are absorbed into the scale units and so need not be estimated. Further, the scale values for Persons A and B are the same by virtue of the symmetry in the design. Thus, there are only four free parameters, the high and low levels of performance and of effort. The values of these four parameters are determined essentially by the sizes of the main effects. Accordingly, the shapes of the two-factor graphs will provide strong tests of the model.

The four parameters were estimated by fitting the model of Equation 3 to the two center two-way graphs for Experiment 2 in the bottom layer of Figure 1. This choice was made because it was evident that the model would provide a good fit to these particular data, as indeed it did. The estimation was done using Chandler's (1969) STEPIT program. The estimated values were 2.32, 1, and .33 for high, medium, and low levels of performance, respectively, and 1.69, 1, and .56 for high, medium, and low levels of effort, respectively.

These parameters were then used to obtain predicted values for all six two-way interaction graphs. These theoretical graphs, shown in the bottom layer of Figure 5, may be compared to the theoretical graphs from the equity integration model in the middle and upper layers. There is considerable disagreement for the two left-hand graphs: These exhibit the slanted barrel shape in the upper layers, but are virtually parallel in the bottom layer. There is also considerable disagreement for the two right-hand graphs: These are virtually parallel in the upper layers, but are markedly nonparallel in the bottom layer. Since the theoretical graphs for equity integration fit the data very well, it follows that the model for input integration fits very ill.

Configural-weight versions of the model for input integration were also considered, but these did not seem to give much improvement. It would seem that the algebraic structure of the model for input integration precludes a fit to these patterns of response.

Some Properties of the Equity Ratio Model

The relative ratio model of Equation 2 has some interesting mathematical properties when the two persons are treated symmetrically. These properties provide a basis for exact tests of fit and also for estimating the effective stimulus values of the inputs. These properties hold both for Equation 2 and for the two corresponding two-factor graphs of the equity integration model of Equation 4. To show these properties, the 3×3 design of Table 2 will be used, in which the input values of both persons are designated as H, M, and L, with $H > M > L$. Most of these properties can be derived using simple algebra, and so no proofs will be given.

The theoretical expression in each cell of Table 2 is simply the equity ratio. These predictions will be illustrated using the numerical values from the left panel of Figure 4, which are listed below each theoretical expression in Table 2. These observed values are given in terms of the 200-mm response measure, so that a theoretical ratio of 1/2 corresponds to 100 mm, and a theoretical ratio of 1 corresponds to 200 mm.

Symmetry. The diagonal elements should equal 1/2, an immediate consequence of the symmetry assumption. All three observed values in Table 2 are very close to 102, slightly larger than the nominal midpoint of the response scale.

Complementarity. Entries in cells that are symmetrically located about the main diag-

Table 2
Theoretical Analysis of Ratio Model, Symmetrical Case

Row level	Column level: H	M	L
H	(a) $H/(H+H)$ 101.9	(b) $H/(H+M)$ 137.2	(c) $H/(H+L)$ 157.8
M	(d) $M/(M+H)$ 65.3	(e) $M/(M+M)$ 102.2	(f) $M/(M+L)$ 133.0
L	(g) $L/(L+H)$ 43.7	(h) $L/(L+M)$ 67.8	(i) $L/(L+L)$ 102.1

Note. H, M, and L denote the three values of the input cues, assumed equal for the two parties. In each cell of the design, the ratio of values of input cues represents theoretical prediction; numerical value represents data from Experiment 2, Part 2. Letters a–i are used for identification in text.

onal should add to unity. In Table 2, therefore,

$$b + d = c + g = f + h = 1. \quad (5)$$

The observed sums are 202.5, 201.5, and 200.8, respectively, slightly greater than the theoretical value of 200. Complementarity is also an immediate consequence of the symmetry assumption. These data, like those just noted, support the further use of the symmetry assumption for this task.

Parallelism subtests. Table 2 contains three 2×2 subtables that should obey the parallelism prediction. These can be designated as follows:

$$\begin{aligned} a - d &= b - e, \\ e - h &= f - i, \\ a - g &= c - i. \end{aligned} \quad (6)$$

Each of these tests can be made as a 2×2 analysis of variance in which the critical term is the interaction. Since the observed deviations from parallelism are small, no formal test was performed.

Barrel shape. In the interaction tests, discussed in a previous section, considerable emphasis was placed on the prediction of a barrel shape. In terms of Table 2, the barrel shape is equivalent to

$$b - h > a - g = c - i. \quad (7)$$

The effect was substantial, since the difference on the left is 69 mm, and the two differences on the right are 58 and 56 mm, respectively. The mean net difference is thus 12 mm, considerably larger than the difference of 3.5 mm that would be required to reach significance.

An asymmetry. There is also an interesting asymmetry that is implied by the ratio model, though its size and sign depend on the exact values of the parameters. For the 3×3 design, the middle curve will generally have greater slope than the top and bottom curves. In terms of Table 2, this asymmetry is measured by

$$(c - f) - (d - g) = (HL - M^2)/(H + M)(M + L). \quad (8)$$

Accordingly, the expression on the left has the same sign as $HL - M^2$. Since M can be set equal to 1 without loss of generality, the expression on the right will be positive when $H > 1/L$. This condition was satisfied by the parameter estimates below, though the effect was not overly large.

In the data of Table 2, $c - f = 24.8$ mm, and $d - g = 21.6$ mm. The difference of 3.2 mm, with $F(1, 30) = 3.47$, was somewhat short of the 3.5-mm value that would be required to reach significance. Considerably stronger asymmetries appear in many of the graphs in Figures 1 and 2.

Parameter-free test. A simple test of fit is available that does not require estimation of any parameters. This test rests on the relation,

$$c/(1 - c) = b/(1 - b) \times f/(1 - f). \quad (9)$$

To perform this test, the terms on the left and right were calculated for each subject, and a t test was applied to the 32 difference scores. The mean difference of .2 did not approach significance, which supports the model. It should be noted, incidentally, that using the means from Table 2 in the expression above would introduce a small bias, since, in general, the ratio of the means is not the mean ratio.

Direct estimation of parameters. Three parameters appear in Table 2, but one of them can be set to any convenient value. This reflects the fact that multiplying all three parameters by a constant does not change the values of the ratios in the model. Accordingly, M was set equal to 1.

An estimate of H is then obtainable by equating the theoretical and observed values in Cell b of Table 2. This yields $H = 2.17$. Similarly, equating the theoretical and observed values in Cell h yields an estimate of $L = .52$. These estimates are not the best, because they do not use all of the data. Also they suffer a slight bias relative to the mean of the corresponding individual estimates. However, they differ little from the improved estimates discussed next, and they would be adequate for many purposes. For example, the mean magnitude of the discrepancies between the observed means of Table 2 and the values predicted from the estimates given above was only 1.4 mm.

Improved parameter estimates. The parameter estimates will be more stable when they are based on all of the data. Such estimates were obtained by fitting the ratio model separately to each subject's data, using a least squares criterion. Owing to the nonlinear nature of the model, an iterative procedure is required, and accordingly, Chandler's (1969) STEPIT program was used. The mean values of the estimates of H and L were 2.15 and .53, respectively—quite close to those obtained by the less efficient direct method given above. The mean magnitude of the nine deviations between the mean observed and the mean predicted value was only 1.2 mm.

Goodness of fit. Getting an exact statistical test of the ratio model is complicated by the nonlinearity, which prevents an exact tally of the degrees of freedom used in estimating the parameters. Fortunately, this complication can be circumvented using a new test procedure introduced by Leon and Anderson (1974). In brief, predicted values were obtained for each subject based on the parameter estimates for that subject. A 3×3 matrix of deviation scores was then obtained for each subject, and these data were tested in an ordinary repeated measurements design. The model implies that rows, columns, and interaction should all be nonsignificant in this analysis, and that prediction was supported by the results: The corresponding F ratios were $F(2, 60) = .15$, $F(2, 60) = .57$, and $F(4, 120) = 2.07$. On the basis of this exact test of goodness of fit, therefore, the ratio model does rather well. This one set of data is only illustrative, of course, but the same methods can be used more generally. With the iterative procedure, parameter estimation and the goodness-of-fit test do not require the assumption of symmetry.

Discussion

Theoretical Implications

Input integration versus equity integration. Multidimensional input is characteristic of everyday equity judgments. Rarely are such judgments made on the basis of a single piece of information. Typically, they entail the use of different kinds or types of information. That creates an obvious difficulty, noted in the introduction, in dividing a one-dimensional outcome in proportion to a multidimensional input.

This problem has been largely avoided in previous discussions of equity by the attractive assumption of input integration. The multiple aspects and dimensions of the stimulus field are assumed to be integrated to form one overall "deservingness" value. The complex stimulus field is thus reduced to a one-dimensional, effective input that mediates the division. Unfortunately, this assumption of input integration ran into difficulty in its first experimental test (Anderson, 1976a, Part 3, p. 295). The present results confirm the

difficulty and indicate that the hypothesis of input integration has limited applicability.

The alternative hypothesis of equity integration has the advantage that it does not require integration of qualitatively heterogeneous dimensions of input. Instead, each input dimension acts independently to determine its own implicit equity division. These implicit divisions, which are themselves qualitatively similar, are then integrated to determine the overt equity division.

This assumption of equity integration did remarkably well for heterogeneous input dimensions (performance and effort) in the present experiments. The one complication, from the higher-order interactions in Experiment 2, seemed to be easily handled in terms of configural weighting. The generality of the present results certainly needs to be assessed, using other scenarios, more than two dimensions of input, and more than two parties. As they stand, however, these data provide strong support both for the hypothesis of equity integration and for the basic relative ratio model.

Unfortunately, equal success cannot be claimed for the case of homogeneous input dimensions (actual performance on two occasions). The hypothesis of input integration had been expected to hold in this case, and the model predictions were verified to a good numerical approximation. Nevertheless, the pattern of discrepancies suggested that the underlying cognitive algebra was different than assumed in the model. Indeed, a linear model based on simple adding and subtracting of the four input dimensions did about as well.

This difficulty with the model for input integration is perplexing. Although the problem of integrating input for a single person has not received much explicit study, there seems little doubt that subjects can do this (see Anderson, 1978a, pp. 40ff). Also, the basic relative ratio model for division between two persons did well for the case of heterogeneous input dimensions, as already noted. Simple concatenation of these two operations would yield the input integration model. In further work on the problem of input integration, therefore, it may be useful to require initial judgments of deservingness about each separate party. That requirement would presumably force input integration as a preliminary to equity division and would help localize the discrepancies from the model.

Molar unitization. When input integration holds, then the effective input can be treated as a molar unit (Anderson, in press-a, in press-b; Anderson & Graesser, 1976) without regard for its molecular composition. No matter how intricate the stimulus field, no matter how complicated the process of input integration, the net result is a single value. This value can be quantified with functional measurement and used for description and for theoretical tests without knowledge of historical antecedents. When input integration holds, therefore, it can markedly simplify the theoretical analysis.

When equity integration holds, however, then input cannot be treated as a molar unit. Each dimension of input must be considered separately. That may be feasible when the input dimensions are under experimental control. However, it could be a serious problem in naturalistic situations in which the input dimensions may not all be known. Despite the present success of the rule of equity integration, the failure of molar unitization is disheartening because it points to unexpected complexities in equity theory.

One such complexity concerns the boundaries of equity theory. Some writers (e.g., Deutsch, 1975; Lerner, 1975) have assumed that parity (all persons treated alike) or need, for example, were distinct concepts that lay outside the boundaries of equity theory. Other writers (e.g., Walster et al., 1976) have claimed equity theory as a general theory, presumably encompassing such concepts as need, effort, and so forth. However, neither group of writers has discussed how this issue may be tested.

Equity theory does provide a simple conceptual test: If need affects input, then need lies within the domain of equity theory. This conceptual test would become operational if input integration held. Then the pattern of data could reveal whether need affected input in the same way as work contribution or acted as a separate force (Anderson, Note 2). But if equity integration holds, then need acts separately, regardless of whether it is a proper

determinant of input. Under equity integration, therefore, it may not be easy to progress beyond phenomenological speculation about the boundaries of equity theory.

Despite these disadvantages, the rule of equity integration does have one notable advantage in relation to the problem of multidimensional outcome (Farkas, 1977). In general, outcome is not one-dimensional. Besides money, it may include praise, status symbols, working conditions, and so forth. A simple rule of input integration would require that all outcomes be divided in the same proportion. However, a rule of equity integration makes direct allowance for weights that represent input–outcome relevance. Thus, more money might be given the person with the greater work contribution, more praise to the person who tried harder, and more status symbols to the person who needed them more. From this standpoint, therefore, the rule of equity integration may be conducive to a more just society.

Overview of Equity Models

The discussion above has been restricted to the equity model from information integration theory, as given by Equations 1 and 2 in the introduction. This section gives an overview of other models that have been proposed in equity theory, followed by a brief summary of negative evidence on the basic ratio assumption employed in many of these models. For simplicity, it will be assumed that inputs are one-dimensional and positive (see also Anderson, Note 2, p. 162).

Three ratio models. The relative ratio model from information integration theory (Equations 1–2) is closely related to ratio models proposed by Aristotle (quoted in Walster & Walster, 1975) and by Adams (1965). Aristotle's definition of equity is given by the equation:

$$O_A/O_B = I_A/I_B. \quad (10)$$

Adams defines equity by the equation:

$$O_A/I_A = O_B/I_B. \quad (11)$$

These models imply different psychological comparison processes. In Aristotle's model, the two ratios involve comparison of like quantities. In Adams' model, the two ratios involve comparison of unlike quantities. Algebraically, however, they are equivalent to each other and to the relative ratio model of integration theory. Equity judgments do not distinguish among these three models.

However, it is possible to distinguish among the three models by obtaining judgments of inequity (Anderson & Farkas, 1975). In these tests, both the relative ratio model and Aristotle's model did quite well. Adams's model, however, did quite poorly, an outcome that has been corroborated in an extensive study of unfairness judgments by Farkas (1977).

Walster, Berscheid, and Walster. Walster et al. (1976; see also Walster, 1975) defined equity with the following equation:

$$(O_A - I_A)/|I_A|^{k_A} = (O_B - I_B)/|I_B|^{k_B}, \quad (12)$$

where the ks are $+1$ or -1 according as $(O - I)$ and I have the same or opposite signs. This model was introduced in an attempt to avoid apparent paradoxes that arise with negative input for the three ratio models considered in the preceding subsection. However, simple thought experiments show that the Walster model also leads to paradoxes, even indeed with positive inputs and outcomes (Anderson, Note 2, p. 161; Note 3, p. 16).

To illustrate, let $I_A = 10$ and $I_B = 5$, in 2:1 ratio. With 18 units to be divided, an equitable distribution is clearly $O_A = 12$ and $O_B = 6$. The three ratio models and the Walster model all agree on this.

Now suppose that there are only half as many outcome units to be divided. It seems obvious that the outcome for each person should also be reduced in half. Indeed, the first three models imply $O_A = 6$ and $O_B = 3$. But the Walster model requires $O_A = 8$ and $O_B = 1$. It seems hardly just that B should lose 4 out of 5 input units, whereas A loses only 2 out of 10.

Finally, suppose that there are only 3 units to be divided. Then the Walster model not only denies that $O_A = 2$ and $O_B = 1$ is equitable, but claims that equity requires $O_A = 6$ and $O_B = -3$. Because B's input was less than A's, B is required to make an additional payment to A. Thus, B's net loss comes to

8 units, whereas A's net loss is only 4 units. This thought experiment shows that the Walster model cannot be valid.

Harris. A somewhat different approach to equity models has been taken by Harris (1976), who considers a number of formulas in terms of rational criteria that they might be expected to satisfy (e.g., outcome should be an increasing function of input). Only Harris's preferred "linear formula" (slightly amended by Harris, Note 4) is considered here. For a two-party division, it defines equity by the equations,

$$O_A = aI_A + r$$
$$O_B = aI_B + r, \qquad (13)$$

where the parameters $a > 0$ and r are the same for both parties, but may depend on I_A and I_B.

This formula is unsatisfactory because it can always give a perfect fit to two-party divisions. The formula is not truly linear because a need not and generally could not be constant across situations. In general, Harris allows a to be some arbitrary, unknown function of the inputs and outcomes of all the parties to a given division. Equations 13 are thus two equations in two unknowns and so can always be solved exactly for a two-party division. In the present two-factor designs of Figures 3 and 4, for example, Harris's formula would provide a perfect fit to every data point. These perfect fits do not support the formula, of course, but instead show that it is not testable with two-party divisions. Perhaps some theoretical foundation for the a parameters could be found that would allow a test similar to that allowed by the models discussed above. As long as the formula is not testable with two-party divisions, however, it does not seem to be very satisfactory.

Subjective values. As has been emphasized by Adams (1965) and by Walster et al. (1976), equity theory in psychology deals with personal values and feelings. In general, therefore, tests of equity models must be capable of measuring psychological values of input and outcome on true linear or ratio scales. Failure to recognize the need for subjective measures can undercut the theoretical reasoning.

This difficulty can be seen in the literature on children's judgments of equity. Nearly all this work has relied on physical counts as measures of input and outcome. Input might be measured by number of problems solved, for example, and outcome by number of candies received. Various theoretical interpretations rest on the implicit assumption that subjective values are proportional to these physical values. This assumption is arbitrary and uncertain, however, and so some of this work does not have any clear meaning (Anderson & Butzin, 1978).

It should be emphasized that many studies of equity do not require linear or ratio scales. A monotone, rank-order response scale can be entirely adequate for certain purposes (see, for example, Anderson, 1976b, p. 689), and it may not then be worthwhile to seek for linear scales. For other purposes, however, linear or ratio scales are vital. One reason for studying algebraic models of equity is that they can provide a base and frame for such measurement scales.

Evidence on the relative ratio model. Although the relative ratio model has done remarkably well in some situations, it has done more or less poorly in others. Existing data point to three areas of poor performance that deserve mention. The most severe discrepancies were obtained in the sole test in which total outcome was not a fixed sum. In this case, the model has a ratio form that implies a linear fan pattern of response. The obtained response pattern showed marked discrepancies from this theoretical linear fan (Anderson, 1976a, Figure 1 and p. 295; 1978b). A second kind of discrepancy appeared in Experiment 2 above for the task in which two pieces of similar performance information were given about each person. Although the rule of input integration was perhaps not too far off, the meaning and seriousness of the discrepancies remain unknown.

The third discrepancy is that rather similar tasks sometimes yield the theoretical pattern of a slanted barrel, but sometimes a pattern of approximate parallelism. Illustrations appear in Figures 3 and 4 above, and in Figures 2 and 3 of Anderson (1976a). Failure to get the theoretical barrel shape does not seem to

be due to lack of power. Although the obtained barrel patterns show only modest deviation from parallelism, they have typically been highly significant statistically. It is possible that the approximate parallelism results from a tendency to simplify the task by using adding or subtracting rules in place of the seemingly more difficult ratio rule. This is only speculation, however, for there is almost no evidence on the causes of this third kind of discrepancy.

This brief summary indicates that the cognitive algebra of equity (and inequity) is in an uncertain state. Still, the relative ratio model has had some substantial successes, and there is hope that the discrepancies will find a simple theoretical interpretation.

Reference Notes

1. Farkas, A. J., & Anderson, N. H. *Input summation and equity summation in multi-cue equity judgments* (Tech. Rep. CHIP 47). La Jolla, Calif.: University of California, San Diego, Center for Human Information Processing, December 1974.
2. Anderson, N. H. *Social perception and cognition* (Tech. Rep. CHIP 62). La Jolla, Calif.: University of California, San Diego, Center for Human Information Processing, July 1976.
3. Anderson, N. H. *Equity judgments as information integration* (Tech. Rep. CHIP 47). La Jolla, Calif.: University of California, San Diego, Center for Human Information Processing, December 1974.
4. Harris, R. J. Personal communication, September 1978.

References

Adams, J. S. Inequity in social exchange. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2). New York: Academic Press, 1965.

Anderson, N. H. Equity judgments as information integration. *Journal of Personality and Social Psychology*, 1976, *33*, 291–299. (a)

Anderson, N. H. How functional measurement can yield validated interval scales of mental quantities. *Journal of Applied Psychology*, 1976, *61*, 677–692. (b)

Anderson, N. H. Cognitive algebra: Integration theory applied to social attribution. In L. Berkowitz (Ed.), *Cognitive theories in social psychology*. New York: Academic Press, 1978. (a)

Anderson, N. H. Progress in cognitive algebra. In L. Berkowitz (Ed.), *Cognitive theories in social psychology*. New York: Academic Press, 1978. (b)

Anderson, N. H. *Information integration theory: A case history in experimental science* (Vol. 1). New York: Academic Press, in press. (a)

Anderson, N. H. Integration theory applied to cognitive responses and attitudes. In R. Petty, T. Ostrom, & T. Brock (Eds.), *Cognitive responses in persuasion*. Hillsdale, N.J.: Erlbaum, in press. (b)

Anderson, N. H., & Butzin, C. A. Integration theory applied to children's judgments of equity. *Developmental Psychology*, 1978, *14*, 593–606.

Anderson, N. H., & Farkas, A. J. Integration theory applied to models of inequity. *Personality and Social Psychology Bulletin*, 1975, *1*, 588–591.

Anderson, N. H., & Graesser, C. C. An information integration analysis of attitude change in group discussion. *Journal of Personality and Social Psychology*, 1976, *34*, 210–222.

Anderson, N. H., & Shanteau, J. Weak inference with linear models. *Psychological Bulletin*, 1977, *84*, 1155–1170.

Chandler, J. P. Subroutine STEPIT—Finds local minima of a smooth function of several parameters. *Behavioral Science*, 1969, *14*, 81–82.

Deutsch, M. Equity, equality, and need: What determines which value will be used as the basis for distributive justice? *Journal of Social Issues*, 1975, *31*, 137–149.

Farkas, A. J. *A cognitive algebra for bystander judgments of interpersonal unfairness*. Unpublished doctoral dissertation, University of California, San Diego, 1977.

Harris, R. J. Handling negative inputs: On the plausible equity formulae. *Journal of Experimental Social Psychology*, 1976, *12*, 194–209.

Leon, M., & Anderson, N. H. A ratio rule from integration theory applied to inference judgments. *Journal of Experimental Psychology*, 1974, *102*, 27–36.

Lerner, M. J. The justice motive in social behavior: An introduction. *Journal of Social Issues*, 1975, *31*, 1–19.

Leventhal, G. S. Fairness in social relationships. In J. W. Thibaut, J. T. Spence, & R. C. Carson (Eds.), *Contemporary topics in social psychology*. Morristown, N.J.: General Learning Press, 1976.

Shanteau, J. C., & Anderson, N. H. Integration theory applied to judgments of the value of information. *Journal of Experimental Psychology*, 1972, *92*, 266–275.

Walster, E., Berscheid, E., & Walster, G. W. New directions in equity research. In L. Berkowitz & E. Walster (Eds.), *Advances in experimental social psychology* (Vol. 9). New York: Academic Press, 1976.

Walster, E., & Walster, G. W. Equity and social justice. *Journal of Social Issues*, 1975, *31*, 21–43.

Walster, G. W. The Walster et al. (1973) equity formula: A correction. *Representative Research in Social Psychology*, 1975, *6*, 65–67.

Received April 17, 1978

Journal of Personality and Social Psychology
1979, Vol. 37, No. 6, 897-912

Race, Sex, and the Expression of Self-Fulfilling Prophecies in a Laboratory Teaching Situation

Marylee C. Taylor
Harvard University

The aims of the study were (a) to identify patterns of teacher behavior that vary with expectations for pupil ability and thus may mediate teacher expectancy effects and (b) to investigate ways in which pupil race and sex may interact with perceived pupil ability in affecting teacher behavior. For the experiment, 100 white female undergraduates enrolled in teacher training programs presented lessons to "phantom" students, who were alleged to be observing and responding from behind one-way glass. Race, sex, and ability manipulations were embedded in descriptions of these phantom pupils. Numerous verbal and nonverbal aspects of teacher behavior were recorded. Subjects with low-ability "pupils" taught less material, allowed less response opportunity, gave lengthier praise after successful performance, and showed less vocal nervousness. Particularly interesting Race × Ability interactions were found, showing that teachers direct most positive affect to high-ability black and low-ability white pupils.

In 1968 Robert Rosenthal and Lenore Jacobson published the report of a study demonstrating that experimentally manipulated teacher expectancies for pupil progress can affect pupils' intellectual growth. Subsequent research has provided steadily accumulating evidence of self-fulfilling prophecy effects in classroom education (Crano & Mellon, 1978; Rosenthal, 1973, 1976). One implication is that the much discussed relation between pupil academic achievement and such background characteristics as race and socioeconomic status may be partially attributable to teacher expectations for pupil performance.

Although little question remains about the existence of expectancy effects, there are gaps in our understanding of how these effects operate. The present study addresses two kinds of questions about the operation of teacher expectancy effects: First, what particular patterns of teacher behavior vary with expectations for pupil ability and thus may mediate the relation between teacher expectations and pupil performance? And secondly, how do teacher expectations about pupil ability interact with other pupil characteristics—particularly pupil race, and also pupil sex—in affecting teacher behavior?

Ability-Expectancy Effects

Systematic investigation of the mediation of teacher expectancy effects has been undertaken by a number of researchers, and an integration of these research findings has been presented by Rosenthal (1973). In his review, Rosenthal offers a four-factor "theory" of teacher expectancy mediation, naming four major categories of mediators, each suggested by at least five empirical studies to be related to pupil-ability expectations held by teachers. Rosenthal's *climate* factor represents general, noncontingent teacher warmth, communicated through such behaviors as smiling, head nod-

This article is based on a doctoral dissertation completed at Harvard University. The research was funded, in part, by the National Science Foundation.

The author would like to thank Robert Rosenthal for his support of this project and for his valuable assistance. The author is also grateful to Gloria I. Joseph, Judith A. Hall, and Michael P. Johnson for their generous contributions at various phases of this project.

Requests for reprints should be sent to Marylee C. Taylor, who is now at the Department of Sociology, The Pennsylvania State University, University Park, Pennsylvania 16802.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3706-0897$00.75

ding, leaning forward, and maintaining eye contact. Greater teacher warmth is generally directed to pupils for whom positive expectations are held. The second factor is *input*; teachers try to teach more material and more difficult material to their high-ability pupils. Represented in Rosenthal's *output* factor is the evidence that teachers encourage greater responsiveness from high-ability pupils. This may take such forms as offering more frequent response opportunities, asking harder questions, or being more patient in waiting for pupil responses.

Research evidence pertaining to Rosenthal's fourth factor, *feedback*, is more complex. In general, it seems to be the case that feedback is given more clearly and more consistently to high-ability than to low-ability pupils. There are qualifications, however: When criticism after unsuccessful performance is the feedback at issue, low-ability pupils have sometimes been the disfavored group; and particularly high levels of teacher praise have occasionally been found to be directed toward low-achieving and disliked pupils (Brophy & Good, 1974). Of the four categories suggested by Rosenthal, feedback seems to be the one most in need of conceptual and empirical clarification.

A goal of the present study was to examine effects of experimentally induced teacher expectations on behaviors representing all four of the mediation types Rosenthal suggests.

Pupil-Race Effects

Previous research provides substantial evidence for effects of pupil race as well as ability expectations on teacher orientation and behavior, with black children the recipients of the less desirable evaluations and treatment at the hands of (typically white) teachers. (See, for example, Antonoplos, 1972; Datta, Schaefer, & Davis, 1968; Leacock, 1969; Yee, 1968). Yet these studies focused on classroom teachers and their actual black and white pupils, so that pupil race was confounded with other pupil characteristics that could plausibly affect teacher orientation. Studies that have failed to find effects of pupil race typically employed a "phantom other" design, in which subjects were asked to rate written descriptions of imaginary black and white students (see Dietz & Parkey, 1969; Long & Henderson, 1974; Mazer, 1971; Miller, 1973), and there is reason for concern about generalizability. In a series of studies that attempted to address both internal and external validity concerns, safeguards against confounding with pupil race and realism were introduced into the design (Coates, 1972; Crowl & MacGinitie, 1974; Harvey & Slater, 1975; Jensen & Rosenfeld, 1974); in these, pupil-race effects were found, but the dependent variables were teacher ratings rather than behavioral measures. Thus, there do remain questions about the effect of pupil race per se on teacher behavior.

The Repression-of-Affect Hypothesis

An intriguing hypothesis specifying a pattern of incongruous race effects comes not from educational research but from a study of interracial interaction. Weitz (1972) administered a racial-attitude questionnaire to her subjects, and later obtained a number of measures of their orientation to a fictitious, soon-to-appear co-worker described to them as being either black or white. The most interesting aspect of the Weitz findings was negative correlations among some of her measures of white-to-black attitudes and behaviors. Weitz concluded that whites with negative racial feelings may, under pressures of a liberal environment, attempt to portray a pro-black posture but leak their underlying negative feelings in behaviors not amenable to conscious control. The result would be the emission of double messages, suggested (Bateson, Jackson, Haley, & Weakland, 1956) to be harder for children to handle than consistent negative messages. The repression-of-affect proposition is certainly important enough in its implications to warrant further testing.

The Interaction of Pupil Race and Ability

Central to the present project is the question of how pupil race may combine with ability expectations to influence teacher be-

havior. Most discussions of pupil-race and ability expectations assume a simple additive model: Whiteness and a high-ability label serve as two strikes in favor of a child, blackness and a low-ability reputation, two strikes against. A disturbing alternate possibility is, however, suggested by the findings of Rubovits and Maehr (1973). When these researchers assigned white university students to teach black and white children who had been randomly assigned labels of "gifted" or "nongifted," there was an interaction of race and ability in their effect on teacher behavior: The gifted label seemed to be an advantage for the white children but a disadvantage for the black children.

The Rubovits and Maehr evidence does not, however, stand without contradiction. Other studies have yielded results suggesting Race × Ability interaction in the other direction, with low-ability black pupils being the recipients of particularly negative teacher orientation (see Datta et al., 1968; Henderson & Long, 1973). In the case of these latter studies, however, the dependent measure has been teacher ratings rather than behaviors. Thus there remains the need to examine Race × Ability effects on teacher behavior, using a design that eliminates the possible confounding of pupil race with other factors. Such an examination was a major goal of the present study.

Pupil Sex and Its Interaction With Race and Ability

As a supplementary benefit, the study offered the opportunity to examine teacher behavior for evidence of pupil-sex effects. From their review of relevant research, Brophy and Good (1974) concluded that there is evidence of teachers' preference for female pupils, as well as evidence that teachers underestimate both the potential and the achievements of male pupils. However, they failed to find convincing evidence of favoritism toward girls in teacher behavior and concluded that previously noted pupil-sex effects on teacher behavior were accounted for by sex differences in the behavior of the children. The present study allowed reconsideration of the no-difference conclusion drawn from previous research on sex effects. Perhaps more interesting, it allowed examination of evidence regarding possible interactive effects on teacher behavior of pupil sex and ability, and of pupil sex and race, topics about which little is, as yet, understood.

Overview of the Study

The present study was designed to address simultaneously a number of the hypotheses and questions suggested by the literature reviewed here. White female undergraduates enrolled in teacher training programs were asked to teach two lessons to a phantom student, who was alleged to be observing and responding to them from behind one-way glass. On a light panel in the experimental room, the subjects received standard responses to their questions from their "6-year-old pupil," described as being either black or white, male or female, and of high or low ability. Numerous kinesic and paralinguistic measures of teacher behavior during the lesson presentation were obtained, along with measures of various aspects of teaching style as well as teacher ratings of the pupils.

The variety of behavioral measures in the present study, including representatives of each of Rosenthal's (1973) four factors, allowed for examination of effects of pupil characteristics on the patterning of those teacher behaviors suggested to be most important in the transmission of self-fulfilling prophecy effects in one-to-one teaching situations. This design allowed testing for interactive effects of ability expectations and pupil race on teacher behavior, without the possibility, present in the Rubovits and Maehr (1973) study, of confounding effects from actual differences in pupil behavior. Built into the present study was a direct search for evidence of Weitz's (1972) repression of affect in subjects teaching black "pupils": At the midpoint of the experimental session there was a manipulation of demand characteristics; half the subjects received a message designed to encourage attempts at problack behavior and thus to maximize the probability of leakage and the contradictory pattern Weitz discusses. Finally,

auxiliary questions, such as those regarding main and interactive effects of pupil sex on teacher behavior, were examined.

Method

Subjects

For the experiment, 105 white female[1] undergraduates at the University of Massachusetts, Amherst, served as subjects. All were either majoring in education or working for teaching credentials. All but 2 were upper-class students, all but 12 had practice teaching experience or the equivalent, and the majority were enrolled in elementary education programs. A few subjects volunteered in response to posted announcements of the study, but the vast majority were identified from university records and recruited by telephone. Recruits were told only that they would receive $2 compensation for 1 hour of participation in an experiment that would concern teaching.

Setting, Equipment, and Materials

The experimental room, in which the subjects performed their teaching task, was simply furnished with a long table, a rolling chair, a floor microphone to be used during lesson presentations, and an armchair and coffee table for the subjects' use during other phases of the experimental routine. Furnishings were positioned so that during their lesson presentations the subjects could sit at an angle from the unobstructed section of a one-way glass window into the observation room, facing the window but unable to see their own reflection in it. Beneath the observation room window stood a light panel on which the "pupil" responses were to be transmitted to the subjects. The panel contained 10 lights, labeled to represent 10 possible responses to the prescribed questions accompanying the lessons.

Two lessons were constructed for the experiment. Both concerned mathematics, and the two paralleled each other as nearly as possible. Each lesson contained four sections, the internal structure of the sections being identical. A given section contained the following: a statement of the concepts and definitions to be read to the "pupil" by the subject; a set of examples, half of which were illustrated; question-answering instructions to be read to the "pupil"; and a set of illustrated questions.

Standard response schedules were followed by an experimental assistant who controlled the light panel. For each lesson question, the schedule specified not only the response(s) to be given but also the latency of the response(s). (In some cases incorrect answers were given initially, and correct responses given only after repeated questioning; for such questions, the sequence of responses to be given by the experimental assistant was specified by the schedule.)

The lessons were constructed in consultation with a specialist in elementary mathematics education. Response schedules were constructed after pretesting the lessons on children the age of the alleged pupils of the experiment. Thus, the appearance of the lessons as legitimate educational vehicles and the plausibility of the allegation that an actual child was responding to the questions were maximized.

Experimental Manipulations and Design

The three major independent variables were manipulated through a Pupil Information Sheet, which contained pupil race, sex, and ability designations, embedded in other information. All pupil information (including age—6 years) was identical except for the pupil's name (John or Betty Powell), sex, ethnic origin (Afro-American or English American), and ability ("at the high end of the class range" and "in one of the highest classroom groups" or "at the low end of the class range" and "in one of the lowest classroom groups"). The awareness manipulation, designed to intensify demand characteristics and possible repression of affect, was included as part of a written reminder for the subjects to try to behave as if they were in a natural teaching situation. For those in the aware condition, this reminder was preceded by the statement: "A major purpose of this study is to compare the ways teachers behave toward black and white students."

A completely crossed repeated-measures design was used. Subjects in the 16 (Ability × Race × Sex × Awareness) conditions were measured twice on each of the dependent variables, once during each lesson presentation. In order to minimize the possibility that any effects of time or order would be confounded with those of the independent variables, assignment to experimental condition was performed within each block of 16 subjects scheduled for consecutive testing. Because some supplementary subjects were tested, and their data retained in the analysis, cell *n*s were slightly uneven, ranging from five to eight.

Procedure

As each subject was greeted by the (female) experimenter, she was presented with written materials containing instructions and an explanation that the research purpose was to examine certain aspects of teaching behavior in a situation where feedback from pupil to teacher was limited. The subject was then taken to the experimental room and left for 5 min to read the Pupil Information Sheet, which

---

[1] With four independent measures of central interest here, it was deemed infeasible to vary experimenter sex and race as well. White females were the group targeted for study because of their historical and continuing prevalence in elementary teaching. There is certainly the need for analogous investigation of black and male populations, an inviting task for the future.

contained the pupil race, sex, and ability manipulations; to review the materials pertaining to the first of the two lessons she would present (the order in which the two lessons were presented was simply alternated from one subject to the next); and, finally, to rate estimated lesson difficulty for the pupil. At the conclusion of this interval, the subject was signaled to move the rolling chair and floor microphone to a position of her choice alongside the large table, and to begin presenting the lesson to the pupil who was allegedly listening and watching through the observation room window. Lesson presentation involved proceeding through the sections of the lesson, repeating a sequence that consisted of exposition, optional presentation of illustrated or nonillustrated examples, and questioning pertinent to the material of that section. At the end of 10 min, the subject was signaled to conclude the lesson presentation, and was given 4 min in which to rate the learning and ability of the pupil.

At this point the subject read the reminder in which the awareness manipulation was included. Then the second lesson was presented, procedures for this parallel presentation being identical to those for the first.

Before being excused, the subjects completed a written manipulation check and probe for hypotheses about the nature of the experiment, and were paid the $2 compensation. Later, after all the subjects had been tested, a thorough debriefing letter was mailed to all the participants, and an announcement was made of a session at which the subjects could discuss the study with the experimenter.

Two assistants were present in the observation room during all testing sessions, one controlling recording equipment and the light panel that allegedly conveyed pupil responses to questions, and the other rating kinesic and proxemic behaviors of the subjects. When possible, the experimenter made independent ratings so that interjudge reliabilities could be computed.

As a safeguard against experimenter expectancy effects, all those at the research site were completely blind to subject condition, except in those cases where the subject used the name of her "pupil," thus revealing the sex of the "pupil" to the assistant who was performing auditory monitoring of the teaching sessions. This assistant was positioned so as to be shielded from visual cues, which might have biased responses on the light-panel controls. Analogously, the assistant performing proxemic and kinesic ratings was shielded from vocal cues, which might have contaminated the ratings.

Aside from the teacher ratings and the kinesic measures recorded by the rater positioned in the observation room, dependent measures were taken from the audio recordings of the teaching sessions. Most involved simple coding performed directly from the audiotapes, the exception being the voice ratings, which required a more elaborate procedure.

For purposes of obtaining voice ratings, four 3-sec clips of each subject's lesson presentation were excerpted, one from the beginning and one from the middle of each lesson presentation. These taped excerpts were content-filtered, a process in which certain frequencies are deleted from the recording, rendering the verbal content of the speech incomprehensible but preserving much of the tonal quality and rhythm of the speech. (For a more detailed description of this technique, see Rogers, Scherer, & Rosenthal, 1971). For raters, 16 female Harvard summer school students were recruited. The raters, half black, and half white, ranged in age from 18 to 21 years; eight had had some teaching experience. Each rater was presented with a stimulus tape containing half of the 3-sec excerpts, one excerpt representing each subject during each lesson presentation. During each of six playings of the tape, these judges rated the content-filtered excerpts on a dimension specified by the researcher.

Dependent Measures [2]

The climate factor. Rosenthal's (1973) climate factor was represented by four behavioral measures:

1. *Smiling* (number of smiles during three 1-min periods of a lesson presentation);
2. *Head nodding* (number of up-and-down head nods during three 1-min periods of a lesson presentation);
3. *Forward leaning* (seconds of leaning toward the observation room window during three 1-min periods of a lesson presentation);
4. *Vocal warmth* (mean of 16 rater judgments of content-filtered excerpts on a 9-point cold–warm scale, where warm is high).

The input factor. The input factor named by Rosenthal (1973) was represented by the following measure:

5. *Teaching pace* (number of nonoptional lesson segments presented during the 10-min session, coded so that for each of the four lesson sections, completion of all nonoptional segments would yield a score of 75).

---

[2] The dependent measures listed here are those to be discussed in the pages to follow. Their selection from a larger set was based on considerations of interpretability, theoretical import, nonredundancy, and empirical interest. Data collected but not discussed here include the following: a proxemic measure and kinesic measures of backward leaning and shifting; various indicators of the way in which teaching time was apportioned; measures of several kinds of assistance given during question-and-answer sessions; voice ratings on dimensions defined by the poles businesslike–personal, condescending–admiring, discouraging–encouraging, and not competent–competent; subjects' ratings of their pupils; and scores on a test of Afro-American history and culture administered to the subjects. Information about effects on these measures and relationships among them can be obtained from the author.

The output factor. Several assessments of the response opportunities offered the pupil were made, including:

6. *Impatience* (average number of times, per question asked, that the subject interrupted the "pupil" response, i.e., interrupted the latency period between question and response called for by the response schedule);

7. *Persistence after "pupil" errors* (percentage of correct answers obtained; since the response schedule dictated that correct answers could be elicited by repeating questioning, this is a measure of persistence in repeating incorrectly answered questions).

The feedback factor. The complexity of previous research findings concerning teacher feedback suggested that multiple measures should be used here. Five were employed: [3]

8. *Praise feedback incidence* (percentage of correct responses followed by explicit praise, e.g., "very good"; statements merely confirming the correctness of a response were not included);

9. *Praise feedback length* (average seconds spent praising correct responses per occasion on which such feedback was given);

10. *Incorrectness feedback length* (average seconds spent responding to an incorrect answer per occasion on which such feedback was given);

11. *Positive feedback omission* (percentage of correct responses followed by either silence or an ambiguous response);

12. *Negative feedback omission* (percentage of incorrect responses followed by either silence or an ambiguous response).

Vocal anxiety. In addition to the behavioral representatives of the four types of mediators suggested by Rosenthal (1973), included in the study were several indicators of another construct, vocal anxiety. Measures included the following:

13. *Speech errors* (total number of times a subject changed the course of a sentence in midstream, corrected herself, made obvious grammatical errors, mispronounced common words, misread sections of the lesson, or "got her tongue twisted");

14. *Pausing* (number of seconds of pausing during a lesson presentation, with 5 sec of silence used as the minimum criterion for a pause);

15. *Vocal nervousness* (mean of 16 rater judgments of content-filtered excerpts on a 9-point nervous-calm scale, where calm is high).

Inadvertent helpfulness. The final behavioral measure to be discussed here had not been planned but was suggested by the contents of the audiotapes:

16. *Helpful slips* (total number of times a subject made an apparent slip of the tongue and gave away the answer while trying to ask a question).[4]

Preliminary Analyses

Manipulation check. On the manipulation check instrument, administered at the conclusion of the study, 86 of the 105 subjects reported the appropriate ability level of their "pupil," all but two of the remaining responses being noncommittal. When asked about ethnic origin of their "pupils," 96 responded appropriately, 8 failed to answer, and 1 subject in the Afro-American condition reported "Anglo-Saxon." This last subject, who apparently misread the manipulation, was the only case excluded from the data analysis on the basis of the manipulation check; four others were excluded because of some deviation from standard experimental procedure. Thus, primary data analyses were carried out on data from 100 subjects.

The subjects' ratings of their "pupils" were used to assess the potency of the ability manipulation. The subjects had estimated "lesson difficulty for the pupil" and "pupil intellectual potential," in each case yielding a score from 0 to 130. Those in the high-ability condition estimated that the lessons would be easier for their pupils ($M = 71.2$) than did those in the low-ability condition ($M = 53.7$), $F(1, 77) = 17.19$, $p < .001$, $MS_e = 807.4$, $d = .945$. And subjects in the high-ability condition also gave higher ratings of their pupils' intellectual potential ($M = 89.1$) than did those in the low-ability condition ($M = 77.1$), $F(1, 82) = 10.29$, $p = .002$, $MS_e = 671.9$, $d = .708$.

Subjects' hypotheses regarding study. When asked for their hypotheses regarding the purposes of the study, nearly one half (44) of the subjects named a theme related in some way to the teacher expectancy idea (e.g., suggesting that the study's purpose was to assess the impact of pupil background on teaching style). Most of the remaining group repeated the cover story told them by the experimenter, or named an unrelated hypothesis. (As will be reported subsequently, where effects of ability on teaching behavior were found, these were confirmed in analyses that excluded the 44 subjects holding accurate suspicions about the research purpose.) Only eight subjects revealed skepticism about the existence

---

[3] Subject behavior after correct "pupil" responses was coded into three all-inclusive categories—praise feedback, correctness feedback (without praise), and feedback omitted. Thus the incidence of praise and feedback omission measures reported here represent all nonredundant information about type of feedback behavior following correct "pupil" responses. Subject behavior after incorrect "pupil" responses was, with a few exceptions (e.g., cases in which subjects enthusiastically confirmed the correctness of incorrect answers), coded into two categories—incorrectness feedback given or omitted. Thus, virtually all nonredundant information about type of feedback behavior following incorrect responses is represented in the feedback omission variable.

[4] Accidentally giving away question responses was not the only type of inadvertent behavior evidenced by the subjects. Two subjects in the black-pupil condition talked to their pupils about the "watermelon" as they looked at a question card containing both a large picture of an apple and a boldface caption referring to the "apple."

Table 1
Summary of Analysis of Variance Effects

Effect	Measure category	Measure	p
Ability (A)	Input	Teaching pace	.022
	Output	Impatience	.028
	Feedback	Praise feedback length	.009
	Vocal anxiety	Speech errors	.033
		Vocal nervousness	.075
Race (B)	Feedback	Positive feedback omission	.056
		Incorrectness feedback length	.054
	Freudian helpfulness	Helpful slips	.017
Sex (C)	Climate	Head nodding	.026
B × C	Feedback	Positive feedback omission	.049
	Freudian helpfulness	Helpful slips	.009
B × Awareness × Time	Output	Impatience	.065
	Feedback	Positive feedback omission	.019
B × A	Climate	Vocal warmth	.023
		Forward leaning	.031
	Freudian helpfulness	Helpful slips	.026
C × A	Feedback	Praise feedback incidence	.009
	Freudian helpfulness	Helpful slips	.026

Note. Relevant Fs and effect size estimates are presented in the text. Subgroup means for main effects are presented in the text, those for interaction effects in Tables 2–5.

of an actual child in the observation room, and many of the subjects responded in a way that confirmed their belief in the existence of an actual pupil, for example, by showing concern for the reactions of a child to the experimental procedure or by suggesting that pupil learning was the dependent variable of interest.

Reliabilities and intercorrelations. Reliabilities of the kinesic ratings were estimated using the independent ratings of a second rater for a subsample of the subjects. For smiling, $r(11) = .577$ and $r(29) = .830$ for the first and second lesson presentations, respectively; for head nodding, $r(12) = .703$ and $r(29) = .850$ for the first and second lesson presentations; and for leaning forward, $r(13) = .863$ and $r(20) = .715$ for the first and second lessons.

Since neither rater race nor timing of excerpt interacted appreciably with experimental condition in affecting voice ratings, analyses were performed on the means of the 16 ratings available for each subject from each lesson presentation. Reliabilities were computed from mean squares by a procedure that yields estimates equivalent to those produced by applying the Spearman-Brown prophecy formula to the mean interrater correlation. (See Guilford, 1954, p. 395.) For the cold–warm dimension, $r(94) = .673$ and $r(92) = .780$ for the first and second lesson presentations, respectively; for the nervous–calm dimension, $r(92) = .735$ and $r(90) = .767$ for the first and second lesson presentations. The correlation between ratings on these two dimensions was $r(98) = .394$ for the first lesson presentation and $r(98) = .502$ for the second lesson presentation.

To assess reliability of the feedback measures, a second set of feedback data was coded from the audiotape for a subsample of subjects giving their first lesson presentation. For the praise incidence measure, $r(35) = .946$; for praise feedback length, $r(35) = .949$; for incorrectness feedback length, $r(28) = .322$; for positive feedback omission, $r(33) = .856$; for negative feedback omission, $r(33) = .450$.[1]

Correlations among feedback measures showed that indeed these indicators are minimally redundant. There is essentially no relation between rates of omission of positive and negative feedback, and relations between feedback incidence and feedback length are weak and inconsistent in direction.

[1] The modest reliability coefficients on the two feedback omission measures reflect the fact that the subjects' verbalizations were often ambiguous, leaving interpretation dependent on the subtleties of vocal inflection. The low incorrectness-feedback-length coefficient reflects the relative rarity of incorrect "pupil" responses, along with the subjects' tendency to follow such responses with lengthy statements in which the line between feedback and exposition was unclear.

Results

An overview of the analysis of variance effects to be reported and discussed is presented in Table 1.

Behavioral Mediators of Ability-Expectancy Effects

Main effects of pupil ability in the expected direction were shown on central indicators of two of the four categories of mediators suggested by Rosenthal (1973)—the input and output factors. Subject teachers in the low-ability condition taught at a slower pace and thus presented less material to their pupils in the alloted time ($M = 103.3$) than did those in the high-ability condition ($M = 115.7$), $F(1, 84) = 5.45$, $p = .022$, $MS_e = 1{,}377.1$, $d = .509$.[6] Perhaps more interestingly, subjects in the low-ability condition were less patient in waiting for responses to questions, interrupting the latency period between question and response ($M = .39$) more frequently than subjects in the high-ability condition ($M = .23$), $F(1, 83) = 5.06$, $p = .028$, $MS_e = 2{,}400.8$, $d = .494$.

Among the several feedback indicators, there were main effects of pupil ability on only one: Statements praising pupils after correct responses were more lengthy among subjects in the low-ability condition ($M = 2.87$) than among those in the high-ability condition ($M = 2.03$), $F(1, 78) = 7.24$, $p = .009$, $MS_e = 447.1$, $d = .609$. The null results on the feedback incidence measures represent a failure to confirm Rosenthal's (1973) suggestion that high-ability pupils receive clearer and more consistent feedback from teachers,[7] but the significant effects on feedback length do not represent a reversal of his prediction: Average length of positive feedback was not an aspect of teacher behavior discussed by Rosenthal or assessed in the studies he reviewed, and intercorrelations among measures in the present study showed feedback length to be unrelated to those incidence measures Rosenthal did discuss.

Rosenthal's fourth factor, climate, was the one category that showed no significant main effects of pupil ability, although, as will be reported later, there was a highly interesting Race × Ability interaction on a primary climate indicator, vocal warmth.

The effects of the ability manipulation on the measures of vocal anxiety suggest a category of expectancy mediator that Rosenthal did not discuss. Low-ability subjects made fewer speech errors ($M = 3.15$) than did high-ability subjects ($M = 4.14$), $F(1, 84) = 4.72$, $p = .033$, $MS_e = 10.05$, $d = .474$. And there was a nearly significant effect of pupil ability on rated vocal nervousness, the content-filtered speech of subjects in the low-ability condition being rated as less nervous ($M = 5.36$) than that of those in the high-ability condition ($M = 5.08$), $F(1, 84) = 3.25$, $p = .075$, $MS_e = 12{,}018$, $d = .393$.[8]

In summary, these data support Rosenthal's hypothesis that low-ability pupils are taught less and are given fewer response opportunities than are high-ability pupils. The low-ability group did not give less clear or consistent feedback in the present study; praise given after correct responses was more extensive in this group than among the high-ability subjects. And the climate provided by the low-ability group as a whole was not less warm than that provided by the high-ability group. Vocal anxiety is suggested by these results to be an additional category of teacher

---

[6] All analysis of variance results reported are from repeated measures analyses with four between-subjects factors (pupil ability, pupil race, pupil sex, and awareness) and one within-subjects factor (Time, i.e., first vs. second lesson presentation).

Throughout the article $d = 2\sqrt{F \div df}$ for error term. Thus any given between-groups difference is standardized by the within-cell standard deviation estimated from the sums of squares of the error term of the corresponding F. Conventionally, $d = .20$ is considered a small effect, $d = .50$ a medium effect size, and $d = .80$ large (Cohen, 1969, p. 38). Minor fluctuations in degrees of freedom for these analyses reflect occasional missing data.

[7] The feedback-omission indicators are the most relevant to Rosenthal's hypothesis, and the modest reliabilities of these measures suggest one possible explanation for the absence of the predicted ability effects.

[8] One reader suggested that the indications of anxiety among subjects in the high-ability condition might be tied to the relatively rapid teaching pace of this group. Intercorrelations show, however, that there was no substantial relation between teaching pace and either speech errors or vocal nervousness.

expectancy mediator; the subjects teaching low-ability pupils manifested less nervousness.[9]

Pupil-Race Effects

The subjects' ratings of their pupils, mentioned earlier as a means of assessing the potency of the ability manipulation, revealed those in the black-pupil condition to profess superior performance expectations of their pupils. Subjects who believed their pupils were black indicated in their ratings that they expected the lesson to be easier for their pupils ($M = 67.4$) than did those in the white-pupil condition ($M = 57.6$), $F(1, 77) = 5.38$, $p = .023$, $MS_e = 807.4$, $d = .529$.

However, the teacher *behaviors* that are the dependent variables of the present study did not reveal a pattern of particularly favorable treatment directed at black pupils. (The only effect plausibly interpretable as favorable to blacks is a trend for those with black "pupils" to give briefer feedback after incorrect responses ($M = 2.94$) than those with white "pupils" ($M = 4.65$), $F(1, 42) = 3.94$, $p = .054$, $ME_e = 2{,}026.2$, $d = .613$. The two other main effects of pupil race on teaching behavior suggest negative teacher orientation toward black students, if anything. Omission of positive feedback after correct responses was more frequent in the black-pupil condition ($M = 5.49$) than in the white-pupil condition ($M = 2.45$), $F(1, 84) = 3.76$, $p = .056$, $MS_e = 119.7$, $d = .423$. And subjects with black "pupils" were less likely to make helpful slips of the tongue that gave away answers ($M = .04$) than were those with white "pupils" ($M = .15$), $F(1, 84) = 6.04$, $p = .017$, $MS_e = .084$, $d = .536$.[10] These effects should be interpreted, however, in conjunction with the Race × Sex interaction on these two measures.

The Interactive Effects of Pupil Race and Sex

Pupil race and sex were found to interact in affecting both positive feedback omission and helpful slips. Relevant cell means are presented in Table 2. The subjects in the black-male and white-female groups were less likely to offer their pupils either kind of assistance—positive feedback after correct responses, $F(1, 84) = 4.00$, $p = .049$, $MS_e = 119.7$, $d = .436$, or helpful slips while asking questions, $F(1, 84) = 7.30$, $p = .009$, $MS_e = .084$, $d = .590$[11]—than were the subjects in the white-male and black-female groups. The combined impact of these interaction effects and the main effects of pupil race on these two variables is evident in Table 2.

Table 2
Race × Sex Interactions

Pupil race	Pupil sex: Male	Pupil sex: Female
Positive feedback omission[a]		
Black	7.78	3.19
White	1.61	3.29
Helpful slips[b]		
Black	.02	.07
White	.24	.06

[a] Scores summarized by these cell means are percentages of correct pupil responses that met with no teacher feedback. The actual range of scores is from 0 to 70.

[b] Scores summarized by these cell means represent total number of times a subject made an apparent slip of the tongue and gave away the answer while trying to ask a question. The range of scores is from 0 to 2.

[9] As mentioned earlier, 44 subjects reported suspicions that the purpose of the study was to investigate effects of pupil background on teacher behavior. Supplementary analyses showed that the ability effects reported here were more pronounced among the naive subjects than in this suspicious group for all except one dependent measure, praise feedback length. On the praise-feedback-length indicator, the ability effect within the naive group was large enough to have been significant had the entire group of subjects remained naive. Apparently, then, the ability effects reported above are not artifacts created by cooperative subjects.

[10] The helpful slips measure is badly skewed, since only 17 of the 100 subjects made any such slips, so the probability associated with the F should be accepted with caution. To confirm the conclusion that the effect was significant, the likelihood of the observed distribution of those 17 cases between pupil-race conditions was assessed using chi-square. For this effect $\chi^2(1) = 7.56$, $p < .01$.

[11] For this effect, $\chi^2(1) = 5.74$, $p < .025$. (See explanation in Footnote 10.)

Table 3
Race × Awareness × Time Interactions

Group	Time 1	Time 2
Positive feedback omission[a]		
Black		
Aware	11.1	5.4
Unaware	3.4	2.0
White		
Aware	1.2	1.7
Unaware	6.8	.0
Impatience[b]		
Black		
Aware	.37	.32
Unaware	.31	.34
White		
Aware	.24	.32
Unaware	.32	.28

[a] Scores summarized by these cell means are percentages of correct pupil responses that met with no teacher feedback. The actual range of scores is from 0 to 70.

[b] Scores summarized by these cell means represent average number of times, per question asked, that subjects interrupted the latency period between question and response called for by the response schedule. The range of scores is from 0 to 1.8.

Positive feedback is most often withheld in the black-male group, most often given in the white-male group; and helpful slips are least frequent in the black-male group, most frequent in the white-male group.

The Awareness Manipulation Test for Repression of Affect

The repression-of-affect idea was directly tested here via the awareness manipulation, and it is the three-way interactions involving pupil race, awareness, and time that announce the verdict. Support for the repression-of-affect notion would be provided by a pattern of such interaction effects showing controllable, positive behaviors to increase disproportionately among subjects in the aware – black-pupil condition after the awareness manipulation was given, and uncontrollable negative behaviors similarly to increase among this group. Cell means relevant to the two significant or nearly significant Race × Awareness × Time interactions are presented in Table 3. It can be seen that the awareness manipulation encouraged the decrease of positive feedback omission among those with black "pupils," while depressing such a decrease among those with white "pupils," $F(1, 84) = 5.74$, $p = .019$, $MS_e = 70.5$, $d = .523$. And there was a nearly significant tendency for the awareness manipulation to discourage impatience in subjects in the black-pupil condition but encourage impatience in subjects in the white-pupil condition, $F(1, 83) = 3.51$, $p = .065$, $MS_e = 319.8$, $d = .411$.

These effects suggest that the awareness manipulation may have produced the intended demand characteristics: The manipulation may have increased, for black-pupil-condition subjects relative to white-pupil-condition subjects, the incidence of behaviors widely assumed to represent good teaching practice. However, there is no evidence that the relative advantages that may have been conferred on black pupils by the awareness manipulation were accompanied by disproportionate increases in uncontrollable disadvantageous teaching behaviors. And there are difficulties even in the interpretation of the two effects that did occur: The pattern of means does not suggest that certain demand characteristics were produced uniquely in the aware – black-pupil condition; and premanipulation (Time 1) differences between aware and unaware groups throw doubt on any conclusions about the manipulation's effect. All in all, it must be said that we fail to find any support for the repression-of-affect idea in these data.[12]

[12] It was in the pattern of correlations among measures that Weitz (1972) found evidence for the repression-of-affect proposition. The correlations among measures for the various Race × Awareness subgroups after the awareness manipulation had been implemented were examined here as well. The repression-of-affect prediction would be the following: for some pairs of behaviors, one member controllable and the other not, there would be a correlation in one direction in the white-pupil group, reflecting affective congruence of the two behaviors; for the unaware-black group there would be a weaker correlation, or a correlation in the opposite direction, reflecting the combination of dissembling and leak-

The Interactive Effects of Pupil Race and Ability

It is the interactive effects of pupil race and ability that are probably the most intriguing results of the present study. Cell means relevant to the three such interactive effects that reached or approached significance are presented in Table 4. Relative to those in the other two Race × Ability cells, subjects who believed they were teaching high-ability – black and low-ability – white pupils spoke in warmer tones, $F(1, 84) = 5.39$, $p = .023$, $MS_e = 16{,}227.6$, $d = .507$. Additionally, those in the high-ability – black and low-ability – white pupil groups were more likely to make helpful slips as they asked questions, $F(1, 84) = 5.19$, $p = .026$, $MS_e = .084$, $d = .497$.[13] And, there was a nearly significant tendency for these groups to lean forward less often as they addressed their pupil, $F(1, 82) = 3.95$, $p = .051$, $MS_e = 8{,}666.1$, $d = .439$.[14]

The direction of the Race × Ability effect on forward leaning is puzzling in light of the

Table 4
Race × Ability Interactions

Pupil race	Pupil ability: High	Pupil ability: Low
	Vocal warmth[a]	
Black	5.29	4.80
White	4.84	5.20
	Helpful slips[b]	
Black	.09	.00
White	.09	.16
	Forward leaning[c]	
Black	121.8	148.4
White	134.9	108.1

[a] Scores summarized by these cell means are means of 16 rater judgments of content-filtered excerpts on a 9-point cold–warm scale, where warm is high. The range of scores is from 2.31 to 7.50.
[b] Scores summarized by these cell means represent total number of times a subject made an apparent slip of the tongue and gave away the answer while trying to ask a question. The range of scores is from 0 to 2.
[c] Scores summarized by these cell means represent seconds of leaning toward the observation room window during three 1-min periods of a lesson presentation. The actual range of scores is from 0 to 180.

age; and for the aware-black group, subject to still stronger demand characteristics, there would be even greater deviation from the white-condition relationship. Pairs of measures were judged to represent this pattern if the correlations of the aware-white and unaware-white groups were not significantly different from each other, if the aware-black group was significantly different from the white group, and if the correlation of the unaware-black group was intermediate and not significantly different from the pooled correlations of the aware – black-pupil and aware – white-pupil conditions.

Requirements for this pattern were technically met by three pairs of variables in the present study, but none of the three cases provides clear support of the Weitz repression and leakage idea. The correlation between speech errors and persistence was negative in the white condition, $r(50) = -.275$, weak and positive in the unaware-black group, $r(19) = .184$, and positive and slightly stronger in the black-aware group, $r(25) = .278$. However, it is not clear that persistence and speech errors have contradictory affective implications. Persistence and head nodding were the second pair of measures that met the pattern: The correlation between them was weak and positive, $r(49) = .157$, in the white-pupil condition, near zero, $r(19) = .005$, in the black-unaware condition, and moderate and negative, $r(25) = -.422$, in the aware-black condition. The third pair was head nodding and positive feedback omission: The correlation was near zero, $r(24) = .022$, in the white-pupil condition, weak and positive, $r(19) = .116$, in the unaware-black condition, and moderate and positive, $r(25) = .440$, in the aware-black condition. In the case of the second and third pairs, it is possible to interpret the pattern of behavior typical of the aware-black group as affectively incongruent—head nodding being associated with lack of persistence in the first case and with omission of positive feedback in the second. However, for both of these pairs of variables, the differences between the correlations for the white and the unaware-black groups were trivial. And the Weitz discussion suggests that evidence of the dissembling and leakage phenomenon should be detectable even in the absence of conditions such as the awareness manipulation in the present study, which may heighten it.

In short, evidence from subgroup correlations offers no stronger support for the repression-of-affect idea than did interactive effects of pupil race, awareness, and time.

[13] $\chi^2(1) = 3.48$, $p < .10$. (See explanation in Footnote 10.)

[14] Supplementary analyses (see Footnote 9) showed these effects to be stronger for naive subjects than for suspicious subjects, except in the case of the helpful-slips measure.

Table 5
Sex × Ability Interactions

Pupil ability	Pupil sex: Male	Pupil sex: Female
Praise feedback incidence[a]		
High	67.3	58.3
Low	56.5	74.2
Helpful slips[b]		
High	.08	.11
Low	.17	.02

[a] Scores summarized by these cell means are percentages of correct responses followed by explicit praise. The actual range is from 0 to 100.
[b] Scores summarized by these cell means represent total number of times a subject made an apparent slip of the tongue and gave away the answer while trying to ask a question. The range of scores is from 0 to 2.

vocal-warmth and helpful-slips effects. It appears that leaning should not be assumed to serve as an alternate indicator of warmth. The negative correlations between the two measures, $r(93) = -.144$, $p < .10$, for the first lesson presentation, and $r(91) = -.249$, $p < .05$, for the second lesson presentation, support this conclusion. Future investigation of the correlates of leaning behavior seems in order. In the meantime, we can note that the more direct and interpretable indicator of climate, vocal warmth, along with inadvertent helpfulness, was mostly likely to exist among those in the high-ability – black and low-ability – white conditions. This finding seems to contradict the Rubovits and Maehr (1973) conclusion that high-ability – black and low-ability – white groups are the recipients of the most detrimental teacher behaviors.

Pupil Sex Effects and the Interactive Effects of Sex and Ability

Main effects of pupil sex are negligible. The only behavior showing a significant effect of pupil sex was head nodding, girls receiving more frequent up-and-down head nods ($M = 2.66$) than boys ($M = 1.62$), $F(1, 82) = 5.19$, $p = .026$, $MS_e = 10.1$, $d = .503$. And even this effect is difficult to interpret. Head nodding had been assumed to be an indicator of climate or warmth, but that assumption is thrown into serious question by the fact that head nodding did not show substantial correlations with any of the other kinesic measures or with ratings of vocal warmth. We thus must say that there were no readily interpretable effects of pupil sex on teacher behavior. This finding is congruent with the conclusions of Brophy and Good (1974) that such pupil-sex effects on teacher behavior as do exist in classroom situations are the product of actual behavior differences among male and female pupils.

Interactive effects of pupil sex and ability, like main effects of pupil sex, are neither numerous nor readily interpretable. Relevant cell means are presented in Table 5. On the one hand, the high-ability – female and low-ability – male groups revealed a lower incidence of praise feedback than did the other two Sex × Ability groups, $F(1, 84) = 7.25$, $p = .009$, $MS_e = 1{,}194.4$, $d = .588$.[15] However, it was also subjects in the high-ability – female and low-ability – male groups who were most likely to make helpful slips of the tongue, $F(1, 84) = 5.19$, $p = .026$, $MS_e = .084$, $d = .497$.[16] Thus, we see the least intentional encouragement, in the form of praise after correct responses, being directed to those groups that receive the greatest unintentional assistance in the form of helpful slips.

Discussion

The Mediation of Teacher Expectancy Effects

The present findings provide strong support for Rosenthal's (1973) conclusion that two of the modes in which teachers may adapt their behavior to reflect differential expectations of their students are by adjusting the

[15] Supplementary analyses (see Footnote 9) showed the effect on helpful slips to be stronger among naive subjects than among suspicious subjects. The interaction effect on praise feedback incidence is slightly stronger for suspicious subjects than for naive subjects, but the effect within the naive group was large enough to have been significant had the entire group of subjects remained naive.

[16] $\chi^2(1) = 3.48$, $p < .10$. (See explanation in Footnote 10.)

pace of their teaching and by varying the response opportunities allowed the children.

The pattern of feedback behavior noted here merits more detailed comment. First, the virtual independence of the various measures of teacher feedback suggests that it is not appropriate simply to speak of feedback as being more or less good, or strong, or clear. Differentiated questions must be asked about rates of teacher feedback, quality of feedback, and extensiveness of feedback. The finding that subjects in the low-ability condition offered longer statements of praise than did those in the high-ability condition was unexpected; however, several plausible interpretations suggest themselves. There may have been a contrast effect at work, a tendency for subjects to express surprise at the accomplishment of low-ability pupils by giving elaborate praise. Or, the subjects may have been taught self-consciously to use praise as a shaping technique, and perhaps were most diligent in the application of this technique when addressing a low-ability pupil. Alternately, focusing on the relative brevity of praise feedback given in the high-ability group, perhaps subjects who believed their pupils to be bright functioned under pressure of demand characteristics dictating that the measure of their teaching ability was the amount of material covered; cutting short the positive verbal feedback was one way of speeding things up. (Correlations between teaching pace and length of praise feedback are negative and moderately strong.)

Similarly, it is not difficult to find plausible explanations for the pupil-ability effect on vocal nervousness. Teachers may see low-ability pupils as less threatening than high-ability pupils, and thus may approach them with less anxiety.

The absence of pupil-ability effects paralleling those found on climate indicators in prior research remains unexplained. Perhaps climate was the aspect of teacher behavior least amenable to natural expression in the absence of actual face-to-face interaction. The results here do suggest that the various kinesic and paralinguistic measures assumed to represent climate do not necessarily correlate strongly or even positively with each other, and therefore should not be assumed to be indicators of the same construct.

It must be remembered that only half of the mediation question was addressed in the present study. We have examined effects of pupil ability on teacher behavior, but behaviors spotlighted as mediators of teacher expectancy effects must have their impact on student performance, as well as show an impact of pupil-ability label. The teacher behaviors that showed pupil-ability effects here probably vary in their influence on pupil performance. In actual classroom situations, it may often be sound pedagogical practice to adjust lesson pacing in accordance with pupil ability. Thus, the effect of pupil-ability label on teaching pace may only be a problem where ability assessments are faulty. In contrast, the observed tendency to give children with low-ability labels less time to answer questions must clearly be dysfunctional for those children, the more so because the correlational analysis showed interruptions of the latency period between question and response often to take the form of additional questions, requiring different responses. Bombardment with a series of follow-up questions, before the pupil has had the chance to respond to the original question, would make successful performance difficult for any child; for the child who may indeed already have learning problems, such teacher behavior may present an insurmountable obstacle to successful performance. Length of praise feedback and vocal nervousness fall into still a third category, showing effects of pupil-ability label but not having prima facie positive or negative implications for pupil performance.

Clearly, the results of numerous field and lab studies, examining effects of teacher behaviors as well as their causes, must be accumulated before we can be satisfied with our understanding of the mediation of teacher expectancy effects.

The Joint Effects of Pupil Race and Sex

The effects of pupil race on teacher behavior in the present study were limited. There does seem to be a tendency for black "pupils" to be recipients of the less positive teacher behaviors. Perhaps more interesting, however, is

the observation that on the two indicators that appeared to show a bias against black pupils —the positive-feedback-omission and helpful-slips variables—there were also Race × Sex interactions, the two effects together implying that white male students receive the most favorable treatment, black male students, the most unfavorable. The parallel between this result and recent Census Bureau statistics is noteworthy: In 1974, educational attainment was highest among white males, lowest among black males (with white and black females intermediate) (Farley, 1977).

The Repressed-Affect Model

It is encouraging that the awareness manipulation, designed to evoke conscious efforts at problack behavior, did not also create detectable leakage of negative affect. The explanation for the incongruence of the present results and those of Weitz (1972) remains to be ascertained.

The Null Results of the Pupil-Sex Manipulation

It is the lack of a pronounced pattern of pupil-sex or Sex × Ability effects that is interesting here. Perhaps Brophy and Good (1974) are correct in suggesting that classroom teachers merely respond to behavioral differences in male and female pupils, differences that originate elsewhere. Other possibilities merit exploration, however. Certain kinds of pressure for sex role specialization intensify as children enter adolescence. It may be the case that differential teacher treatment of male and female pupils is pronounced, but for children older than the phantom 6-year-old of this study. It may also be the case that effects of pupil sex do exist, even among teachers of young children, but on dimensions of behavior different from those selected for attention in the present study. For example, Dweck (1976) discusses the differential rates of praise and criticism teachers direct at male and female children for nonintellectual aspects of their performance. Dweck believes that experience with nonspecific feedback affects pupil sensitivity to positive and negative feedback on intellectual tasks, and that the sex difference in receipt of nonspecific feedbac creates a sex difference in response to fail ure. This potentially important dimension o teacher behavior, nonspecific feedback, is on of those that obviously could not be represente here, since our teacher subjects had the op portunity to give feedback specific only t intellectual tasks.

The Race by Ability Interaction Effect o Vocal Warmth

One of the most interesting observations o the present study is that pupil race and ability interacted to influence teacher affective orientation in a direction favoring high-ability – black and low-ability – white pupil groups. As noted earlier, this effect is opposite in direction from that found by Rubovits and Maehr (1973). Why might teachers be positively disposed toward high-ability – black and low-ability – white pupils? Findings of Datta et al. (1968) suggest an explanation.

When Datta et al. asked a predominantly white group of seventh-grade teachers to rate selected black and white pupils in their classes on a variety of dimensions, the high-ability – black and low-ability – white groups were rated higher on positive task orientation and lower on verbal aggression than were low-ability blacks and high-ability whites.

Intuition suggests that most teachers would prefer pupils to manifest the orderly and submissive patterns of behavior represented by high task orientation and low verbal aggressiveness, and in this case there is ample empirical evidence to support the intuitive conclusion. Feshbach (1971) finds, for example, that teachers most prefer rigid, conforming, and orderly pupil descriptions, and least prefer those that are independent, active, and assertive. The interaction effect noted in the present study may, then, reflect teacher anticipation that high-ability blacks and low-ability whites will be relatively trouble-free students, and low-ability blacks and high-ability whites relatively troublesome.

External Validity and the Need for Naturalistic Research

The limitations of this phantom pupil design must be acknowledged. In the present

study we tried to simulate one-to-one interaction instead of the typical group teaching situation, and for one of the interactants we depended on our cover story's credibility, the subjects' imaginations, and a remote-control light panel. The reader is invited to be as concerned about the external validity of the study as we are. On the other hand, as in virtually all social psychology research, we necessarily had to accept some limitations in order to avoid others. A great deal of research has been done in more naturalistic settings where, typically, the lack of control precluded some behavioral measures and meant that extraneous pupil characteristics were possibly confounded with those of primary interest. In order to employ the range of refined measures represented here, and to be assured that we could look at clean effects of pupil race, sex, and ability, a design was selected that leaves open some questions about generalizability. If we are to reap maximum value from this study's strengths and minimize the ramifications of its potential weaknesses, the present results must be considered along with evidence from more naturalistic classroom studies.

Concluding Comments

Directions that could profitably be taken in future research are undoubtedly obvious to the reader. There is the need for further data to be gathered in actual classrooms and in controlled laboratory settings, focusing on black teachers and males, as well as on white female teachers. Conceptualization and empirical work relevant to the dimensionality of teacher behavior will provide valuable additions to existing understandings. And, if we are to talk of the impact that various teacher orientations may have on pupil performance, there needs to be much more extensive evidence collected concerning the effect of teacher behaviors on pupils.

References

Antonoplos, D. P. Interactions of teacher–pupil sex as expressed by patterns of reinforcement and judgment about pupils: A national study. (Doctoral dissertation, Indiana University, 1971). *Dissertation Abstracts International*, 1972, *32*, 6117-A.

Bateson, G., Jackson, D., Haley, J., & Weakland, J. Towards a theory of schizophrenia. *Behavioral Science*, 1956, *1*, 251–264.

Brophy, J. E., & Good, T. L. *Teacher–student relationships: Causes and consequences*. New York: Holt, Rinehart & Winston, 1974.

Coates, B. White adult behavior toward black and white children. *Child Development*, 1972, *43*, 143–154.

Cohen, J. *Statistical power analysis for the behavioral sciences*. New York: Academic Press, 1969.

Crano, W. D., & Mellon, P. M. Causal influence of teachers' expectations on children's academic performance: A cross-lagged panel analysis. *Journal of Educational Psychology*, 1978, *70*, 39–49.

Crowl, T. K., & MacGinitie, W. H. The influence of students' speech characteristics on teachers' evaluation of oral answers. *Journal of Educational Psychology*, 1974, *66*, 304–308.

Datta, L., Schaefer, E., & Davis, M. Sex and scholastic aptitude as variables in teachers' ratings of the adjustment and classroom behavior of Negro and other seventh-grade students. *Journal of Educational Psychology*, 1968, *59*, 94–101.

Dietz, S. M., & Parkey, W. W. Teacher expectations of performance based on race of student. *Psychological Reports*, 1969, *24*, 694.

Dweck, C. Children's interpretation of evaluative feedback: The effect of social cues on learned helplessness. *Merrill-Palmer Quarterly*, 1976, *22*, 105–123.

Farley, R. Trends in racial inequalities: Have the gains of the 1960s disappeared in the 1970s? *American Sociological Review*, 1977, *42*, 189–208.

Feshbach, N. D. Student teacher preferences for elementary school pupils varying in personality characteristics. In M. L. Silberman (Ed.), *Experience of schooling*. New York: Holt, Rinehart & Winston, 1971.

Guilford, J. P. *Psychometric methods*. New York: McGraw-Hill, 1954.

Harvey, D. G., & Slater, G. T. The relationship between child's SES and teacher expectations: A test of the middle-class bias hypothesis. *Social Forces*, 1975, *54*, 140–159.

Henderson, E. H., & Long, B. H. Academic expectancies of black and white teachers for black and white first graders. *Proceedings of the 81st Annual Convention of the American Psychological Association*, 1973, *8*, 685–686. (Summary)

Jensen, M., & Rosenfeld, L. B. Influence of mode of presentation, ethnicity, and social class on teachers' evaluations of students. *Journal of Educational Psychology*, 1974, *66*, 540–547.

Leacock, E. *Teaching and learning in city schools*. New York: Basic Books, 1969.

Long, B. H., & Henderson, E. H. Certain determinants of academic expectancies among southern and non-southern teachers. *American Educational Research Journal*, 1974, *11*, 137–147.

Mazer, G. E. Effects of social-class stereotyping on teacher expectation. *Psychology in the Schools*, 1971, *8*, 373–378.

Miller, H. L. Race vs. class in teachers' expectations. *Psychological Reports*, 1973, *32*, 105–106.

Rogers, P. L., Scherer, K. R., & Rosenthal, R. Content-filtering human speech. *Behavioral Research Methods and Instrumentation*, 1971, *3*, 16–18.

Rosenthal, R. On the social psychology of the self-fulfilling prophecy: Further evidence for Pygmalion effects and their mediating mechanisms. *MSS Modular Publications*, 1973 Module 53, 1–28.

Rosenthal, R. *Experimenter effects in behavioral research* (enlarged edition). New York: Irvington, 1976.

Rosenthal, R., & Jacobson, L. *Pygmalion in the classroom*. New York: Holt, Rinehart & Winsto 1968.

Rubovits, P. C., & Maehr, M. L. Pygmalion blac and white. *Journal of Personality and Social Psychology*, 1973, *25*, 210–219.

Weitz, S. Attitude, voice, and behavior: A represse affect model of interracial interaction. *Journal Personality and Social Psychology*, 1972, *24*, 14–21

Yee, A. Interpersonal attitudes of teachers and ad vantaged and disadvantaged pupils. *Journal of Hu man Resources*, 1968, *3*, 327–345.

Received June 16, 1978

Journal of Personality and Social Psychology
1979, Vol. 37, No. 6, 913-929

Attitude Organization and the Attitude–Behavior Relationship

Richard P. Bagozzi and Robert E. Burnkrant
University of California, Berkeley

The validity of a single-component model of attitude was assessed and compared with a two-component (affective/cognitive) conceptualization of attitude. Confirmatory factor analysis was employed in a reanalysis of data previously reported by Fishbein and Ajzen in the *Psychological Review*. Convergent validity, in the sense of uniform consistency of responses, was found not to hold for the single-component model but was achieved for the two-component model. Further, the two-component model of attitude was found to predict scaled, but not unscaled, multiple-act behavior criteria; and the predictive validity of the relationship between the two-component model and single-act behavior criteria was not supported.

Considerable research attention has been directed at the attitude–behavior relationship. Much of this research has yielded disappointingly weak results (e.g., Berg, 1966; Bray, 1950; Kutner, Wilkins, & Yarrow, 1952; La Piere, 1934; Nemeth, 1970). Wicker (1969) reviewed over 30 studies concerned with the attitude–behavior relationship and concluded that in most cases attitude is either unrelated or only slightly related to behavior.

The failure to find consistent, strong relationships between attitude and behavior has led to elaboration and clarification of the attitude construct and to further specification of its relationship to behavior. It is frequently argued that attitude is a multidimensional construct and that simultaneous consideration of these several dimensions will provide better predictions of behavior than the consideration of a single attitudinal dimension or component alone (Krech, Crutchfield, & Ballachey, 1962; Norman, 1975). Other researchers argue that attitude may be appropriately regarded as a single dimension of affect and that the relationship between attitude and behavior would be stronger if both attitudinal predictors and behavioral criteria were measured at an equivalent level of generality (e.g., Fishbein, 1967b; Fishbein & Ajzen, 1975). Fishbein and his associates maintain that a general attitude should be predictive of a general set of behaviors, but it should bear no necessary relationship to any specific behavior the researcher might be interested in (Ajzen & Fishbein, 1977; Fishbein, 1973; Fishbein & Ajzen, 1974).[1]

This article addresses both of these issues. In doing so, it provides empirical tests of the validity of the multicomponent and single-component models of attitude. Second, it tests the validity of the relationship between a general attitude and single-act criteria and between a general attitude and multiple-act criteria. Finally, it examines the relationship between attitude and scaled measures of behavior.

The ordering of the authors' names is alphabetical; both authors contributed equally to the research.

Robert E. Burnkrant is now at the Ohio State University.

Requests for reprints should be sent to Richard P. Bagozzi, who is now at the Alfred P. Sloan School of Management, Massachusetts Institute of Technology, 50 Memorial Drive, Cambridge, Massachusetts, 02139.

[1] Many other factors have been cited to explain the lack of consistently strong positive relationships between attitude and behavior. These will be treated in the discussion section of this article, but for a fuller treatment see Calder and Ross (1973), Fishbein and Ajzen (1975), Kelman (1974), and Wicker (1969).

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3706-0913$00.75

Attitude Organization

Those who hold a multidimensional view of attitude consider it to be a complex construct comprised of an affective (i.e., feeling) component, a cognitive (i.e., knowledge) component, and a conative (i.e., behavioral) component (Krech et al., 1962; Smith, 1947). Rosenberg and Hovland (1960) regard attitude as a predisposition to respond to some class of stimuli with cognitive, affective, and behavioral responses. They suggest that each response class is mediated by a separate component of attitude. Greenwald (1968) maintains that these components may have separate antecedents. He further suggests that affect may be formed through classical conditioning, cognitions through cognitive learning, and conations through instrumental learning or operant conditioning.

It has been pointed out, in accordance with a multicomponent view of attitude, that people may hold the same attitude as measured by one index of attitude and yet hold very different attitudes in terms of their standing on the other unmeasured components (e.g., Rosenberg & Hovland, 1960). It is implied that the attitude–behavior relationship will be stronger when the components are consistent than when they are inconsistent (e.g., Rosenberg, 1968; Rosenberg & Hovland, 1960). Failure to find a consistent, direct relationship between attitude and behavior may be due, then, to failure to measure people's standing on all three components of attitude and employ these as simultaneous predictors of behavior. The three-component view suggests that cognitive, affective, and conative evaluations of objects are distinguishable aspects of attitude and that simultaneous consideration of all three components should be most predictive of overt behavior.

The major alternative to the three-component view treats attitude as a single dimension of affect for or against an object. In accordance with this position, Fishbein (1967b) points out that although definitions of attitude usually refer to three components, it is the evaluative or affective component that is usually measured and treated as attitude. He argues that all attitude-scaling techniques have in common the characteristic that they place individuals on a dimension of affect for or against a psychological object. According to this view, alternative approaches to the measurement of attitude provide alternative measures of the same thing (i.e., affect) and should yield the same results. Obtained differences among alternative attitude instruments would be due, therefore, to measurement error and not the assessment of valid alternative components.

A third and intermediate position is maintained in the present article. Attitude is considered to be a complex construct comprised of cognitive and affective components. These components are believed to simultaneously account for behavioral intentions or actual behavior.[2]

Katz and Stotland (1959) and Rosenberg (1968) assert that all true attitudes must have both cognitive and affective content although they need not include a conative component. Rosenberg (1968) notes that with the exception of cognitive dissonance,

> The authors and proponents of most of the other consistency theories pay ready lip service to the definition of attitude as an internally consistent structure of affective, cognitive and behavioral components. But, in practice, the last of these components is usually slighted. Behavior (in the sense of externally visible, overt action) toward the attitude object is usually relegated to the status of a dependent variable; implicitly it is assumed that the person will simply act toward an attitude object in a manner consistent with his coordinated affective–cognitive orientation toward that object. (p. 101)

The two-component attitude position taken here recognizes and is consistent with the fact that self-reported behavior and stated intentions to respond have frequently been treated as dependent effects of affective and/or cognitive variables (e.g., DeFleur & Westie, 1958; Linn, 1965; Nemeth, 1970; Rogers & Thistlethwaite, 1970; Sample & Warland, 1973; Tittle & Hill, 1967). Intentions seem to be at a

[2] It is also recognized that behavior may lead to changes in the affective and cognitive components of attitude as a result of learning and self-perception or dissonance effects. However, explicit consideration of these feedback effects is beyond the scope of this research.

lower level of abstraction (i.e., closer to observable behavior) than cognitions or affect.

Several attemps have been made to provide empirical support for a multicomponent conceptualization of attitude. The results of these studies have been equivocal at best. Woodmansee and Cook (1967) factor analyzed scales designed to measure attitude toward blacks. They failed to obtain unique factors that could be interpreted as representing cognitive, affective, and conative components of attitude. Based on judgment procedures, Ostrom (1969) and Kothandapani (1971) developed Guttman, Likert, Thurstone, and Guilford self-rating attitude scales for each component of attitude. People then responded to cognitive, affective, and conative scales. Using Campbell and Fiske's (1959) multitrait–multimethod matrix approach and criteria, both researchers concluded that obtained correlations among the attitude component scores provided evidence for convergent and discriminant validity for the affective, cognitive, and conative components. However, in a reanalysis of both data sets, Bagozzi (1978) found that discriminant validity was obtained for the Ostrom, but not the Kothandapani, data. Thus, only the Ostrom study provides clear support for a multicomponent attitude model.

An objective of the present article is to provide empirical tests of the validity of the two-component and single-component models of attitude. This research involves a reanalysis of data presented by Fishbein and Ajzen (1974). In that study, each subject's attitude toward "being religious" was assessed on Thurstone, Guttman, Likert, semantic differential, and Guilford self-rating scales. In addition, subjects provided responses to either 100 behavioral intention items or 100 self-reported behaviors. While the approach taken by Fishbein and Ajzen was consistent with a unidimensional approach to attitude in that all five attitude scales were treated as alternative measures of attitude, it is our contention that the evaluative dimension of the semantic differential and the Guilford self-rating scale represent alternative measures of the affective component of attitude and that the Guttman, Likert, and Thurstone scales represent alternative measures of the cognitive component of attitude.

Theory and research tend to support this distinction. Katz and Stotland (1959) identified the affective component with attributions of good or bad qualities. Osgood, Suci, and Tannenbaum (1957) defined attitude as the projection of a concept on the evaluative dimension of semantic space. Similarly, various other researchers have presented arguments and evidence for using the evaluative dimension of semantic space as a measure of affect (cf. Davidson & Jaccard, 1975; Fishbein & Raven, 1962; McGuire, 1969; Norman, 1975). In a parallel fashion, the Guilford self-rating favorable/unfavorable scale can be used to measure affect. Conceptual and empirical support for this claim can be found in Ostrom (1969), Kothandapani (1971), and Norman (1975).

The affective component measures the degree of emotional attraction toward an attitude object. The cognitive component, in contrast, accounts for the perceived relationship between attitude object and other objects or concepts. The cognitive component typically covers "beliefs about the object, characteristics of the object, and relationships of the object with other objects" (Ostrom, 1969, p. 16). With regard to the specific cognitions employed in the present research, Fishbein and Ajzen (1974) maintain that "the other three attitude measures [i.e., the Guttman, Likert, and Thurstone scale] were standard religiosity scales based on opinion items" (p. 62). Elsewhere, Fishbein and Ajzen (1975, p. 12) consider opinions to be cognitions, and they distinguish them from affect and conation. In their study, Fishbein and Ajzen (1974) employed a Likert scale from Bardis (1961), a Guttman scale from Faulkner and DeJong (1969), and a Thurstone scale from Poppleton and Pilkington (1963). Examination of these scales shows them to be consistent with the preceding definitions of the cognitive component. For instance, the statements employed are of the following form:

1. "Religious faith is merely another name for belief which is contrary to reason" (Poppleton & Pilkington, 1963, p. 34).

2. "Religious truth is higher than any other

form of trust" (Faulkner & DeJong, 1969, p. 567).

It is possible to distinguish empirically between the single-component and two-component views of attitude using the data of Fishbein and Ajzen (1974). To do this, one may examine and compare the convergent validity of each model. Although the Campbell and Fiske (1959) criterion of high correlations among the measures of each component provides one indication of convergent validity, the procedure is an arbitrary one, and the more rigorous confirmatory factor analysis methodology was employed in the present study. Among other benefits (cf. Bagozzi, 1978), Kenny (1976) notes the following advantages of confirmatory factor analysis:

> The application of confirmatory factor analysis to the multitrait–multimethod matrix has a number of advantages over the traditional Campbell-Fiske set of criteria: (a) Confirmatory factor analysis gives estimates of parameters while the Campbell-Fiske criteria are only rules of thumb. (b) Significance tests are possible with confirmatory factor analysis. (c) Given marked differences in the reliability of measures, the Campbell-Fiske criteria are misleading . . . while confirmatory factor analysis takes into account differential reliability. (p. 248)

To test for convergent validity using confirmatory factor analysis, the null hypotheses expressed in Figure 1 must be examined. Panel a of Figure 1 shows the null hypothesis for the single-component model. Under the single-component model, for convergent validity to be sustained, it is necessary that the responses of subjects to all attitude measures achieve a high degree of correspondence. Rather than using the criterion of correspondence suggested by Campbell and Fiske, which is not as stringent as confirmatory factor analysis and which can be misleading, as noted above, the likelihood ratio chi-square goodness-of-fit test can be used. Similarly, Panel b of Figure 1 displays the null hypothesis for the two-component model. Under this model, convergent validity will be obtained when the responses of subjects achieve a high degree of correspondence for (a) the affective component and (b) the cognitive component *and* when the correlations of attitude measures across the two components are both consistent in their pattern and significantly lower in magnitude than the within-component attitude measure correlations. Again, the likelihood ratio chi-square goodness-of-fit test can be used to test this hypothesis. The formal statement of the convergent validity null hypotheses will be made below under Method.

The data of Fishbein and Ajzen (1974) do not permit one to distinguish between the prediction that behavior is accounted for by intentions alone and the prediction that behavior is accounted for by the simultaneous consideration of the conative, cognitive, and affective components of attitude. This is because behavior and behavioral intention measures were not both obtained from the same individuals.[3] However, a finding that the cognitive and affective components of attitude simultaneously account both for behavioral intention and for verbal reports of past behavior would be consistent with the two-component view suggested here.

Attitude–Behavior Relationship

It has been argued (e.g., Ajzen & Fishbein, 1977; Fishbein, 1967a, 1973; Fishbein & Ajzen, 1974) that a major reason for the weak relationship obtained between attitude and behavior is the lack of correspondence between attitudinal predictors and behavior criteria. Fishbein (1973) notes that although a relationship is frequently assumed between attitude and whatever behavior the researcher happens to be interested in, there is no necessary relationship between a general attitude toward an object and the performance of a specific behavior with respect to that object. However, a general attitude toward an object should be predictive of the overall pat-

[3] Fishbein and Ajzen (1974) obtained self-reports of behavioral intentions from a sample of 63 subjects and self-reports of past behavior from a separate sample of 62 subjects. They argue that while both of these behavioral criteria are verbal responses, self-reports of past behavior go beyond intentions to provide meaningful approximations of behavior. More recently, Weigel and Newman (1976) employed overt behaviors as criteria, and they obtained results that are consistent with those obtained by Fishbein and Ajzen.

terns of behavior engaged in with respect to that object. In their most complete elaboration of this view, Ajzen and Fishbein (1977) maintain that "the strength of an attitude–behavior relationship depends in large part on the degree of correspondence between attitudinal and behavioral entities" (p. 891). When correspondence is high, the strength of the relationship between attitude and behavior should also be high. They suggest that attitudinal and behavioral entities may be considered in terms of four different elements: an action element, a target toward which the action is directed, a context in which the action is performed, and a time at which the action is performed. As a result, an attitude toward a specific act to be engaged in at a given time in a specific context should be predictive of the actual performance of that act in that time and situation.

However, an attitude toward an object (i.e., time, action, and context are unspecified) should not necessarily predict any particular behavior logically related to the object. One would not, therefore, necessarily expect a correlation between attitude toward an object and single-act behavior criteria or even multiple-act behavior criteria unless the content of the criteria were scaled at comparable levels of generality and/or the multiple-act criteria encompassed a wide range of relevant behaviors. The attitude employed in the present study was attitude toward "being religious" and is thus nonspecific with respect to action, context, and time. Consequently, it may be hypothesized that attitude will predict scaled multiple-act criteria but not necessarily single-act criteria or unscaled multiple-act criteria.

Fishbein and Ajzen (1974) have shown support for their contention that high correlations between measures of attitude and behavior should be expected only under conditions of high attitude–behavior correspondence. Average correlations between attitude and single-act criteria were low, ranging from .121 to .202. When a multiple-act criterion was constructed by summing responses to 100 behavioral intention statements or 100 self-reports of past behavior, the correlations between attitude and this criterion were high, ranging from .604 to .749. Similar results were obtained by Weigel and Newman (1976) in an assessment of overt behaviors obtained 3 months after administration of an attitude instrument. Fishbein and Ajzen also found that correlations between attitude and multiple-act criteria tended to be equally high and significant when the behavioral criteria were scaled in accordance with Guttman, Likert, or Thurstone scaling procedures.

It is rather surprising that correlations between attitude and scaled multiple-act criteria were not stronger than correlations obtained between attitude and a simple summative criterion. As Fishbein and Ajzen (1974) pointed out, "Many (if not most) behaviors that are performed with respect to or in the presence of a given object, *cannot* be considered valid indicants of a person's attitude towards that object" (p. 66). Use of a standard scaling procedure "is one way of insuring that the behavioral items selected constitute valid behavioral criteria, that is, that they are in some way related to the attitude under consideration" (Fishbein & Ajzen, 1974, p. 65). Fishbein and Ajzen reported that when scaling was employed, most of the 100 items were rejected by all three scaling procedures. This would suggest that most behaviors (or intentions) were unrelated to people's attitudes toward being religious. These behaviors may be performed for essentially nonattitudinal reasons. The inclusion of these behaviors in a multiple-act criterion would incorporate considerable error in the instrument. We would expect, therefore, that a strong relationship would be obtained between attitude and behavior only when both attitude and the multiple-act criterion are appropriately scaled (see discussion below and under Method).

To test for the predictive validity of the attitude–behavior relationship using confirmatory factor analysis, the null hypotheses of Figures 2 and 3 must be investigated. Although a formal, detailed description of these hypotheses will be provided in the Method section, for now it should be noted that the path diagrams provide a representation of the relationship between attitude as a *construct* and the appropriate behavior criterion. The construct to use in these hypotheses is the one achieving convergent validity. Intuitively, the

null hypotheses entailed by Figures 2 and 3 state that the more favorable the attitude, A, the greater the number of behaviors, B, performed. The parameter γ represents the degree of predictive validity and is analogous to a beta coefficient in regression. It should be stressed, however, that the hypotheses in confirmatory factor analysis are more stringent than those suggested by, say, correlation or regression analyses. That is, a failure to reject the null hypotheses implied by Figures 2 and 3 requires that the relationship γ be statistically significant *and* that (a) the intercorrelations among the attitude measures of A achieve a high degree of correspondence, (b) each measure of A relate to the behavioral criteria in a statistically significant and sufficiently large way, and (c) the pattern of intercorrelations between measures of A and B be uniform. The likelihood ratio chi-square goodness-of-fit test can be used to evaluate this hypothesis.

Notice that it is possible for each individual attitude measure to correlate highly with the behavior criterion, yet if the pattern of intercorrelations is not consistent, then one would reject the null hypothesis. This condition is analogous to Campbell and Fiske's (1959) rule of thumb that "the same pattern of trait interrelationship be shown in all of the heterotrait triangles of both the monomethod and heteromethod blocks" (p. 83). This aspect of the null hypotheses is also what is meant, in part, by Kenny's (1976) observation that confirmatory factor analysis "takes into account differential reliability" (p. 248). With regard to the attitude–behavior relationship, it is hypothesized that the attitude *construct* will not predict the single-act or multiple-act behavior criteria even though correlations between individual measures of attitude and single acts or multiple acts might be significant. Because the correspondence between the attitude measures and both the single-act criterion (i.e., the mean correlation with 100 behaviors) and the multiple-act criterion (i.e., the sum over 100 behaviors) is low (in the sense of being at different levels of specificity with regard to time, action, and context), one would not expect the pattern of correlations between measures of A and B to be uniform. Consequently, it is hypothesized that the relationship between attitude and single-act and multiple-act behavior criteria should be rejected. In contrast, it is hypothesized that the relationship between attitude and *scaled* multiple-act criteria will be sustained, because the attitude and behavior measures are formed at comparable levels of specificity.

In sum, we expected that a reanalysis of the Fishbein and Ajzen (1974) data using confirmatory factor analysis would support the following hypotheses:

1. Convergent validity will be obtained when the cognitive and affective measures of attitude are treated as separate components, but convergent validity will not be obtained when all five instruments are treated as alternative measures of the same underlying construct.
2. The attitude *construct* will not predict single-act behavior criteria.
3. The attitude *construct* will not predict multiple-act behavior criteria.
4. The attitude *construct* will predict scaled multiple-act behavior criteria.

Method

Data Source

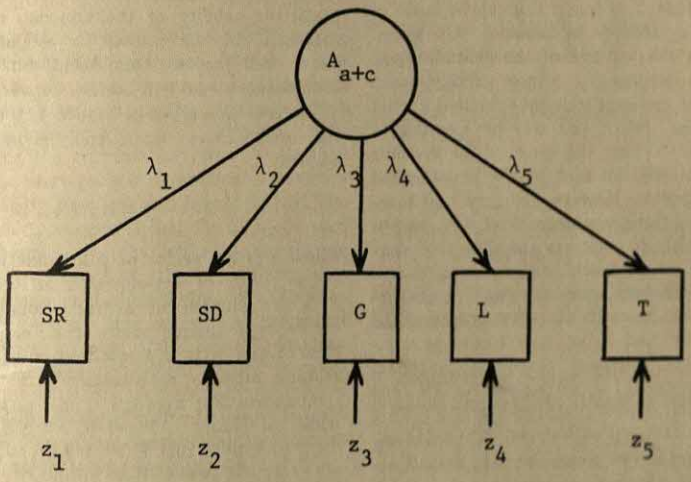
To test the hypotheses, a reanalysis of data reported in Fishbein and Ajzen (1974) was performed. Fishbein and Ajzen had used two samples of respondents. In the first, 62 male and female undergraduates indicated which items from a set of 100 behaviors they had performed (the self-reported behaviors sample). The set of behaviors consisted of a list of 70 actions dealing with religious matters (e.g., pray before or after meals, donate money to a religious institution, etc.) and 30 additional actions in a refusal format selected from the original 70 (e.g., refuse to state a religious preference during university registration). The second sample was composed of 63 male and female undergraduates who indicated which items from the set of 100 behaviors they would perform (the behavioral intentions sample).

In addition, all subjects completed five scales measuring attitudes toward religion. A Guilford self-rating scale measured attitudes towards being religious on an 11-point scale ranging from extremely favorable to extremely unfavorable. A semantic differential scale measured evaluations of "being religious" on five 11-point bipolar scales having the following endpoints: good–bad, harmful–beneficial, wise–foolish, pleasant–unpleasant, and sick–healthy. The Guilford and semantic differential scales are considered in the present study to tap the affective dimensions

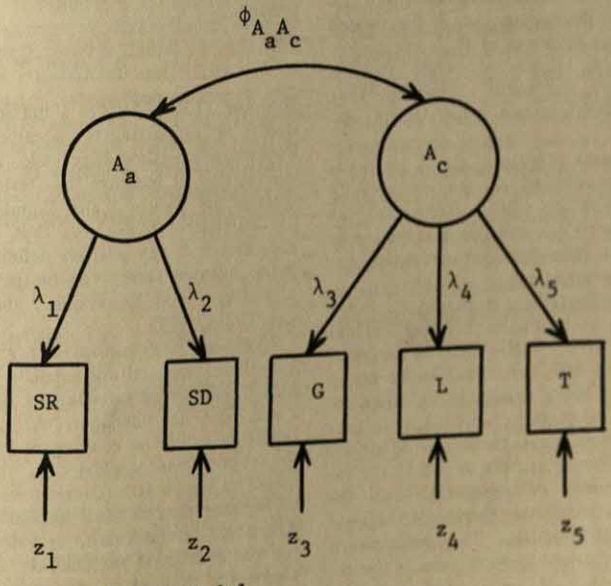
of attitudes. Support for this contention was provided in the preceding sections of this article.

To measure the cognitive component of attitude toward religion, three standard religiosity scales were employed. These included Likert (Bardis, 1961), Guttman (Faulkner & DeJong, 1969), and Thurstone (Poppleton & Pilkington, 1963) scales.

The Fishbein and Ajzen (1974) data set provide a better test of the predictive validity of attitude than the data of either Kothandapani (1971) or Osrom (1969). This is because Kothandapani used a single dichotomous measure of behavior (i.e., used or did not use contraceptives) and Ostrom employed eight measures of religious behaviors but had to omit 44



a. Single-Factor Model



b. Two-Factor Model

Figure 1. Path diagrams for determining convergent validity of attitude measures. (A_{a+c} = single-component attitude construct; A_a = affective attitude component; A_c = cognitive attitude component; SR = Guilford self-report, SD = semantic differential, G = Guttman, L = Likert, and T = Thurstone scales. The remaining symbols represent parameters to be estimated and are defined in the text.)

subjects who completed the attitude scales but failed to complete all of the behavioral self-reports. The Fishbein and Ajzen data offer the advantages of 100 measures of behavior or behavioral intentions with no attrition.

Methods of Analysis

Convergent validity. Before a test can be made of hypotheses relating attitude to behavior, it is necessary to establish the validity of the attitude measures. Although a full test of construct validity would require the use of the multitrait–multimethod matrix design (cf. Ostrom, 1969), this was not possible in the present study because the same set of multiple measures was not used on each of the hypothesized attitudinal components. However, because two measures of affect and three measures of beliefs were obtained, it is possible to ascertain the degree of convergent validity for the attitudinal measures.

Two rival models bear upon the issue of convergence in the present research. The first is termed the single-factor model and hypothesizes that the variance in individual responses (except for random error in measurement) is accounted for by one underlying attitudinal construct. Panel a of Figure 1 depicts the single-factor model, in which the five attitudinal scales are shown measuring one overall attitudinal construct, affective and cognitive (i.e., A_{a+c}). Random error in measurements is indicated by the *z*s. The single-factor model is the attitude model implicitly employed by Fishbein and Ajzen (1974), in that they did not differentiate the five scales into affective and cognitive dimensions but rather treated all five as independent measures of the same underlying "attitude towards religion" construct. Panel a of Figure 1 represents a diagram of a hypothesis that can be used to test Fishbein and Ajzen's (1974) assumption.

The two-factor model (Figure 1, Panel b) posits that attitudes are represented as two conceptually independent, yet empirically related, constructs: (a) an affective component (A_a) and (b) a cognitive or belief dimension (A_c). The empirical correlation between A_a and A_c is shown as $\phi_{A_aA_c}$.

As a point of interpretation, it should be noted that the path diagrams in Figure 1 follow current practice in the confirmatory factor analysis literature in that they represent null hypotheses to be tested on data (cf. Jöreskog, 1969; Kenny, 1976). Although the application of confirmatory factor analysis does not permit one to discriminate *between* the single- and two-factor models per se, it does allow the researcher the opportunity of examining both the single-factor model as a null hypothesis and the two-factor model as a null hypothesis. The confirmatory factor analysis in the present context permits the researcher to test whether the one-factor, the two-factor, or both the one-factor and two-factor models are consistent with the data. In this way, the procedures yield an empirical test of the convergent validity for each attitude model and can be used to identify internally consistent constructs. The test is important to apply before investigation of the relationship between attitude and behavior because it is a matter of logical and empirical necessity that convergent validity be achieved in order for one to accept the theoretical and predictive utility of the attitude model. If one or both of the convergent validity hypotheses were rejected, then it would not be legitimate to examine further hypotheses as to the predictive validity of the nonconvergent attitudinal construct. In the present investigation, it is suggested that the two-component, but not the single-component, model will achieve convergent validity.

To test the models of Figure 1, the following general confirmatory factor analysis model can be examined:

$$y = \Lambda x + z, \tag{1}$$

where y is the vector of p attitude measures; x is a $k < p$ vector of hypothesized attitude components; k is the number of attitude components; z is a vector of p unique scores (i.e., errors in measurement); and Λ is a $p \times k$ matrix of factor loadings relating attitude measurements to their respective components. For Figure 1, Panel a, $p = 5$ and $k = 1$, while for Figure 1, Panel b, $p = 5$ and $k = 2$. With the assumptions that $E(x) = E(z) = 0$ (i.e., measures are expressed as deviations from their means), $E(xx') = \phi$ (where ϕ is the intercorrelation between attitude components), and $E(zz') = \psi$ (where ψ is a diagonal matrix of error variances), the variance-covariance matrix of measurements, Σ, may be expressed as

$$\Sigma = \Lambda\phi\Lambda' + \psi. \tag{2}$$

Jöreskog (1969) derives a maximum likelihood procedure for estimating the parameters in Λ, ϕ, and ψ. Further, the methodology yields an overall chi-square goodness-of-fit measure that tests the null hypothesis specified by Figure 1 and Equations 1 and 2 against the alternative hypothesis that the variance–covariance matrix is any positive definite matrix. The computer program LISREL can be used to test the models of Figure 1 and Equations 1 and 2 (cf. Jöreskog & van Thillo, 1972).

Single-act and multiple-act criteria. To test the hypotheses that attitude predicts single-act and multiple-act behavioral criteria, the path model of Figure 2 was used. In this figure, A represents the attitude construct achieving convergent validity; B stands for the appropriate single-act or multiple-act criterion measures; and n attitude measures (i.e., y_is) are drawn to indicate that the test applies to the general case. The single-act behavioral criterion measures are simply the individual responses to each of the 100 self-reports of behaviors or behavioral intention. The multiple-act behavioral criterion measure is the sum over 100 self-reports of behaviors or behavioral intentions.

In the present study, it is hypothesized that the attitude construct will not act as a significant predictor of either single-act or multiple-act behavioral

criteria (see the earlier discussion on attitude–behavior hypotheses). Looking first at the relation between attitude and the single-act behavioral criterion, it is expected that the relationship will be weak or nonsignificant because the degree of correspondence between attitude and behavior is minimal (see earlier comments and comments in Discussion section). This hypothesis is the same as that posed by Fishbein and Ajzen (1974).

Looking next at the relation between attitude and the multiple-act behavioral criterion, however, it is hypothesized that the overall predictive model of Figure 2 will be rejected despite the presence of high absolute correlations between each individual attitude measure and the multiple-act criterion measure. This perhaps counterintuitive hypothesis is contrary to that suggested by Fishbein and Ajzen (1974). The rationale for this hypothesis is that the high observed correlations between each individual attitude measure and the multiple-act criterion arise as a consequence of the definition and operationalization of the multiple-act criterion. Because this criterion is defined as the sum of occurrences over 100 different behaviors, it is probable that each respondent's score expresses a different set of behaviors, and thus the numerical value for the multiple-act criterion assigned to each person would represent an inconsistent mapping in measurement. Two or more people can have identical scores on the multiple-act criterion yet exhibit vastly different behaviors as a function of the content of their responses. As a result, the procedure used to construct the criterion violates the rule of numerical assignment theory that requires a 1:1 correspondence between number and referent. Even though one expects high correlations between individual measures of attitude and the criterion, it can be anticipated that the entire pattern of correlations among attitude measures and the criterion would not be consistent with the model hypothesized in Figure 2.

To test the model of Figure 2, Jöreskog's (1970) analysis of covariance structures methodology can be used in conjunction with LISREL. The appropriate model to test is

$$B = \gamma A + \zeta \tag{3}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \\ y_b \end{bmatrix} = \begin{bmatrix} \alpha_1 & 0 \\ \alpha_2 & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \\ \alpha_n & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} + \begin{bmatrix} z_1 \\ z_2 \\ \cdot \\ \cdot \\ \cdot \\ z_n \\ 0 \end{bmatrix}, \tag{4}$$

where B is the appropriate single-act or multiple-act behavioral criterion; A is the attitude component achieving convergent validity in the previously described tests; γ is a parameter analogous to a regression coefficient relating attitude to behavior; ζ is a random error in equation term; $y_1, y_2, \ldots, y_n$ are n attitude measurements; y_b is the actual behavioral criterion measure; the αs are parameters relating A to its measures; and the zs are random measurement errors.

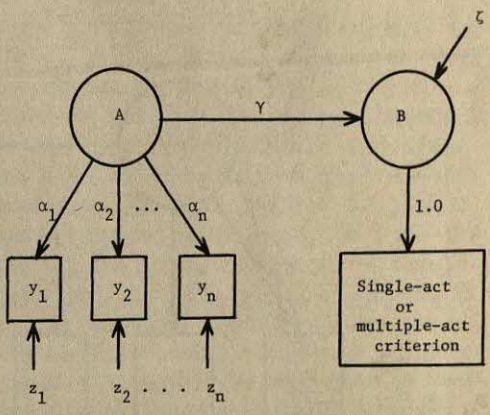


Figure 2. Path diagram for testing validity of single-act and multiple-act criteria models of behavior. (A = attitude construct; B = behavior criterion. The remaining symbols represent parameters to be estimated and are defined in the text.)

Prediction of scaled behavior. To test the hypothesis that attitude predicts scaled behavioral measures, the path diagram of Figure 3 may be used. In contrast to the single-act and multiple-act behavioral criteria, the scaled behaviors are meaningful measures and approach the same level of specificity as the attitudinal scales. For a description of the scaling procedures used, the reader is referred to Fishbein and Ajzen (1974). Although Fishbein and Ajzen hypothesized that attitude would predict scaled behavioral measures, they tested the hypothesis only by examining the individual correlations between each attitude measure and each scaled behavior measure. A more rigorous and meaningful test of the hypothesis can be performed by using the general analysis of covariance structures model, which is a generalization of confirmatory factor analysis. This method examines the entire pattern of intercorrelations as a unit rather than examining only the individual relationships. Further, it provides a chi-square goodness-of-fit test, estimates parameters, and takes into account differential reliability, as noted earlier. The appropriate equations to test are

$$B_s = \gamma A + \zeta \tag{5}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \\ y_{n+1} \\ \\ y_{n+2} \\ \cdot \\ \cdot \\ \cdot \\ \\ y_{n+m} \end{bmatrix} = \begin{bmatrix} \alpha_1 & 0 \\ \alpha_2 & 0 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \alpha_n & 0 \\ 0 & \lambda_1 \\ \\ \cdot & \\ \cdot & \lambda_2 \\ \cdot & \cdot \\ & \cdot \\ 0 & \lambda_m \end{bmatrix} \begin{bmatrix} A \\ B_s \end{bmatrix} + \begin{bmatrix} z_1 \\ z_2 \\ \cdot \\ \cdot \\ \cdot \\ z_n \\ \epsilon_1 \\ \epsilon_2 \\ \cdot \\ \cdot \\ \cdot \\ \epsilon_m \end{bmatrix}, \tag{6}$$

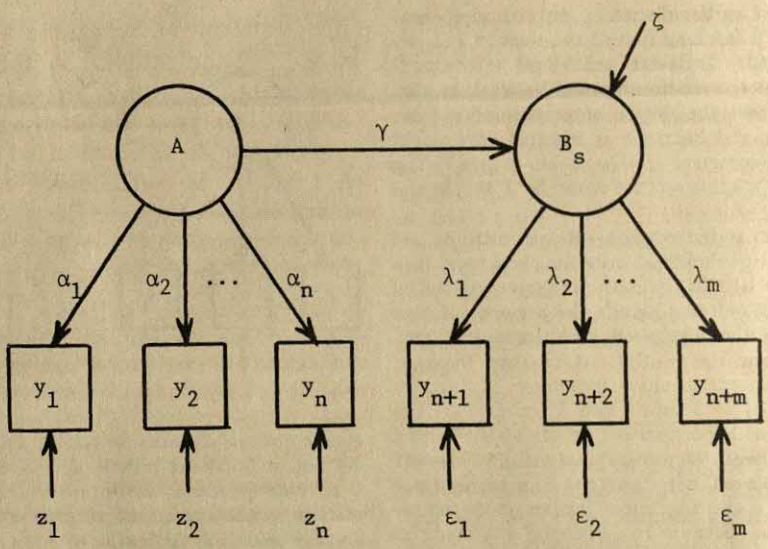


Figure 3. Path diagram for determining the relationship between attitudes and scaled measures of behavior. (A = attitude construct; B_s = scaled multiple-act behavior criterion. The remaining symbols represent parameters to be estimated and are defined in the text.)

where B_s is the scaled behavioral construct; A is the attitude component; $y_1, y_2, \ldots, y_n$ are the n measures of A: $y_{n+1}, y_{n+2}, \ldots, y_{m+n}$ are the m measures of B_s; the αs and λs are parameters relating A and B_s to their respective measures; the z and ϵ are random measurement errors; and γ and ζ are defined as above.

Results

Table 1 summarizes the intercorrelations of variables for both samples. Looking at these correlations, Fishbein and Ajzen (1974) concluded that the five verbal attitude scales showed "a high degree of convergent validity" and "all attitude scales correlated highly with the multiple-act criterion, while the prediction of single-act criteria tended to be low and nonsignificant" (p. 62). The latter was taken as evidence that a person's attitude towards an object should be related to the overall pattern of one's behaviors but not necessarily to

Table 1
Intercorrelations of Verbal Attitude Scales, Single-Act and Multiple-Act Behavioral Criteria, and Behavioral Scales[a]

Variable	1	2	3	4	5	6	7	8	9	10
1. Self-report	—	.800	.519	.652	.584	.137	.640	.451	.582	.701
2. Semantic differential	.765[b]	—	.644	.762	.685	.149	.714	.591	.688	.727
3. Guttman	.688	.773	—	.790	.744	.121	.608	.531	.647	.570
4. Likert	.743	.837	.878	—	.785	.142	.684	.660	.656	.611
5. Thurstone	.672	.666	.818	.786	—	.131	.628	.575	.624	.542
6. Single-act criterion[c]	.162	.178	.176	.202	.170	—	—	.143	.178	.156
7. Multiple-act criterion[d]	.604	.658	.656	.749	.648	—	—	.699	.898	.789
8. Guttman behavior	.444	.517	.501	.563	.483	.189	.750	—	.776	.631
9. Likert behavior	.594	.640	.656	.727	.649	.229	.915	.851	—	.792
10. Thurstone behavior	.545	.566	.663	.677	.628	.202	.797	.553	.793	—

[a] Above diagonal for respondents indicating self-reported behaviors (n = 62); below diagonal for respondents indicating behavioral intentions (n = 63).
[b] $r_{.05}$ = .250; $r_{.01}$ = .325.
[c] Mean correlation with 100 behaviors or intentions.
[d] Sum over 100 behaviors or intentions.

any single behavior that may be performed with respect to the object. Each of these hypotheses and the ones relating attitude to scaled behavior measures are examined below using the more rigorous confirmatory factor analysis methodology. An extension of the attitude model is also investigated.

Convergent Validity

Single-factor model. Application of the confirmatory factor analysis model of Figure 1, Panel a and LISREL to the data shows that the hypothesis of convergent validity must be rejected for both the self-reported behaviors sample, $\chi^2(5) = 23.91$, $p < .01$, and the behavioral intentions sample, $\chi^2(5) = 13.98$, $p < .01$. Thus, contrary to the original claim made by Fishbein and Ajzen (1974), one cannot accept the hypothesis that the five attitudinal scales are measures of a single underlying attitude construct.

Two-factor model. Given the failure to achieve convergent validity for the single-factor model, the two-factor, two-component model was tested (Figure 1, Panel b). Application of LISREL to the data indicates that the hypothesis of convergent validity is supported for the self-reported behaviors sample, $\chi^2(4) = 2.30$, $p \simeq .68$, and the behavioral intentions sample, $\chi^2(4) = 8.30$, $p \simeq .08$.[4] Hence, convergent validity is established for the two-component model but not for the single-component model. Further analysis of the predictive validity of attitude in the present study is thus justified only for the two-component conceptualization of attitude.

Prediction of Behavior: Single-Act and Multiple-Act Criteria

Single-act criterion. As posited, the hypothesis maintaining that attitude will be a significant predictor of single-act criterion measures was not supported. Looking first at the self-reported behaviors sample, it can be seen that—although the pattern of correlations among attitude and behavioral criterion measures was consistent with predictions using either affect, $\chi^2(1) = .02$, $p \simeq .89$, or cognitions, $\chi^2(3) = .05$, $p \simeq .99$, as predictors[5] —the standardized parameter estimates for the γs, relating attitude to behavior, were insignificant, $\gamma_a = .157$, *ns*; $\gamma_c = .148$, *ns*. Similarly, looking at the behavioral intentions sample, it can be observed that—despite finding a pattern of correlations consistent with predictions for affect, $\chi^2(1) = .04$, $p \simeq .83$, and cognitions, $\chi^2(3) = .34$, $p \simeq .95$—the standardized parameter estimates, γs, were insignificant, $\gamma_a = .190$, *ns*; $\gamma_c = .193$, *ns*. In sum, it was found that attitude does not act as a significant predictor of single-act criterion measures. These results are typical of that found in the literature relating attitude to single-act behaviors (cf. Wicker, 1969).

Multiple-act criterion. Although Table 1 indicates that high correlations exist between each of the five measures of attitude and the multiple-act behavior criterion, this does not necessarily imply that the attitude construct itself will be related validly to the multiple-act behavior criterion. The prediction of behavior indicated by Figure 2 requires that the entire pattern of correlations among all measurements meet a set of logical and empirical conditions. Not only must the correlations among

---

[4] Although the chi-square value for the behavioral intentions sample indicates a borderline fit (i.e., $p \simeq .08$), when Bartlett's (1951) small-sample correction factor is applied, the model reaches acceptable levels of significance, $\chi^2(4) = 7.65$, $p \simeq .10$.

[5] To test the prediction of behavior hypotheses, each attitude component was examined as a separate predictor rather than observing both simultaneously. This was done in order to avoid what Burt (1976) terms "interpretational confounding." Interpretational confounding is similar in form to the problem of multicollinearity in multiple regression analysis, and the condition can affect the precision of the parameter estimates relating attitude to behavior. Briefly, the problem arises because of the structure of intercorrelations among attitude and behavior measures. Logical considerations dictate that measures of the same construct should correlate highly among themselves and that most, if not all, correlations between measures of different constructs should be lower than the within-construct correlations (Campbell & Fiske, 1959). However, as shown in Table 1, many of the cross-construct correlations that include the criterion measure violate this rule, even though the criterion correlates highly with all measures of attitudes in an absolute sense. As a result, to examine the validity of attitude as a predictor of behavior, it is more meaningful to examine the predictive ability of the affective and cognitive components separately.

attitude measures be significant and achieve convergence, but simultaneously the composition of correlations between each attitude measure and the criterion must be consistent with the structure implied by Figure 2. It is hypothesized, however, that the construction of the multiple-act criterion will not sustain the prediction of behavior because the content of the criterion measure fails to achieve construct validity and is not at the same level of specificity as the attitude construct.

Application of Jöreskog's analysis of covariance structures to the data and the model of Figure 2 shows that the hypothesis referring to the prediction of the multiple-act criterion must be rejected, as maintained. Looking first at the self-reported behaviors sample, the hypothesis is rejected for both the model using affect as a predictor, $\chi^2(1) = 14.05$, $p < .01$, and the model using cognitions as a predictor, $\chi^2(3) = 12.89$, $p < .01$. Similarly, for the behavioral intentions sample, the hypothesis must be rejected for both the model using affect as a predictor, $\chi^2(1) = 10.93$, $p < .01$, and the model using cognitions as a predictor, $\chi^2(3) = 21.41$, $p < .01$.

Prediction of Behavior: Scaled Behavior Measures

As posited, the hypothesis that attitude will predict scaled behavior was supported (see Figure 3). Looking first at the self-reported behaviors sample, it can be seen that the hypothesis [6] is supported for (a) the model with affect as a predictor and either Guttman and Likert scaled behavior as a criterion, $\chi^2(1) = 1.01$, $p \simeq .32$, or Likert and Thurstone scaled behavior as a criterion, $\chi^2(1) = 2.77$, $p \simeq .10$, and (b) the model with cognitions as a predictor and either Guttman and Likert scaled behavior as a criterion, $\chi^2(4) = 5.38$, $p \simeq .25$, or Likert and Thurstone scaled behavior as a criterion, $\chi^2(4) = .85$, $p \simeq .93$. Similarly, for the behavioral intentions sample, the hypothesis is supported for (a) the model with affect as a predictor and either Guttman and Likert scaled behavior as a criterion, $\chi^2(1) = 1.38$, $p \simeq .24$, or Likert and Thurstone scaled behavior as a criterion, $\chi^2(1) = .14$, $p \simeq .71$, and (b) the model with cognitions as a predictor and either Guttman and Likert scaled behavior as a criterion, $\chi^2(4) = 6.35$, $p \simeq .17$, or Likert and Thurstone scaled behavior as a criterion, $\chi^2(4) = 4.79$, $p \simeq .31$.

To examine the joint effect of affect and cognitions on scaled behavior, the path model of Figure 4 can be used.[7] The structural equations for this model can be written as

$$B_s = (\gamma_1 \gamma_2) \begin{bmatrix} A_a \\ A_c \end{bmatrix} + \zeta \tag{7}$$

$$\begin{bmatrix} L_B \\ T_B \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} B_s + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \end{bmatrix} \tag{8}$$

$$\begin{bmatrix} SR \\ SD \\ G \\ L \\ T \end{bmatrix} = \begin{bmatrix} \alpha_1 & 0 \\ \alpha_2 & 0 \\ 0 & \alpha_3 \\ 0 & \alpha_4 \\ 0 & \alpha_5 \end{bmatrix} \begin{bmatrix} A_a \\ A_c \end{bmatrix} + \begin{bmatrix} z_1 \\ z_2 \\ z_3 \\ z_4 \\ z_5 \end{bmatrix}, \tag{9}$$

where B_s is the scaled behavior construct; A_a and A_c are the affective and cognitive components of attitude, respectively; L_B and T_B

---

[6] Predictive validity models are run with the Guttman/Likert and Likert/Thurstone operationalizations rather than with a single Guttman/Likert/Thurstone operationalization of behavior because inspection of the correlation matrix in Table 1 reveals that the Guttman and Thurstone measures correlate at a lower level than the other two pairs of correlations in both samples. As Burt (1976) notes, such a pattern of intercorrelations among the measures of the same construct can result in "interpretational confounding," pariticularly if each measure correlates at a different degree with a criterion measure. The details of this confounding are quite mathematically complex, and for a complete description of the problem, the reader is referred to Burt (1976). At an intuitive level, what interpretational confounding means in the present context is that the covariances of the measures of behavior with the attitude measures are not proportional but reflect some unmeasured influence such as excessive measurement error. The condition is present only when all three measures of behavior are present or when the Guttman and Thurstone scales are used. The condition is not present when the Guttman/Likert and Likert/Thurstone combinations are employed. Hence, the predictive validity models were examined with the latter two measures of behavior.

[7] The joint effect of affect and cognitions on scaled behavior should be interpreted with caution. The procedure should be employed only when interpretational confounding is not a serious problem (Burt, 1976). See Footnote 5.

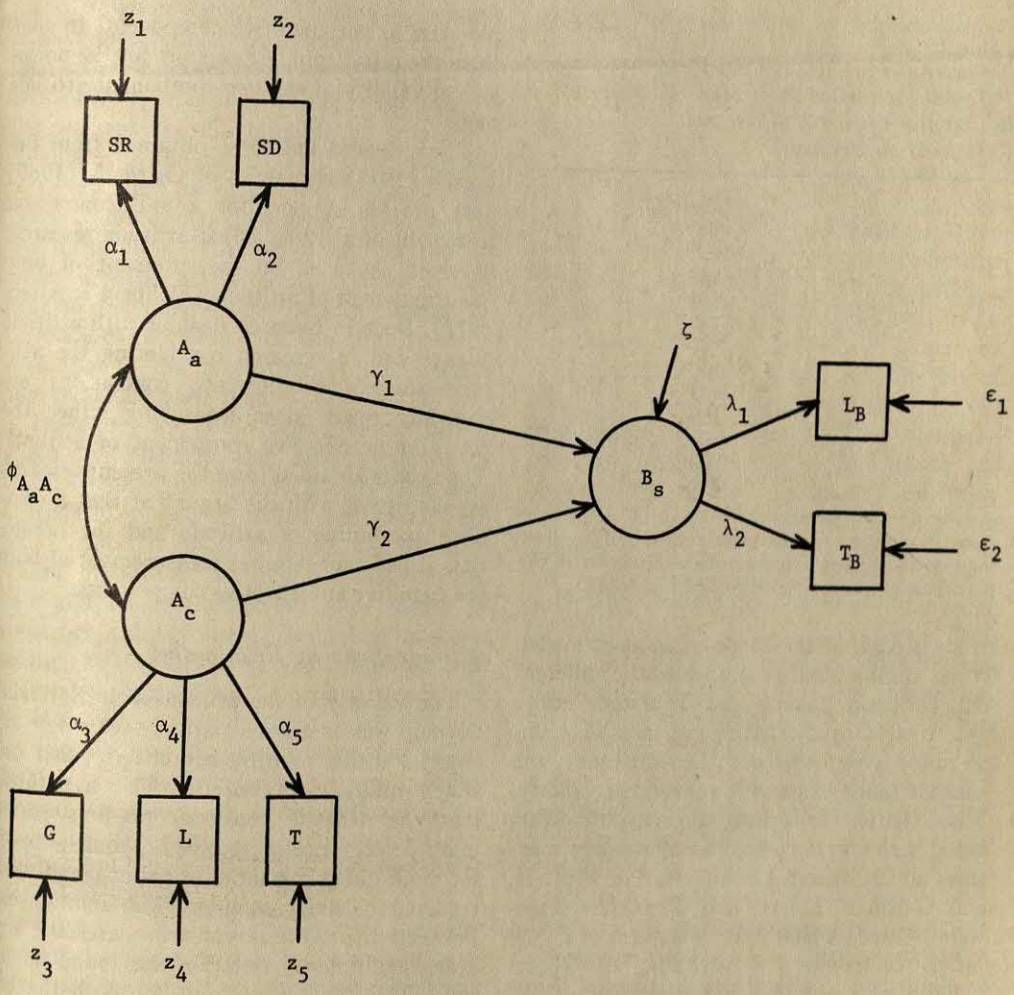


Figure 4. Testing for the relative impact of affect and cognitions on behavior. (A_a = affective attitude component; A_c = cognitive attitude component; B_s = scaled multiple-act behavior criterion; L_B = Likert scaled behavior measure; T_B = Thurstone scaled behavior measure; SR = Guilford self-report, SD = semantic differential, G = Guttman, L = Likert, and T = Thurstone scales. The remaining symbols represent parameters to be estimated and are defined in the text.)

are the Likert and Thurstone scaled behavior measures, respectively; SR = Guilford self-report, SD = semantic differential, G = Guttman, L = Likert, T = Thurstone scales; and the remaining symbols are as defined earlier.

Application of Jöreskog's analysis of covariance structures to the data and model of Figure 4 yields the results shown in Table 2 where, because of problems of interpretational confounding, only the findings for the self-reported behaviors sample are displayed. In general, the model fits the data well, $\chi^2(10) = 13.11$, $p \simeq .22$. Further, the affective component can be seen to be approximately three times as forceful in its impact on behavior as the cognitive component ($\gamma_1 = .651$ and $\gamma_2 = .226$).

Discussion

Attitude Organization

The results obtained from a reanalysis of the Fishbein and Ajzen (1974) data support a two-component model of attitude and lead

Table 2
Goodness-of-Fit Tests and Standardized Parameter Estimates for a Model Testing the Relative Impact of Affect and Cognitions on Behavior

Parameter	Self-reported behavior
α_1	.838
α_2	.955
α_3	.849
α_4	.930
α_5	.853
γ_1	.651
γ_2	.226
λ_1	.879
λ_2	.901
$\phi_{A_1A_2}$	.834

Note. Parameters in this table correspond to those indicated in Figure 4 and are defined further in the text. Test of model: $\chi^2(10) = 13.11$, $p \simeq .22$.

to a rejection of the single-component model. When Guilford self-rating, semantic differential, Guttman, Likert, and Thurstone scales were treated as alternative measures of a single underlying construct (i.e., attitude), the analysis failed to provide convergent validity. When Guilford self-rating and semantic differential scales were treated as alternative measures of an affective component of attitude, and Guttman, Likert, and Thurstone scales were treated as alternative measures of a cognitive component of attitude, convergent validity was obtained. The validity of the two-component model was supported for the self-report and behavioral intention samples.

Further support for the validity of the two-component model is provided by a consideration of nomological validity. Following Campbell (1960), nomological validity is defined as the degree to which predictions from a concept in a theoretical system of concepts are confirmed. It was found that each component of the two-component model separately predicted scaled multiple-act criteria. Furthermore, both components contributed simultaneously to the prediction of behavior. Because of interpretational confounding, only self-report data were employed in the latter analysis. While both components simultaneously accounted for behavior, the affective component was approximately three times as powerful as the cognitive component. In sum, then, the results provide support for the nomological validity of the two-component attitude model.

These results and those obtained from Bagozzi's (1978) reanalysis of Ostrom's (1969) data provide support for a multicomponent treatment of attitude. Most attitude research, however, is based on measurement of only one component of attitude (Fishbein & Ajzen, 1975). Both in research dealing with attitude change and in research considering the attitude–behavior relationship, conclusions are generally made after measuring either the cognitive or affective component of attitude. The results obtained from the present research are consistent with the argument that a complete accounting of attitude and the prediction of behavior requires measurement of both the cognitive and affective components.

Attitude–Behavior Relationship

The validity of the attitude–behavior relationship was assessed separately for single-act criteria, scaled multiple-act criteria, and unscaled multiple-act criteria. The hypothesis that the attitude construct predicts scaled multiple-act criteria received strong support for both the behavioral intentions and self-reported behavior samples. The relationship between the attitude construct and the unscaled multiple-act criterion was found not to hold. This finding was true for both the behavioral intention and self-reported behavior samples. Finally, as predicted, the relationship between the attitude construct and single-act criteria was found not to hold in either sample.

These results provide partial support for Ajzen and Fishbein's (1977) contention that the strength of the attitude–behavior relationship is directly related to the degree of correspondence between "attitudinal and behavior entities." According to Ajzen and Fishbein's conceptualization, however, correspondence should also be high for the multiple-act criterion derived from summation across all 100 behaviors, because the action, context, and time elements are left unspecified in both the attitudinal and behavioral entities. Even though correspondence was high in this sense,

the summation across 100 behaviors is likely to have incorporated considerable error in the final criterion score. In an earlier paper, Fishbein (1973) notes that scaling is necessary in the construction of a multiple-act criterion to eliminate invalid items (i.e., items unrelated to favorability for or against the concept under consideration). The error introduced by the inclusion of "invalid items" in the unscaled multiple-act criterion is likely to have lead to an inconsistent pattern among the correlations between attitude and behavior even though the magnitude of the relationship remained quite high.

Therefore, we would argue, in contrast to Ajzen and Fishbein, that correspondence between attitude and behavior is only high when attitude is related to scaled multiple-act criteria. The attitudinal construct was based on scaling procedures that ensured that only beliefs and feelings achieving some degree of construct validity were used. The unscaled multiple-act criterion was not based on such a restrictive procedure. Correspondence will be high only when the behavioral entity is also based on scaling procedures that ensure that the criterion achieves convergent validity and the content of the attitudinal and behavioral items is at comparable levels of generality. This was found only for the scaled multiple-act criteria.

Failure to sustain the validity of the relationship between attitude and single-act criteria provides further support to Fishbein and Ajzen's argument that a general attitude bears no necessary relationship to single-act criteria. This does not alter the frequently obtained finding of significant relationships between attitude and single acts (i.e., Campbell et al., 1960; Fishbein & Ajzen, 1974). It does point out, however, that these relationships cannot be automatically assumed. The question remains as to how we can determine which single acts can be expected to bear a valid relationship to attitude.

Failure to obtain direct relationships between general attitude and single acts may be due to the fact that frequently little a priori thought is given to why a relationship should logically be expected between attitude and action in a given context and time. Not only should a rationale exist for hypothesizing a specific link between attitude and single acts, but other nonattitudinal factors should also be considered. Normative prescriptions and social constraints have frequently been identified as factors that moderate the strength of the attitude–behavior relationship (e.g., Gross & Niman, 1975; Kelman, 1974; Liska, 1974; Wicker, 1969).[8] We may suggest that the attitude–behavior relationship will be strong to the extent that the behavior does not involve inconsistent social or normative constraints. Warner and DeFleur (1969) provide support for the contention that the attitude–behavior relationship will be stronger when normative pressures are consistent with attitude and weaker when normative pressures are inconsistent with attitude. Similar results have been obtained by Schofield (1975). In contrast to such public behaviors as appearing in a photograph with a member of another racial group (cf. Linn, 1965), voting behavior takes place in the privacy of the voting booth. Strong relationships have been obtained between attitude toward a candidate and the act of voting for that candidate (e.g., Campbell et al., 1960). It may be, then, that those behaviors most directly related to attitude toward religion are behaviors toward religion that are either private or in other respects unlikely to be subject to inconsistent normative pressures. While it is beyond the purpose of this article to develop and test the implications of a priori criteria that will affect predictions of whether a behavior will be related to attitude, it is suggested that continued consideration of the meaning of single acts in conjunction with these and other moderating variables would further our understanding of the attitude–behavior relationship.

Further study of behaviors unrelated to attitude would also be desirable. Most behaviors employed in the Fishbein and Ajzen (1974) study were rejected by all three scal-

[8] It has also been maintained that other attitudes and motives are important mediators of the attitude–behavior relationship and that the behaviors must be perceived to be relevant to the attitude (i.e., Gross & Niman, 1975; Insko & Schopler, 1967; Kelman, 1974; Liska, 1974; Wicker, 1969).

ing procedures and were essentially unrelated to attitude. Nevertheless, performance of these behaviors may be strongly related to other psychologically relevant dimensions such as motivation, attribution processes, or personality variables. If we are to enhance our understanding of and ability to predict behavior, then it would seem desirable to direct further attention to psychological dimensions (other than attitude) accounting for behavior.

Conclusion

This study was directed at further clarifying attitude organization and attitude–behavior issues. Evidence was obtained to support an affective–cognitive conceptualization of attitude. Convergent validity was obtained for this two-component model, and it was found that attitude accounted for self-reported behavior and for behavioral intentions. It was suggested that further research with attitude should include measures of both components and treat them as simultaneous predictors of behavioral criteria.

Strong support was found for the hypothesis that attitude would predict scaled multiple-act criteria. Criteria based on summation over a large set of behaviors were found not to be related systematically to the attitude construct. Furthermore, the hypothesis that attitude would predict single-act criteria was not supported. These findings provide partial support for Ajzen and Fishbein's concept of attitude–behavior correspondence and lead to further specification of this concept. It was suggested that correspondence (in the sense of comparable levels of specificity in context, time, and action between attitude and behavior) is high, and hence a strong empirical attitude–behavior relationship can be expected, only when both attitudinal and behavioral entities are appropriately scaled.

The failure to support a valid general attitude – single-act criteria relationship—even though some single acts were directly related to attitude—led to a consideration of factors likely to result in strong behavior relationships. It was speculated that a strong relationship would be forthcoming when the target elements of the attitude and behavior correspond and when the behavior under consideration is relatively devoid of normative constraints.

References

Ajzen, I., & Fishbein, M. Attitude–behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, 1977, *84*, 888–918.

Bagozzi, R. P. The construct validity of the affective, behavioral and cognitive components of attitude by analysis of covariance structures. *Multivariate Behavioral Research*, 1978, *13*, 9–31.

Bardis, P. D. A religion scale. *Social Science*, 1961, *36*, 120–123.

Bartlett, M. S. The effect of standardization on an approximation in factor analysis. *Biometrika*, 1951, *38*, 337–344.

Berg, K. E. Ethnic attitudes and agreement with a Negro person. *Journal of Personality and Social Psychology*, 1966, *4*, 215–220.

Bray, D. W. The prediction of behavior from two attitude scales. *Journal of Abnormal and Social Psychology*, 1950, *45*, 64–84.

Burt, R. S. Interpretational confounding of unobserved variables in structural equation models. *Sociological Methods and Research*, 1976, *5*, 3–52.

Calder, B. J., & Ross, M. *Attitudes and behavior*. Morristown, N.J.: General Learning Press, 1973.

Campbell, A., Converse, P. E., Miller, W. E., & Stokes, D. E. *The American voter*. New York: Wiley, 1960.

Campbell, D. T., & Fiske, D. W. Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin*, 1959, *56*, 81–105.

Davidson, A. R., & Jaccard, J. J. Population psychology: A new look at an old problem. *Journal of Personality and Social Psychology*, 1975, *6*, 1073–1082.

DeFleur, M. L., & Westie, F. R. Verbal attitudes and overt acts: An experiment on the salience of attitudes. *American Sociological Review*, 1958, *23*, 667–673.

Faulkner, J. W., & DeJong, G. F. Religiosity in 5-D: An empirical analysis. In J. P. Robinson & P. R. Shaver (Eds.), *Measures of social psychological attitudes*. Ann Arbor: University of Michigan, Institute of Social Research, 1969.

Fishbein, M. Attitude and the prediction of behavior. In M. Fishbein (Ed.), *Readings in attitude theory and measurement*. New York: Wiley, 1967. (a)

Fishbein, M. A consideration of beliefs and their role in attitude measurement. In M. Fishbein (Ed.), *Readings in attitude theory and measurement*. New York: Wiley, 1967. (b)

Fishbein, M. The prediction of behaviors from attitudinal variables. In C. D. Mortensen & K. K. Sereno (Eds.), *Advances in communication research*. New York: Harper & Row, 1973.

Fishbein, M., & Ajzen, I. Attitudes towards objects as predictors of single and multiple behavioral criteria. *Psychological Review*, 1974, *81*, 59–74.

Fishbein, M., & Ajzen, I. *Belief, attitude intention*

and behavior: An introduction to theory and research. Reading, Mass.: Addison-Wesley, 1975.

Fishbein, M., & Raven, B. H. The AB scales: An operational definition of belief and attitude. *Human Relations*, 1962, *15*, 35–44.

Greenwald, A. G. On defining attitude and attitude theory. In A. G. Greenwald, T. C. Brock, & T. M. Ostrom (Eds.), *Psychological foundations of attitudes*. New York: Academic Press, 1968.

Gross, S. J., & Niman, C. M. Attitude–behavior consistency: A review. *Public Opinion Quarterly*, 1975, *39*, 358–368.

Insko, C. A., & Schopler, J. Triadic consistency: A statement of affective-cognitive-conative consistency. *Psychological Review*, 1967, *74*, 361–376.

Jöreskog, K. G. A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 1969, *34*, 183–202.

Jöreskog, K. G. A general method for analysis of covariance structures. *Biometrika*, 1970, *57*, 239–251.

Jöreskog, K. G., & van Thillo, M. LISREL: *A general computer program for estimating a linear structural equation system involving multiple indicators of unmeasured variables*. Princeton, N.J.: Educational Testing Service, 1972.

Katz, D., & Stotland, E. A preliminary statement to a theory of attitude structure and change. In S. Koch (Ed.), *Psychology: A study of a science* (Vol. 3). New York: McGraw-Hill, 1959.

Kelman, H. C. Attitudes are alive and well and gainfully employed in the sphere of action. *American Psychologist*, 1974, *29*, 310–324.

Kenny, D. A. An empirical application of confirmatory factor analysis to the multitrait–multimethod matrix. *Journal of Experimental Social Psychology*, 1976, *12*, 247–252.

Kothandapani, V. Validation of feeling, belief, and intention to act as three components of attitude and their contribution to prediction of contraceptive behavior. *Journal of Personality and Social Psychology*, 1971, *19*, 321–333.

Krech, D., Crutchfield, R. S., & Ballachey, E. L. *Individual in society*. New York: McGraw-Hill, 1962.

Kutner, B., Wilkins, C., & Yarrow, P. R. Verbal attitudes and overt behavior involving racial prejudice. *Journal of Abnormal and Social Psychology*, 1952, *47*, 649–652.

La Piere, R. T. Attitudes versus actions. *Social Forces*, 1934, *13*, 230–237.

Linn, L. S. Verbal attitudes and overt behavior: A study of racial discrimination. *Social Forces*, 1965, *43*, 353–364.

Liska, A. E. Emergent issues in the attitude–behavior consistency controversy. *American Sociological Review*, 1974, *39*, 261–272.

McGuire, W. J. The nature of attitudes and attitude change. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (Vol. 3). Reading, Mass.: Addison-Wesley, 1969.

Nemeth, C. Effects of free versus constrained behavior on attraction between people. *Journal of Personality and Social Psychology*, 1970, *15*, 302–311.

Norman, R. Affective–cognitive consistency, attitudes, conformity, and behavior. *Journal of Personality and Social Psychology*, 1975, *32*, 83–71.

Osgood, C. E., Suci, G. J., & Tannenbaum, P. H. *The measurement of meaning*. Urbana: University of Illinois Press, 1957.

Ostrom, T. M. The relationship between the affective, behavioral, and cognitive components of attitude. *Journal of Experimental Social Psychology*, 1969, *5*, 12–30.

Poppleton, P. K., & Pilkington, G. W. The measurement of religious attitudes in a university population. *British Journal of Social and Clinical Psychology*, 1963, *2*, 20–36.

Rogers, R. W., & Thistlethwaite, D. L. Effects of fear arousal and reassurance on attitude change. *Journal of Personality and Social Psychology*, 1970, *15*, 227–233.

Rosenberg, M. J. Hedonism, inauthenticity, and other goads toward expansion of a consistency theory. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A source-book*. Chicago: Rand McNally, 1968.

Rosenberg, M. J., & Hovland, C. I. Cognitive, affective and behavioral components of attitudes. In M. J. Rosenberg, C. I. Hovland, W. J. McGuire, R. P. Abelson, & J. W. Brehm (Eds.), *Attitude organization and change: An analysis of consistency among attitude components*. New Haven, Conn.: Yale University Press, 1960.

Sample, J., & Warland, R. Attitudes and prediction of behavior. *Social Forces*, 1973, *51*, 292–304.

Schofield, J. W. Effect of norms, public disclosure, and need for approval on volunteering behavior consistent with attitudes. *Journal of Personality and Social Psychology*, 1975, *31*, 1126–1133.

Smith, M. B. The personal setting of public opinions: A study of attitudes toward Russia. *Public Opinion Quarterly*, 1947, *11*, 507–523.

Tittle, C. R., & Hill, R. J. Attitude measurement and prediction of behavior: An evaluation of conditions and measurement techniques. *Sociometry*, 1967, *30*, 199–213.

Warner, L. G., & DeFleur, M. L. Attitude as an interactional concept: Social constraint and social distance as intervening variables between attitudes and action. *American Psychological Review*, 1969, *34*, 153–169.

Weigel, R. H., & Newman, L. S. Increasing attitude–behavior correspondence by broadening the scope of the behavioral measure. *Journal of Personality and Social Psychology*, 1976, *33*, 793–802.

Wicker, A. W. Attitudes versus actions: The relationship of verbal and overt behavioral responses to attitude objects. *Journal of Social Issues*, 1969, *25*, 41–78.

Woodmansee, J. J., & Cook, S. W. Dimensions of verbal racial attitudes: Their identification and measurement. *Journal of Personality and Social Psychology*, 1967, *7*, 240–250.

Received June 22, 1978 ■

Special Section on Interpretive Controversies in Personality and Social Psychology

Editor's Foreword

The convergence of 13 articles on five interpretive controversies in this issue of the *Journal of Personality and Social Psychology* has been a product of coincidence and necessity more than design. Several of the articles were ready for appearance in this issue in the normal course of publication scheduling. The others were made ready in time for this issue as a consequence of its being the last issue to appear under the present editorship. (Acting Editor Clyde Hendrick and a new editorial board assume responsibility starting with next month's issue.) The several authors, JPSP's Topic Editors, and the American Psychological Association's Central Office staff are due thanks for their cooperation in preparing several of the articles in time for appearance in this special section.

The Special Section was edited using a novel strategy for preparing and presenting controversial exchanges. In preparing the articles, parties to each exchange were asked to read each others' prepublication drafts in order to bring to the editor's attention misstatements of each others' positions. Insofar as was possible—and because of disagreements between authors, it was not always possible—apparent misstatements were modified or deleted prior to publication. It was hoped that this procedure would permit the exchange to be focused on the substance of the interpretive issues and would minimize debate over peripheral matters. In publishing the articles, a novel format was used to avoid an obvious last-word advantage for any author. The author (Author A) of the first-submitted article in an exchange was not invited to write a rebuttal article in response to an accepted reply (by Author B). Rather, Author A was instead given the opportunity to add a postscript to the originally accepted article. The major ground rule for such postscripts was that they could indicate responses to arguments presented in Author B's reply, but not in such a way as to bring in new issues that might legitimate further comment by Author B.

Experience with the articles in this special section suggests that the editorial strategy that was employed requires more effort by authors and editors than does the more usual article–reply–rejoinder format. This effort is justified only if readers find the exchanges prepared in this fashion to be especially useful. Readers are therefore requested to direct evaluative reactions to the Chief Editorial Advisor, American Psychological Association, 1200 Seventeenth Street N.W., Washington, D.C. 20036.—A.G.G.

Journal of Personality and Social Psychology
1979, Vol. 37, No. 6, 931-949

A Reply to Norman H. Anderson's Critique of the Subject–Verb–Object Approach to Social Cognition

Harry F. Gollob
University of Denver

The Subject–Verb–Object (S–V–O) approach to social cognition provides a method of conceptualizing several social cognition problems within the context of item sets that contain eight subject–verb–object sentences, such as "The kind man avoided the intelligent swindler" and "Conservative voters enjoy golf." Anderson recently criticized the S–V–O model on numerous grounds that were primarily technical in nature. The present reply refutes all of Anderson's major criticisms. In discussing an alternative to the S–V–O model, Anderson recommends that problems raised by the pervasive tendency of subjects to respond to information configurally be dealt with by collecting data that can be described by simple algebraic models. It is argued that Anderson's alternative avoids, rather than deals with, the important issues raised by the configural responding of subjects. Although the specific content of the present article is primarily concerned with the S–V–O model, much of the discussion is relevant to issues of general importance to investigators who use analysis of variance models and/or simple algebraic models in their work. A postscript replies to the article by Anderson that follows this one.

Anderson (1977) recently wrote a detailed critique of Gollob's (1974a, 1974b) Subject–Verb–Object (S–V–O) approach to social cognition as well as of Wyer's (1975) and Insko, Songer, and McGarvey's (1974) applications of that approach. Wyer and his co-workers (Wyer, 1974, 1975; Wyer, Henninger, & Wolfson, 1975) have extended Gollob's model and have applied it in a broad variety of areas concerned with social inference and attribution theory. Insko et al. applied the S–V–O model to the analysis of pleasantness judgments in a traditional balance-theory paradigm.

In his critique, Anderson claims that Gollob and his colleagues have used analysis of variance techniques inappropriately by failing to use tests of goodness of fit, by failing to recognize that monotonic transformations can affect results, by failing to realize that a complete analysis of variance decomposition can perfectly account for any set of data, by failing to realize that correlation coefficients can be misleading, by failing to recognize that people often respond configurally to information in sentences, and by committing various other foolish or careless errors. Anderson also suggests that all would have been fine if only Gollob and his colleagues had used Anderson's preferred methods to guide their data collection and analyses.

The present reply argues that all of Anderson's major criticisms of the S–V–O model are either wrong, irrelevant, or misleading. Since Anderson's criticisms focus primarily on

Work on this article was supported by National Science Foundation Grant SOC75-13303.

I wish to thank Robert S. Wyer, Jr., Robyn M. Dawes, Betty B. Rossman, and Robert M. McFatter for their extremely helpful comments on a draft on this article. Additionally, I thank Kenneth Purcell and Maureen Gollob for helpful discussions concerning the postscript.

Requests for reprints should be sent to Harry F. Gollob, Department of Psychology, University of Denver, Denver, Colorado 80208.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3706-0931$00.75

Gollob's work, this reply emphasizes comparisons of Gollob's (1974a, 1974b; Rossman & Gollob, 1976) claims and practices with those attributed to him by Anderson. Although the specific content of the reply is concerned with the S–V–O model, much of the discussion is relevant to issues of general importance to investigators who use analysis of variance models or other relatively simple mathematical models in their work.

Anderson's critique frequently ignores the necessity of considering the theoretical and psychological meaningfulness of a model in evaluating its usefulness. To enable meaningfulness to be given adequate consideration, the present article summarizes the major characteristics of the S–V–O model before giving detailed consideration to Anderson's criticisms. In typical applications of the S–V–O approach, considerations of judges' phenomenology play a crucial role, but since most of Anderson's criticisms are primarily technical in nature, the following description emphasizes technical aspects of the model.

The S–V–O Approach to Social Cognition

It should be noted at the outset that although the historical links between balance theory and the S–V–O approach to social cognition are strong, the similarity between the two systems is much weaker than Anderson implies by his decision to refer to the S–V–O model as the "balance triad model." The S–V–O approach differs from balance theory in several ways: The S–V–O approach considers six cognitive biases in addition to Heiderian balance and proposes a framework within which the joint effects of the hypothesized biases can be investigated; it deals with stimuli expressed as subject–verb–object sentences rather than as triads; it emphasizes that the relative importance of cognitive biases in an item set will vary greatly depending on the specific content of the item set and on the type of inference (or other judgment) being made; it focuses on the role of informational processes in affecting social judgments; and it makes several theoretical claims that are not part of traditional balance theory.

The S–V–O approach to social cognition provides a systematic method of investigating and conceptualizing several social cognition problems within the context of subject–verb–object sentences such as "Friendly Bill dislikes criminals" or "Intelligent men oppose antiobscenity laws." The substantive examples Anderson uses in his critique involve social inferences made on the basis of information that is highly evaluative in nature. Therefore, in order to facilitate discussion, I shall summarize only those aspects of the S–V–O approach that are relevant when dealing with item sets in which social inference judgments are obtained, and in which each of the sentence components is described either by a positively evaluated term (+1) or by a negatively evaluated term (−1).

Although Anderson considers only verb inferences, the S–V–O model distinguishes between three major types of social inference judgment: subject inferences, verb inferences, and object inferences. When making verb inferences a perceiver judges the probability that an actor with specified characteristics behaves or feels in a particular way toward a recipient who also has some specified characteristics—for example, "How likely is it that the kind man hates Mr. B who is a psychologist?" Examples of subject and object inferences, respectively, are "The man hates Mr. B who is a psychologist. How likely is it that the man is kind?" and "The kind man hates Mr. B. How likely is it that Mr. B is a psychologist?"

To apply the S–V–O approach, *item sets* are constructed in which two sentence subjects, two verbs, and two objects are factorially combined to yield the eight basic sentence types shown in Table 1. For example, the code "+ + −" represents the sentence type in which the subject and the verb are positively evaluated and the object is negatively evaluated. As shown in Table 1, each of the eight sentence types can be described according to the presence or absence of seven different cognitive biases—in the case of evaluative content there are three positivity biases and four balance biases. The presence or absence of each cognitive bias is easily determined. Let S′, V′, and O′, respectively, denote the sign of the sentence subject, verb, and object. Then, whenever S′, V′, or O′ is

positive, or the product S'V', S'O', V'O', or S'V'O' is positive, the specified sentence type is said to possess the corresponding cognitive bias. The names of the seven biases correspond to the names of the seven sources of variation one would have in a Subject × Verb × Object analysis of variance.

I shall now briefly discuss interpretations of the seven cognitive biases that often seem appropriate when dealing with evaluative content that is relevant to interpersonal relationships. S, V, and O positivity are cognitive biases that should be important in affecting inferences to the extent that judges tend to assume that, on the average, an actor, behavior, or object will have positively evaluated characteristics rather than negatively evaluated characteristics. It is hypothesized that SV balance will be important in affecting an inference to the extent that the perceiver expects the sentence subject and verb to have the same evaluative sign regardless of the characteristics of the recipient of the act (e.g., perceivers expect good people to do good things and bad people to do bad things). Similar lines of reasoning are used to relate SO balance and VO balance to relevant aspects of perceivers' phenomenology. SVO balance is simply traditional Heiderian balance. For more detailed discussion of the psychological significance of the biases in both evaluative and nonevaluative cases, see Gollob (1974b, pp. 288–292, 316–319).

In the S–V–O approach, each of the cognitive biases is thought of as providing a psychologically meaningful unit of information that is combined with other biases to yield a likelihood judgment. Formally, the most general version of the model assumes that the likelihood judgment, y_i, associated with the ith sentence type in an item set is given by the equation

$$y_i = \sum_{j=1}^{7} w_j B_{ij} + \text{constant}, \qquad (1)$$

where B_{ij} is equal to $+1$ or -1, depending on whether the jth cognitive bias is present or absent in the ith sentence in the item set, and w_j is the weight given to the jth cognitive bias for the item set (Gollob, 1974b, p. 292). The absolute value of the weight assigned to a bias reflects the importance of the bias in describing the inferences in the item set.

Table 1
Presence (+1) and Absence (−1) of Cognitive Biases in Each Sentence Type

Sentence type	Positivity			Balance			
	S	V	O	SV	SO	VO	SVO
+ + +	1	1	1	1	1	1	1
+ + −	1	1	−1	1	−1	−1	−1
+ − +	1	−1	1	−1	1	−1	−1
+ − −	1	−1	−1	−1	−1	1	1
− + +	−1	1	1	−1	−1	1	−1
− + −	−1	1	−1	−1	1	−1	1
− − +	−1	−1	1	1	−1	−1	1
− − −	−1	−1	−1	1	1	1	−1

Note. S = subject, V = verb, O = object.

When no restrictions are placed on the weights, Equation 1 is simply one way of writing the standard analysis of variance decomposition for a $2 \times 2 \times 2$ table. Significance tests of corresponding sources of variation in an analysis of variance are tests of the significance of the regression weights in Equation 1. In order to give the model some predictive power, some restrictions must be placed on the weights. The major restrictions discussed by Gollob (1974b, pp. 292–295) are summarized in the following section.

Weight Restrictions for the S–V–O Model

Nonnegative weight hypothesis. This hypothesis states that any cognitive bias useful in describing the judgment will be given a positive weight. The interpretations given above for the biases implicitly assumed that this hypothesis will hold when dealing with evaluative content.

Negatively accelerated weight hypothesis. This hypothesis states that when the absolute values of the weights in Equation 1 are put in rank order from largest to smallest, the size of the nonzero weights decreases at a slower and slower rate (Gollob, 1974b, p. 295; Rossman & Gollob, 1976, p. 377). Empirical results suggest that this negatively accelerated curve usually is fairly steep, so that typically only two or three cognitive biases

in a particular item set are given substantial weights (Gollob, 1974b; Rossman & Gollob, 1976; Wyer, 1974).

Relevant-bias hypothesis. This hypothesis states that the three cognitive biases that do not involve the sentence component to be inferred will be given zero weights. For example, in the case of verb inferences the hypothesis states that S positivity, O positivity, and SO balance will be given weights of zero, and that only V positivity and SV, VO, and SVO balance can be relevant in accounting for verb inferences. The relevant-bias hypothesis is expected to apply in most evaluative item sets in which each of the sentence components is described in terms of bipolar opposites (e.g., friendly–unfriendly, likes–dislikes, supports–opposes, liberals–conservatives). Use of such bipolar opposites, however, does not ensure that the relevant-bias hypothesis will hold (see Gollob, 1974b, pp. 294, 318–319).

Some intuitive feeling for the psychological meaningfulness of the relevant-bias hypothesis may be gained by considering an example. In making a verb inference, values along the verb dimension (e.g., helps–harms) often have implications for the judgment, both when they are considered alone and when they are considered in conjunction with other sentence components. More specifically, to the extent that a judge feels that it is generally true that people help rather than hurt others, V positivity should contribute to his or her inference; to the extent that he or she believes that the evaluative characteristics of the actor (i.e., the sentence subject) affect the likelihood of the actor helping another person, SV balance should contribute; to the extent that he or she believes that the evaluative characteristics of the recipient (i.e., the sentence object) affect the likelihood that the recipient will be helped, VO balance should contribute; finally, to the extent that the evaluative characteristics of both the sentence subject and object are jointly considered, SVO balance should contribute.

When the nonnegative weight hypothesis, the negatively accelerated weight hypothesis, and the relevant-bias hypothesis all hold, a different special case of the general S–V–O model is obtained for subject, verb, and object inferences. Gollob (1974b) has shown that each of these submodels implies that only 24 different orderings of the eight basic sentence types can be obtained in data. Since there are 40,320 possible orderings of the eight sentence types, it is clear that these special cases of the general model have considerable predictive power.

Other weight restrictions. In addition to the above hypotheses, which have rather broad applicability to social inference item sets, it is of course possible to use additional theoretical and intuitive considerations to make more specific a priori predictions concerning the relative importance of the biases. Such considerations depend on the content of the item sets, the experimental manipulations used, the characteristics of the judges, and the types of judgments obtained. For example, since the phenomenological basis for the relevant-bias hypothesis is applicable only to inference judgments, Rossman and Gollob (1976, pp. 379–380) did not use this hypothesis to make predictions concerning pleasantness judgments. Instead they used the nonnegative weight hypothesis and various theoretical and empirical considerations of subjects' phenomenology to predict that VO and SVO balance would dominate pleasantness judgments and would be given positive weights. The resulting special case of the general model was reasonably well supported by the data. In general, most S–V–O research to date has involved testing various special cases of the general model given by Equation 1.

Having reviewed the main ideas in the S–V–O formulation, let us now consider the specific issues Anderson raises in his critique.

Technical Issues

Most of Anderson's technical criticisms of the S–V–O work of Gollob and his colleagues are based on issues concerned with tests of goodness of fit.

Testing Goodness of Fit

Can the S–V–O model fit data perfectly? Anderson (1977) begins his formal criticism of the S–V–O model with the comment: "An

odd difficulty faces the user of the balance triad [S–V–O] model: It can always fit the data perfectly. As is well known, the complete analysis of variance model holds by statistical definition and so is not in general a model of substantive process" (p. 142). Proceeding from this base, Anderson repeatedly claims that the S–V–O model has no way of testing goodness of fit and therefore is defenseless against various types of measurement problems. These claims are wrong because most S–V–O work by Gollob, Wyer, and their co-workers has investigated special cases of the complete analysis of variance model, for which appropriate goodness-of-fit tests are available. Incidentally, the methods of analysis typically advocated by Anderson also involve testing special cases of the complete analysis of variance model.

Gollob (1974b, pp. 293–301) spends eight pages discussing plausible restrictions that can be placed on the weights in the complete S–V–O model. Moreover, most of the empirical results reported in Gollob (1974a, 1974b) were directed toward testing special cases of the model that resulted when various combinations of the nonnegative weight, negatively accelerated weight, and relevant-bias hypotheses were used to restrict the parameters of the general model. Other intuitive and theoretical considerations concerning subjects' phenomenology also have been used to make a priori predictions concerning the relative importance of biases in different item sets and under different experimental manipulations (e.g., Rossman & Gollob, 1976; Wyer, 1974, 1975; Wyer, Henninger, & Wolfson, 1975). Each different set of predictions yielded a different special case of the model, which was not forced to fit the data perfectly. Since Anderson (1977, p. 146) explicitly acknowledges that special cases of the complete model can readily be tested, it is clearly misleading for him to ignore the fact that it is precisely such special cases of the complete model that have been the focus of most S–V–O research done to date.

In addition to making the misleading claim that the S–V–O model can always fit data perfectly, Anderson misrepresents the specific methods that Gollob and his colleagues actually have used to assess the goodness of fit of the S–V–O model.

Specific tests of goodness of fit used in S–V–O research. In this section I shall briefly describe several published analyses that were specifically designed to assess the goodness of fit of various special cases of the complete S–V–O model. In discussing the verb inference submodel, Gollob (1974a, pp. 162–163, 165, 167–168; 1974b, p. 304) presented several figures that compared obtained mean ratings with the best fitting predicted order of the mean ratings. These figures showed all occurrences of errors in the predicted order and indicated whether or not each error was statistically significant. This enabled discrepancies from prediction to be examined in detail. These are precisely the kinds of goodness-of-fit tests that Anderson (1977, pp. 145–146) claims are missing from evaluations of the S–V–O model.

Other methods of evaluating goodness of fit also have been used in S–V–O research. Gollob (1974a, pp. 170–171) and Rossman and Gollob (1976, pp. 382–385) presented clearly labeled sections that discussed several analyses explicitly directed toward evaluating the degree to which various hypotheses concerning the general S–V–O model held; failures of the hypotheses to hold would help to identify specific aspects of lack of fit of the associated special cases of the model. In particular, both of these papers reported indices of the magnitude and/or statistical significance of the data's discrepancies from the patterns predicted by the relevant-bias hypothesis, the nonnegative weight hypothesis, and the negatively accelerated weight (or a closely related) hypothesis. Wyer (1975) and Wyer, Henninger, and Wolfson (1975) also have used appropriate methods to test the statistical significance of sources of variation attributable to biases that they predicted would be unimportant in accounting for their data.

Although there are some exceptions (cf. Gollob, 1974b, pp. 294, 318–319), the overall results of the above tests of goodness of fit provide strong support for various special cases of the model that were investigated. The following section rebuts Anderson's

Table 2
S–V–O Weights for Anderson's Item Sets A and C

Bias	Item Set A (helps/harms) Weight	%	Item Set C (helps/hinders) Weight	%
S	−.92	4	−.03	0
V	1.30	9	.12	0
O	.39	1	−.11	0
SV	3.70	71	2.50	59
SO	.79	3	.07	0
VO	1.09	6	.78	6
SVO	1.05	6	1.94	35

Note. S = subject; V = verb; O = object. The weights were taken from Anderson's (1977, p. 149) Table 2. Figures in the columns labeled "%" give the percentage of variation in the item set that is accounted for by each bias. In both item sets the levels for the sentence subject are *kind/cruel* and the levels for the object are *physicians/criminals*.

(1977) criticisms of Gollob's use of correlationlike indices.

Appropriate uses of correlationlike indices in S–V–O research. Anderson belabors the well-known fact that correlations between model predictions and data can be extremely high even when the model is seriously in error. He then, without presenting specific examples or any qualifying conditions, uses this innocuous premise as the basis for various criticisms concerning Gollob's use of correlation coefficients, percentages of variation, and other correlationlike indices. Anderson's criticisms include such sweeping statements as "Clearly, correlations can be extremely misleading in model analysis, for they obscure and misrepresent the facts . . . . The correlation coefficient and similar statistics are not valid as tests of fit . . . . The not infrequent practice of comparing several models in terms of correlations simply compounds the problem" (1977, pp. 145–146). So strong is his damnation of correlations that Anderson does not give even a hint that correlations have appropriate, as well as inappropriate, uses in model analysis.

We will now use the results of S–V–O analyses presented by Anderson (1977, p. 149) to illustrate the major uses that Gollob (1974a, 1974b; Rossman & Gollob, 1976) has made of correlationlike indices in working with the S–V–O model. The data are presented in Table 2, but the descriptions of the specific item content are deferred to a point in the discussion where they are more relevant. For convenience, the following discussion assumes that the percentages being compared are significantly different from each other.

In one or more of the papers cited above, Gollob has made statements analogous to the following: (a) In Anderson's Item Sets A and C, respectively, SVO balance accounts for 6% and 35% of the item-set variation, therefore SVO balance is relatively more important in accounting for variation in Item Set C than in Item Set A. (b) In Anderson's Item Set A, SV balance and SVO balance, respectively, account for 71% and 6% of the item-set variation, therefore SV balance is relatively more important than SVO balance in accounting for variation in Item Set A. (c) The S–V–O verb inference submodel accounts for 92% of the variation in Anderson's Item Set A and for 98% of the variation in his Item Set C, therefore the verb inference submodel is more effective in accounting for variation in Item Set C than in Item Set A. These are appropriate uses of percentage of variation indices; they provide useful summary descriptions of various aspects of the data.

Finally, Anderson (1977) asserts that the common practice of comparing several models in terms of correlations or percentages of variation is not a valid procedure. For a counterexample to Anderson's assertion, consider the fact that the verb inference submodel and the object inference submodel account for 92% and 6%, respectively, of the variation in Anderson's Item Set A. Assuming the verb inference submodel is at least as psychologically meaningful as the object inference submodel, surely these data provide legitimate grounds for concluding that the verb inference submodel provides a better account of the data than does its competitor.

Goodness of fit is not enough. Although no evaluation of a model can be complete without considering the results of detailed tests of goodness of fit, more is needed. In general, it is necessary to evaluate not only

discrepancies from the fit of the model, but also to evaluate the fit itself. Knowledge that there are no reliable or important deviations from a model's fit does not imply that the model is useful in accounting for the data. One typically also wants to know how well the model accounts for the data and whether it accounts for a statistically significant amount of variation. Moreover, contrary to the flavor of Anderson's remarks, it is not enough to consider only statistical tests in evaluating the fit of a model. It is also necessary to consider the psychological meaningfulness of the values obtained for the model's parameters. Work using the S–V–O approach has been encouraging in this regard.

Summary. In summary, all of Anderson's claims concerning methods Gollob has used to test goodness of fit can be refuted. In particular, Gollob's S–V–O articles have focused on studying special cases of the model, which need not fit the data perfectly; appropriate techniques have been used to test the goodness of fit of these special cases; and correlationlike indices have been used appropriately to provide useful summaries of various aspects of the model's relation to data. Let us now discuss scaling issues—the second major category of technical issues that Anderson raises.

The Importance of Using Appropriate Scales

The need for interval response scales. Anderson uses a set of fictitious S–V–O data to illustrate that a perfect additive fit can be destroyed by moving a single point, and that when this is done in a $2 \times 2 \times 2$ design all four interactions, which were initially zero, are changed to have the same nonzero value. He correctly points out that in this instance it would make little sense to interpret the four interactions as indicating that four distinct balance tendencies were operating. He makes the traditional suggestion that it would be more parsimonious to interpret the interactions as resulting from an aberrant response, and in this example the aberrant response could well be the result of a floor effect. Anderson's (1977, p. 143) general argument is that scaling difficulties can cause observed interactions to contain undecipherable confoundings, and that the S–V–O model has no way to guard against such confoundings because it lacks a test of goodness of fit. Let us now examine some of the details of his argument.

It is argued above that in spite of Anderson's claim to the contrary, the S–V–O model does have tests of goodness of fit. For example, a simple test shows that the fictitious verb inference data that Anderson (1977, pp. 142–143) uses to illustrate a floor effect cannot be fit adequately by the S–V–O submodel for verb inferences. In this application the relevant-bias hypothesis predicts that the S, O, and SO sources of variation will be unimportant. In fact, however, these sources account for over 70% of the variation in Anderson's example, and the submodel obviously fails to meet the goodness-of-fit test. Consequently, the S–V–O model would not be expected to be helpful in deciphering the meaning of the observed interactions; other interpretations would be sought—including, of course, interpretations that allow for the possibility of scaling artifacts.[1]

Anderson is correct in pointing out that rating scales are subject to biases which can make it risky to give substantive interpretations to main effects and/or interactions. Fortunately, he also is correct when he later points out that "relatively modest experimental precautions, such as the use of stimulus end anchors, can largely eliminate these biases in many experimental situations" (1977, p. 144). Gollob and Wyer agree and have consistently used such precautions in their research.

Although a floor effect or other artifact can parsimoniously account for Anderson's single set of fictitious data, it seems highly unlikely that scaling artifacts can explain the

[1] Examples of content-based interpretations of some verb inference item sets in which the relevant-bias hypothesis was severely disconfirmed were given by Gollob (1974b, pp. 318–319). These examples involved item sets whose sentence subjects and objects were nonevaluative. Such severe disconfirmations of the relevant-bias hypothesis have rarely been observed in item sets whose sentence components are bipolar and primarily evaluative in content (but for a probable exception see Gollob, 1974b, p. 294).

results that have been observed in the many S–V–O item sets studied by Gollob, Wyer, and their co-workers. In fact, Anderson makes no effort to apply "scaling artifact" interpretations to even a single set of actual data.

In summary, it is agreed that measurement issues must be considered when any model is used to help understand psychological data. By using appropriate experimental precautions to guard against scale artifacts, by treating rating-scale data as though they are interval data, and by assessing the goodness of fit of special cases of the model, Gollob, Wyer, and their co-workers have repeatedly obtained psychologically meaningful results within the S–V–O framework.

The need for ratio stimulus scales. In Anderson's section on "The Need for Ratio Stimulus Scales," he points out that the areas of a set of rectangles can be represented (a) as the product of Base × Height or (b) by an analysis of variance decomposition that expresses the area of each rectangle as the sum of a grand mean, two main effects, and an interaction. He also points out that an analogous situation exists for the volume of rectangular solids. Anderson argues that the analysis of variance decomposition, if taken literally, would imply that the area and/or volume is the sum of several independent processes, and that the main effects and interactions in the analysis of variance decomposition are terms "whose meaning is essentially artificial. . . . These examples, of course, depend on the assumption that the true integration rule is multiplicative in the two or three variables. . . . Thus, the problem of stimulus zeros seems to be relevant to at least some applications of the balance triad model" (1977, p. 145).

Anderson seems casual about assuming which of two mathematically equivalent models describes "the true integration rule." Rather than choosing between two mathematically equivalent models on the basis of an assumption that one knows "the true integration rule," it probably is more reasonable to ask which of the two models is more helpful to use in thinking about the problems of interest. In the case of Anderson's rectangles, it is obvious that the multiplicative model is more helpful. In the case of S–V–O item sets, on the other hand, the multiplicative model discussed by Anderson rarely provides a reasonable alternative to the S–V–O model. For example, using conjoint-measurement techniques (Krantz & Tversky, 1971, pp. 160–163), it can be shown that even if one allows arbitrary monotonic transformations of the response scales, it is impossible for a multiplicative model to account for the data in even one of the six S–V–O item sets Anderson presents in his Table 1 (1977, p. 148).[2] More important, except when grossly implausible restrictions are met, the three-way multiplicative model and the relevant-bias hypothesis are inconsistent with each other. For example, in the case of verb inferences, the relevant-bias hypothesis implies that S, O, and SO sources of variation are all equal to zero. Except in trivial cases, the resulting S–V–O submodel for verb inferences and the three-way multiplicative model can both be correct if, and only if, $w_{svo} \neq 0$ and $w_v w_{svo} = w_{sv} w_{vo}$.[3] These conditions imply that SVO

[2] In order for a multiplicative rule to fit the data in an S–V–O item set, it is necessary that the ordering for the four combinations of S and V either be the same or be exactly reversed for the two levels of O. Similar relations must hold for the ordering of the four combinations of S and O at each level of V, and for the four combinations of V and O at each level of S. None of Anderson's item sets satisfies these necessary conditions; many of the failures are substantial and seem unlikely to be due to error in the data.

[3] This footnote shows that the claim made in the text is correct. If the multiplicative model holds, we have

$$y_{ijk} = S_i V_j O_k + \text{constant},$$

where y_{ijk} is the likelihood judgment of the ijkth sentence in the item set, and where i, j and k each are either positive or negative. Thus, y_{+-+} denotes the sentence with a positive subject, negative verb, and positive object. S_+ and S_-, respectively, refer to the scale values for the positive and negative sentence subject, and V_+, V_-, O_+, and O_- are analogously defined. In order to avoid trivial cases of the multiplicative model we assume that $S_+ \neq S_-$, $V_+ \neq V_-$, and $O_+ \neq O_-$, that is, we assume that the positive and negative scale values for any given sentence component are not equal to each other.

The weight for an S–V–O bias is equal to half the difference between the mean of the four sentences that have the bias and the mean of the four sen-

balance can never have a weight of zero. Moreover, if SV balance and/or VO balance have a weight of zero, then V positivity must also have a weight of zero, and vice versa. Since these highly restrictive and psychologically vacuous conditions are rarely met, the multiplicative model usually cannot account for the data obtained in the many studies that have found strong support for the relevant-bias hypothesis (Gollob, 1974a, 1974b; Rossman & Gollob, 1976; Wyer, 1974, 1975; Wyer, Henninger, & Wolfson, 1975). Moreover, since we seek parsimonious understanding of many sets of data—not merely of isolated cases—the S–V–O representation usually would be preferable even in those rare cases where the multiplicative model could adequately describe the data.

---

tences that do not have the bias. Thus, when the multiplicative model holds, the S positivity weight can be expressed as

$$w_s = (y_{+..} - y_{-..})/2 = (S_+V_.O_. - S_-V_.O_.)/2 = V_.O_.(S_+ - S_-)/2,$$

where a dot replacing a subscript indicates that a mean has been computed over the levels of the replaced subscript. When similar calculations are done for w_o and w_{so}, we obtain

$$w_o = S_.V_.(O_+ - O_-)/2$$

$$w_{so} = V_.(S_+ - S_-)(O_+ - O_-)/4.$$

Since $S_+ \neq S_-$ and $O_+ \neq O_-$, we see that the verb inference submodel that assumes $w_s = w_o = w_{so} = 0$ can hold if, and only if, $V_. = 0$. Proceeding as we did for w_s, and using the fact that $V_. = 0$ implies that $V_- = -V_+$, we obtain the following expressions for the weights for V positivity, SV, VO, and SVO balance:

$$w_v = S_.O_.V_+$$

$$w_{sv} = O_.V_+(S_+ - S_-)$$

$$w_{vo} = S_.V_+(O_+ - O_-)$$

$$w_{svo} = V_+(S_+ - S_-)(O_+ - O_-).$$

Since $V_+ = -V_-$ and $V_+ \neq V_-$, we see that $V_+ \neq 0$, and therefore the equation for w_{svo} implies that the weight for SVO balance can never be zero. It is easy to verify that the above equations for the weights also imply that $w_v w_{svo} = w_{sv} w_{vo}$. Thus, if these conditions are not met, then the multiplicative model and the S–V–O submodel for verb inferences cannot both be correct. Similar conclusions hold for the subject inference and object inference submodels.

Summary

Anderson, of course, is correct in asserting that measurement issues and tests of goodness of fit must be considered when evaluating models that are intended to help us understand psychological data. The above discussion demonstrates that contrary to the charges Anderson makes in his critique, Gollob and his colleagues have dealt with these matters in careful, effective, and responsible ways. The technical portion of Anderson's critique, however, ignores the fact that no matter how sound one's measurement methods or how well one's model fits some data, it is necessary to consider the question of whether the model makes sense. Surely an investigator oriented toward using models that make good psychological sense would not seriously suggest using an analysis of variance decomposition to analyze areas of rectangles! For a model to be most useful, the values obtained for its parameters should have sensible interpretations in terms of our best understanding of the psychological processes involved. In the case of the S–V–O model, it has been found repeatedly that substantive interpretations of the weights given to the hypothesized cognitive biases make good psychological sense. At this point it is natural to consider Anderson's criticisms concerning substantive issues.

Substantive Issues

Configurality

Anderson's position concerning ways in which the S–V–O model is affected by configural response patterns is based on a misunderstanding of how weights in the S–V–O model should be interpreted. For this reason, I shall first discuss correct methods of interpreting the weights in S–V–O analyses, and shall then discuss ways that the S–V–O model deals with configural response patterns.

Interpretation of weights in S–V–O analyses. Anderson considers the possibility that the weights in an S–V–O analysis can be interpreted as estimates of the weights of biases in *individual sentences*. He concludes

that this interpretation makes no sense. Gollob and Wyer agree, and they neither made nor intended this interpretation. Although the S–V–O model assumes that the judgment of a given sentence is a linear combination of the various psychological biases, the S–V–O method of analysis does not yield direct estimates of the weights for that linear combination. Gollob (1974b, p. 292) and Wyer (1974) emphasize that the weights obtained in an S–V–O analysis are interpreted as reflecting the direction and importance of biases in accounting for variation *within individual item sets*—not within individual sentences. To compare Gollob's interpretation of the weights obtained in an S–V–O analysis with the interpretation attributed to him by Anderson, let us consider the results of S–V–O analyses of Anderson's Item Sets A and C, which are presented in Table 2.

The data for Anderson's Item Set A consist of college students' responses to eight sentences of the following form: "Bill is *kind* (*cruel*); how probable is it that Bill *helps* (*harms*) *physicians* (*criminals*)?" The italicized words denote evaluatively positive (negative) levels of the sentence subject, verb, and object, respectively. Responses were made orally using a 0–20 rating scale whose ends were defined by the terms *extremely improbable* and *extremely probable*. Anderson's Item Set C was based on responses to eight sentences that were the same as those in Item Set A except that the verb *hinders* was substituted for the verb *harms*. Thus, the four sentences that contain the verb *helps* are common to both item sets.

Anderson reminds the reader that the ratings associated with each sentence in each item set can be recovered by applying the weights in Table 2 to the contrasts that define the biases (see Table 1) and emphasizes that there are large and statistically significant differences in the patterns of weights in the two item sets. On the basis of this observation, he correctly points out that each of the four sentences common to Item Sets A and C (i.e., the sentences that contain the verb *helps*) have two distinct representations, one in Item Set A and one in Item Set C. He then argues that there is no way to choose between these two representations and from this draws the conclusion that the S–V–O method of analysis "would seem to have no use" (1977, p. 149). The confusion here seems to stem from Anderson's failure to recognize that the two different representations are not inconsistent with each other, and therefore it is not necessary to choose between them—both can be correct. The weights obtained in the S–V–O analysis of the two item sets express the four common sentences in terms that are *relative* to the other content in each item set. Thus, in Item Set A, the four sentences that contain the verb *helps* are expressed in terms relative to the four sentences that contain the verb *harms*, whereas in Item Set C, they are expressed in terms relative to the four sentences that contain the verb *hinders*. Obviously the results of an analysis of variance are changed when the data input to it are changed. Since the issue of how to interpret the weights in an S–V–O analysis is at the core of the matter, the following paragraphs discuss interpretation of the weights in detail.

To illustrate the relativistic interpretation of the weights obtained in an S–V–O analysis, let us consider the results Anderson obtained for verb positivity. In Item Set A, verb positivity was given a weight of 1.30, and in Item Set C it was given a weight of .12. Using correct methods of interpretation, one would view the different weights given to verb positivity in Item Sets A and C as indicating that for the particular sentence subjects and objects used, verb positivity is more important in accounting for verb inferences when the verbs are *helps* and *harms* (Item Set A) than when the verbs are *helps* and *hinders* (Item Set C). That is, on the average, the tendency for judges to infer that a person will help rather than harm, regardless of whether the person is described as kind or cruel, and regardless of whether the object is physicians or criminals, is stronger than the comparable tendency for judges to infer that a person will help rather than hinder.

Further insight into the meaning of the weights obtained in an S–V–O analysis can be gained by recalling that the analysis of variance contrast that defines a particular

bias assigns +1 values to the sentences that have the bias and assigns −1 values to the sentences that do not have the bias. Thus, the weight for a bias is closely related to the difference between the mean of the four sentences that have the bias and the mean of the four sentences that do not have the bias. Specifically, the weight given to a bias is equal to half the difference between these two means. In his uses of the S–V–O model, Wyer does not report the S–V–O weights per se; instead, he reports the difference between the mean rating of the sentences having the bias and that of the sentences not having the bias in question. Let us now use this perspective to look once again at the difference between the verb positivity weights in Anderson's Item Sets A and C. The mean rating of the four sentences containing the verb *helps* was 9.94 for Set A and 9.49 for Set C. (Since the same four sentences were used in both item sets, probably the difference between these two means is due to measurement error.) The mean rating given to the sentences that used the verb *harms* was 7.34, whereas that for the four sentences that used the verb *hinders* was 9.25. Thus, as one would expect, the average judged probability that kind or cruel people will harm physicians or criminals is considerably less than the average judged probability that kind or cruel people will hinder physicians or criminals. The difference between the weights for verb positivity in the two item sets reflects these characteristics of the data. Thus, the weight for verb positivity in Set A was $(9.94 - 7.34)/2 = 1.30$, and the weight in Set C was $(9.49 - 9.25)/2 = .12$.

In a less technical way, let us now consider interpretation of Anderson's results concerning SV and SVO balance in the two item sets. Suppose that the tendency among judges to think that a person who harms someone is bad is much stronger than their corresponding tendency to think that a person who hinders someone is bad, that is, the evaluative implications of hindering someone depend to a much greater extent on the characteristics of the person who is hindered. This phenomenological assumption would lead SV balance to contribute more to judgments in the *help/harm* item set, and because SVO balance allows for the importance of the characteristics of the object, the assumption would lead SVO balance to be more important in the *help/hinder* item set. As shown in Table 2, this is what occurred in Anderson's data. Thus, when interpreted correctly, findings that Anderson emphasizes as invalidating the S–V–O model are seen in reality to be amenable to a psychologically plausible interpretation in terms of the model.

Having shown that Anderson's position is based on an incorrect interpretation of the weights in S–V–O analyses, and having discussed correct methods of interpreting the weights, I shall now discuss more specifically the question of how the S–V–O model deals with configural responding.

Configural effects and the S–V–O model. As an example of a configural response, Anderson emphasizes that with respect to verb inference judgments, the difference between harming and hindering physicians is much less than the difference between harming and hindering criminals. The judged likelihood of the verb depends in part on the object. Anderson's discussion conveys the incorrect impression that S–V–O researchers have not been concerned with such configural effects. Gollob and Wyer have discussed dozens of S–V–O analyses of item sets that involve highly configural responses. For example, Item Sets 1 and 4 in Gollob's (1974a, p. 165) Table 2 include data showing that with respect to verb inference judgments, the difference between helping and admiring friendly people is much less than the difference between helping and admiring unfriendly people. His subjects inferred that a friendly person was reasonably likely to help both friendly and unfriendly people, but they inferred that admiration was considerably more likely to be directed toward friendly, rather than unfriendly, people. This help–admire example closely parallels Anderson's harm–hinder example of configural responding. Gollob (1974a, pp. 166–168; 1974b, p. 303) grouped together several item sets with configural patterns similar to the help–admire example and found that the hypothesized cognitive biases of the S–V–O model were helpful in thinking about the results.

In general, since the S–V–O model and its hypothesized cognitive biases are psychologically meaningful and can account for highly configural data such as those presented by Anderson, it follows that configural responding poses no special problems for the formulation. The model simply conceptualizes configural responding in terms of the effect such responding has on the relative importance of the hypothesized cognitive biases of the model.

We have seen that the S–V–O system can handle configurality issues of the type raised by Anderson. There also is another sense in which the S–V–O system is intimately concerned with configural aspects of judges' responses. The S–V–O conceptualization recognizes that a given piece of information can contribute to judgments simultaneously in a variety of ways, and a basic assumption of the model is that judges often use configurations of information (combinations of S, V, and O) as single informational units. In the present article, these informational units are referred to as cognitive biases; Wyer often refers to them as informational cues. At one level, the fact that notions of configurality are essential in defining the hypothesized biases is one reason that nonconfigural models, such as the multiplying model advocated by Anderson, typically are unable to account for data that can be conveniently conceptualized in terms of the S–V–O model (see Footnotes 2 and 3).

In summary, if one is to obtain sensible interpretations of the weights obtained in S–V–O analyses, one must interpret them in the context of the item sets from which they are obtained. When this is done, configural response patterns pose no special problem, and the weights often can be used to compare the contributions of hypothesized biases across different item sets in a systematic and psychologically meaningful way.

Nonadditivity

Anderson argues that if the hypothesized S–V–O biases are psychologically real, they probably are combined by an averaging, rather than an adding, rule. The basic argument Anderson presents in support of his position is that because the averaging rule has worked well in numerous investigations, it probably also would work well if applied to cognitive biases. The question is one that can be investigated empirically, and there seems to be little value in engaging in further speculation on the issue. It is important, however, to realize that the ratio of the weight of one bias to another is the same regardless of whether the biases are combined in accordance with an additive or with a simple weighted-average rule. Thus, the "adding versus averaging" question raised by Anderson is irrelevant if one's objective is to compare the *relative* importance of the hypothesized biases in affecting judgments.

An Alternative Approach

Anderson's (1977) section entitled "An Alternative Approach" presents (a) an alternative approach for the analysis of S–V–O item sets, (b) a criticism of a procedure Gollob (1974b, pp. 305–306, 313–316) used to transform 2×2 item sets to make them amenable to S–V–O analyses, (c) a multiplicative model that is sometimes useful in the analysis of verb inference data when sentences are constructed by using one verb in combination with several levels of sentence subject and object, and (d) a set of general arguments advocating use of simple algebraic models to study problems in implicit personality theory. I shall consider each of these topics in turn.

Anderson's Method for Analyzing S–V–O Item Sets

In his subsection entitled "A Multiplying Model for Subject–Object Integration" (1977, pp. 151–153), Anderson implicitly advises that one should avoid collecting data in the $2 \times 2 \times 2$ form of an S–V–O item set. If, however, one ends up with verb inference data in this form, he recommends that the S–V–O item sets be investigated as two separate 2×2 designs—one for each of the two verbs. He recommends that each of the 2×2 designs be conceptualized and analyzed in terms of a multiplicative model

that states that the judged likelihood of the verb can be expressed as the product of a number assigned to the sentence subject and a number assigned to the sentence object. He then says that "the 2×2 design does not provide a proper test of goodness of fit for the multiplying model" (p. 153), and he goes on to argue that a better test of the multiplying model can be obtained by quantifying the subject and object term (see his Figure 3, p. 151).

Anderson's comment highlights a limitation of using the procedure he recommends for the analysis of S–V–O item sets. The limitation, however, does not apply to S–V–O analyses of intact $2 \times 2 \times 2$ item sets because, as was emphasized above, appropriate tests of goodness of fit are possible and have been used routinely in S–V–O research. Moreover, the S–V–O model enables one to conceptualize the data in an intact item set in terms of the hypothesized cognitive biases, whereas the analysis method proposed by Anderson breaks the design into two separate pieces and provides no natural way of integrating the results of the two analyses. In addition, the S–V–O model is used not only in the investigation of verb inferences, which are discussed by Anderson, but also in the investigation of subject inferences, object inferences, pleasantness, and other types of social judgments.

Transforming 2×2 Analyses into Equivalent $2 \times 2 \times 2$ Analyses

Anderson (1977, pp. 152–153) uses the terms "artificial," "misleading," and "awkward" to describe a procedure Gollob (1974b, pp. 305, 313) used to enable Rodrigues's (1968) verb inference data and Osgood and Tannenbaum's (1955) subject and object inference data to be analyzed using the S–V–O model. The procedure involved transforming two-way analyses into *equivalent* three-way analyses. Rodrigues's verb inference data will be used to describe the procedure and the reasoning that motivated its use; generalization of the procedure to subject and object inference data is obvious.

The transformation. In each of three different conditions, Rodrigues presented experimental subjects with four items that fit the following format: "You dislike [or like] *o*; how does *o* feel about *x* [an issue the experimental subject either favored or opposed]?" In each condition 20 subjects responded to each item by indicating whether he or she predicted that *o* had a positive feeling or a negative feeling toward the issue. When both the person, *o* (i.e., the sentence subject), and the issue, *x* (i.e., the sentence object), were negatively evaluated, 13 of Rodrigues's 20 experimental subjects inferred a positive verb and 7 inferred a negative verb; Gollob expressed the triads as S–V–O sentences and in this case assigned a "probability score" of 13 to the $- + -$ sentence and a score of 7 to the $- - -$ sentence. The other three items were dealt with in a similar manner, so that an S–V–O item set consisting of eight sentences was generated. The S–V–O model was then used to analyze the resulting $2 \times 2 \times 2$ item set.

When S–V–O item sets are generated in this manner, the relevant-bias hypothesis is forced to hold perfectly and therefore cannot be tested in data of the type collected by Rodrigues. Gollob (1974b, pp. 305–306, 313) emphasized this point and discussed ways of taking this fact into account in interpreting results of S–V–O analyses that are conducted on data of this type. It should be emphasized, however, that the transformation is rarely needed because S–V–O research typically involves collection of the type of $2 \times 2 \times 2$ data sets for which it was designed.

Anderson's recommendation. Rather than apply the transformation used by Gollob, Anderson (1977, pp. 151–153) recommends that 2×2 analyses of variance be applied to data of the type collected by Rodrigues. Let us use Rodrigues's data to discuss the interpretation of such an analysis. Averaging over Rodrigues's three conditions, the mean probability scores for the sentences containing a positive verb were $+ + +$ (18.7), $+ + -$ (4.0), $- + +$ (7.7), and $- + -$ (13.0). For convenience of exposition we subtract the "midpoint" of 10 from each of these scores before doing the analysis of variance; if the grand mean is larger than zero on this new scale, we conclude that

over half of the experimental subjects inferred the positive verb. In the two-way analysis recommended by Anderson, the weights given to the standard analysis of variance contrasts that define the grand mean, S, O, and SO effects are .85, .50, 2.35, and 5.00, respectively.[4] The large weight given to the SO interaction suggests that on the average, people we like are judged more likely to agree with us on issues than are people whom we dislike. Thus, in the 2×2 design Rodrigues used, the predicted effects of Heiderian balance are shown by the presence of a large SO interaction. In an S–V–O analysis, on the other hand, Heiderian balance leads to a prediction of an SVO interaction. The following discussion shows that this difference is no cause for alarm.

Comparison of the two analyses. The most important thing to realize in comparing the two analyses of verb inferences is that there is an exact correspondence between the grand mean, S, O, and SO sources, respectively, in the two-way analysis recommended by Anderson, and the V, SV, VO, and SVO sources in the three-way analysis used by Gollob. Standard analysis of variance procedures give identical weights to the contrasts that define appropriately paired sources of variation in the two designs. The weights given to the grand mean, S, O, and SO in the three-way design are all necessarily equal to zero. *Thus, except for the names given to the sources of variation, the two methods of analysis yield identical results.* Since the two analyses have identical psychological implications, the choice between them should be made on the basis of which labels are most convenient to use.

When Gollob's three-way analysis is used, one obtains results directly interpretable in terms of the hypothesized cognitive biases. When Anderson's two-way analysis is used, however, much unnecessary confusion results. Consider, for example, a situation in which Heiderian balance is predicted to have an important effect on subject, verb, and object inference judgments. When using three-way analyses, an SVO interaction is implied by the Heiderian balance prediction regardless of the type of inference. When using two-way analyses, however, the same hypothesized balance tendency leads one to predict a VO interaction in the case of subject inferences, an SO interaction in the case of verb inferences, and an SV interaction in the case of object inferences. The confusion is further compounded by the fact that in standard S–V–O item sets, SO, SV, and VO refer to three different hypothesized cognitive biases—none of which are equivalent to traditional Heiderian balance. The method Gollob used to transform two-way analyses into equivalent three-way analyses avoids these difficulties and enables the S–V–O model's concepts and notation to be used in a consistent manner in investigating data involving different types of judgments and different types of data collection procedures.[5]

In summary, in order to use the S–V–O model to reanalyze their data, Gollob transformed Rodrigues's (1968) and Osgood and Tannenbaum's (1955) 2×2 data sets into equivalent $2 \times 2 \times 2$ data sets. Contrary to Anderson's (1977, p. 152) assertion, there is nothing artificial or misleading about the transformation. It avoids the notational confusion that Anderson's method introduces, and provides a convenient way to conceptualize the results in a 2×2 data set in terms of the hypothesized cognitive biases used in the S–V–O system.

A Multiplicative Model for Verb Inference Data

With the purpose of illustrating an alternative to the S–V–O approach, Anderson recommends using Oden and Anderson's (1974) multiplicative model to study verb inference judgments that are obtained in response to sets of sentences constructed by using one verb in combination with several quantified levels of sentence subject and object. The data Anderson uses to illustrate the model consist of judges' responses to stimuli of the

[4] When an analysis is done on the probability scores for the sentences containing negative verbs, the grand mean and weights have the same absolute value but are reversed in sign.

[5] Although Anderson used a rating scale example to present his criticism of the procedure Gollob used to apply the S–V–O model to two-day designs, the issues involved are identical to those discussed above.

following type: "Mr. X is *very sociable*; how likely is it that he will *socialize with moderately unsociable* people?" In both the sentence subject and object positions, the adjective was graded in five levels from very unsociable to very sociable and the verb *socialize with* was used in all sentences. Although the method Anderson (1977) and Oden and Anderson (1974) used to test the goodness of fit of their multiplicative model is not entirely appropriate,[6] it is clear that if a constant term is added to their model, it can provide a reasonable account of the data in Anderson's Figure 3 (1977, p. 151).

Anderson recommends use of a nonconfigural model. Without attempting to explain the apparent paradox, Anderson presents the *nonconfigural* multiplicative model as providing a method of attacking problems due to configurality (1977, pp. 147–148, 156). Anderson seems to be saying we should study problems due to subjects' configural responding by avoiding them. For example, to study verb inferences using the approach Anderson recommends, the data sets investigated must meet several restrictions explicitly designed to eliminate configural responding on the part of the judges: (a) The data set must contain only one verb, (b) the semantic relationship between the sentence subject and verb, and also the semantic relationship between the verb and object, must somehow be qualitatively fixed in all sentences in the data set, (c) the sentence subject and verb must be capable of being quantitatively varied without disturbing the fixed qualitative semantic relationships, and, finally, (d) the likelihood judgment must be a multiplicative function of the scale values assigned to the sentence subjects and objects. Oden and Anderson's results are of interest in their own right because they show that data sets in which these stringent requirements are met can in fact be constructed. Since, however, none of the conditions required by Oden and Anderson's method need to be met when using the S–V–O approach, it is not clear what relevance their results have for assessing the usefulness of the S–V–O formulation.

Anderson's approach and the S–V–O approach are concerned with different problems for which different models and different data collection procedures are appropriate. If one's interest is in identifying sets of social judgment data that are highly quantified and can be fit by nonconfigural models, the approach recommended by Anderson is preferable to the S–V–O approach. On the other hand, the S–V–O approach is preferable if one's interest is in studying social judgments in the context of $2 \times 2 \times 2$ item sets in which configural responding can be investigated in terms of the hypothesized cognitive biases.[7]

Use of Simple Algebraic Models in the Analysis of Social Judgment

In the last few pages of his critique, Anderson (1977, pp. 153–156) argues that problems in implicit personality theory and general social inference can best be investigated by use of the methods of data analysis and data collection that he typically uses in his own work (cf. Anderson, 1970, 1971). Although

[6] Except for a change of notation, the multiplicative model presented by Anderson (1977) and Oden and Anderson (1974) is

$$y_{ij} = A_i B_j,$$

where y_{ij} denotes the judged likelihood for the sentence containing the ith sentence subject and the jth object, and A_i and B_j denote the scale values associated with the stimuli. Anderson (1977, p. 151) notes that the data in his Figure 3 form a diverging fan of straight lines, which is characteristic of the multiplying model, and on this basis he concludes that the data support the model given above. Oden and Anderson (1974) report that an analysis of variance test of goodness of fit indicated that the data did not deviate significantly from the linear fan pattern. However, unless a transformation of the response scale is allowed (an option not mentioned by Anderson or by Oden and Anderson), these methods test the goodness of fit of the model

$$y_{ij} = A_i B_j + \text{constant},$$

rather than the model proposed by Oden and Anderson. Thus, in the absence of some independent a priori certainty that the constant term in the model is zero, the test of goodness of fit used by Anderson (1977) and by Oden and Anderson (1974) is inappropriate.

[7] Although some work relevant to quantifying various notions of cognitive bias has been reported (e.g., Gollob, 1968b; Gollob & Fischer, 1973; Gollob, Rossman, & Abelson, 1973, p. 29; Heise, 1969), at present the S–V–O model assesses the biases only at the level of "present" or "absent."

these methods are based on simple algebraic models that are essentially nonconfigural in nature, it is clear that many important problems in social judgment can be effectively approached by their use. Gollob has used such methods when they have been appropriate for the problems under investigation.

One such example is provided by Gollob's (1965, 1968b) study of impression formation and word combination in simple subject–verb–object sentences. In that study, judges made evaluative ratings of men described by sentences that were constructed from factorial combination of several adjectives, verbs, and objects (e.g., "The vicious man likes beggars," "The kind man praises communists"). As part of the larger study, Gollob used several different models to analyze the mean evaluative ratings (averaged over observations and adjectives) of verb–object combinations. One set of analyses compared two multiplicative models of verb–object integration. One model treated the direct evaluative ratings of the isolated verbs and objects as interval-level scale values, and the other treated the verb and object main effects as interval-level scale values in the model. (The latter model is the type of algebraic model whose use is advocated by Anderson.) Interestingly, although the model based on the main effects estimated 2.5 times as many free parameters as did the model based on direct evaluative ratings of isolated sentence components, it was slightly *less* effective in accounting for variation in the ratings of the verb–object combinations.

Gollob (1965, 1968b) also tested models more complex than the ones just described. Using Gollob's (1968a, 1968c; Tucker, 1968) Factor Analysis of Variance (FANOVA) model, it was found that two multiplicative terms were needed in order to adequately account for the variation in ratings of the verb–object combinations. Each of the multiplicative terms suggested a different aspect of meaning that contributed to the evaluative ratings. The types of model whose use is advocated by Anderson are nonconfigural and do not allow for the possibility that more than one multiplicative term can affect judgments. Wyer and Hinkle (1976) also have reported social judgment data in which the FANOVA heuristic of representing the data by a series of multiplicative components was found to be useful. And finally, Gollob and Rossman (1973) found that both evaluative and potency judgments of sentence components were important in accounting for potency judgments of sentence subjects. These few examples emphasize that although simple algebraic models such as those advocated by Anderson often are useful, they are far from complete in their ability to describe social judgment data. One must be prepared to modify one's models and one's methods of analysis as dictated by the characteristics of the problems of interest and the data at hand.

Anderson's critique implies that if a model has a simple algebraic form and passes appropriate tests of goodness of fit, then that model and its associated scale values are correct, valid, and subjectively real. Anderson seems to forget that radically different models can be roughly equivalent in their ability to fit data and that the choice between such models must be made on the basis of convenience, theoretical considerations, or other substantive grounds. Anderson might agree, however, that it is important to temper one's enthusiasm at finding a model that fits data well by considering Tukey's (1977, p. 586) admonition concerning model fitting: "Even when we see a very good fit—something we know has to be a very useful summary of the data—we dare not assume that what is fitted is of exactly the right form; we dare not believe we have found a natural law."

In summary, Anderson's critique ends by advising researchers in implicit personality theory to focus their work sharply on the investigation of problems that are amenable to quantitative analysis by use of simple algebraic models. In contrast to Anderson's recommendation that we tailor our questions to fit a particular research paradigm and method of analysis, the present article recommends that we tailor our research paradigms and data analysis methods to fit our questions.

References

Anderson, N. H. Functional measurement and psychophysical judgment. *Psychological Review*, 1970, *77*, 153–170.

Anderson, N. H. Integration theory and attitude change. *Psychological Review*, 1971, *78*, 171–206.

Anderson, N. H. Some problems in using analysis of variance in balance theory. *Journal of Personality and Social Psychology*, 1977, *35*, 140–158.

Anderson, N. H. Indeterminate theory: Reply to Gollob. *Journal of Personality and Social Psychology*, 1979, *37*, 950–952.

Gollob, H. F. *A combined additive and multiplicative model of word combination in sentences* (Doctoral dissertation, Yale University, 1965). (University Microfilms, No. 65–9674)

Gollob, H. F. Confounding of sources of variation in factor-analytic techniques. *Psychological Bulletin*, 1968, *70*, 330–344. (a)

Gollob, H. F. Impression formation and word combination in sentences. *Journal of Personality and Social Psychology*, 1968, *10*, 341–353. (b)

Gollob, H. F. Rejoinder to Tucker's "Comments on confounding of sources of variation in factor-analytic techniques." *Psychological Bulletin*, 1968, *70*, 355–360. (c)

Gollob, H. F. Some tests of a social inference model. *Journal of Personality and Social Psychology*, 1974, *29*, 157–172. (a)

Gollob, H. F. The subject–verb–object approach to social cognition. *Psychological Review*, 1974, *81*, 286–321. (b)

Gollob, H. F., & Fischer, G. W. Some relationships between social inference, cognitive balance, and change in impression. *Journal of Personality and Social Psychology*, 1973, *26*, 16–22.

Gollob, H. F., & Rossman, B. B. Judgments of an actor's "power and ability to influence others." *Journal of Experimental Social Psychology*, 1973, *9*, 391–406.

Gollob, H. F., Rossman, B. B., & Abelson, R. P. Social inference as a function of the number of instances and consistency of information presented. *Journal of Personality and Social Psychology*, 1973, *27*, 19–33.

Heise, D. R. Affectual dynamics in simple sentences. *Journal of Personality and Social Psychology*, 1969, *11*, 204–213.

Insko, C. A., Songer, E., & McGarvey, W. Balance, positivity, and agreement in the Jordan paradigm: A defense of balance theory. *Journal of Experimental Social Psychology*, 1974, *10*, 53–83.

Krantz, D. H., & Tversky, A. Conjoint-measurement analysis of composition rules in psychology. *Psychological Review*, 1971, *78*, 151–169.

Oden, G. C., & Anderson, N. H. Integration of semantic constraints. *Journal of Verbal Learning and Verbal Behavior*, 1974, *13*, 138–148.

Osgood, C. E., & Tannenbaum, P. H. The principle of congruity in the prediction of attitude change. *Psychological Review*, 1955, *62*, 42–55.

Rodrigues, A. The biasing effect of agreement in balanced and imbalanced triads. *Journal of Personality*, 1968, *36*, 138–153.

Rossman, B. B., & Gollob, H. F. Social inference and pleasantness judgments involving people and issues. *Journal of Experimental Social Psychology*, 1976, *12*, 374–391.

Tucker, L. R. Comments on "Confounding of sources of variation in factor-analytic techniques." *Psychological Bulletin*, 1968, *70*, 345–354.

Tukey, J. W. *Exploratory data analysis*. Reading, Mass.: Addison-Wesley, 1977.

Wyer, R. S., Jr. *Cognitive organization and change: An information processing approach*. Potomac, Md.: Erlbaum, 1974.

Wyer, R. S., Jr. Some informational determinants of one's own liking for a person and beliefs that others will like this person. *Journal of Personality and Social Psychology*, 1975, *31*, 1041–1053.

Wyer, R. S., Jr., Henninger, M., & Wolfson, M. Informational determinants of females' self-perceptions and observers' perceptions of them in an achievement situation. *Journal of Experimental Social Psychology*, 1975, *32*, 556–570.

Wyer, R. S., Jr., & Hinkle, R. L. Informational factors underlying inferences about hypothetical persons. *Journal of Personality and Social Psychology*, 1976, *34*, 481–495.

Received October 17, 1977 ■

Postscript: Anderson's Reply Is Not Responsive to the Issues

This postscript is a rejoinder to the immediately following article by Anderson (1979). The preceding part of the present article directly responds to every major criticism Anderson (1977) presented in his original critique of the S–V–O formulation. Anderson's (1979) article on the other hand, does not attempt to rebut the points made above; it merely repeats a few of the many questionable claims Anderson made in his original paper. This postscript considers (a) Anderson's shift of emphasis away from technical issues, (b) Anderson's insistence that weights obtained from S–V–O analyses be applied to single sentences, whereas Gollob has specifically said that they apply to item sets, and (c) Anderson's claims concerning goodness of fit.

Indeterminate Criticism

Apparently the central thrust of Anderson's criticism of the S–V–O formulation has shifted. In his reply, Anderson (1979) complains that the main point of his original paper is becoming lost in a mass of technical detail. This is a surprising complaint. Anderson's (1977) original paper spent a dozen pages presenting technical criticisms, but only three pages on what he now describes as "the main issue." Thus, contrary to the implication of his reply, Anderson's critique was primarily technical in nature. In the first part of the present article I have shown that

Anderson's (1977) technical criticisms of the S–V–O approach are refutable, and that the methods Anderson proposes for dealing with problems of configurality are ineffectual. Rather than discuss my refutations of his criticisms concerning technical issues, Anderson now states that technical issues played a minor part in his original critique.

Choice of Problem

Anderson begins his reply by saying that the "main point" of his original critique and of his reply is that an S–V–O analysis "yields a theoetical representation of the response to a *single sentence* that is arbitrary and indeterminate" (p. 950, emphasis added). Thus, Anderson has ignored the explicit statement that "the weights obtained in an S–V–O analysis are interpreted as reflecting the direction and importance of biases in accounting for variation *within individual item sets*—not within individual sentences" (p. 940 above). Moreover, at the end of his reply, Anderson (1979) implies that Gollob (1974b, Equation 1, p. 292) presented the basic S–V–O model as applying to single sentences, regardless of context. Contrary to Anderson's implication, Gollob (1974b, p. 292) specifically stated that the model equation and weights apply to sentences in the context of item sets. Since the weights obtained in S–V–O analyses do not apply to single isolated sentences, Anderson's "main point" is irrelevant in evaluating the usefulness of S–V–O analyses.

Anderson (1979) complains that Gollob has not emphasized that the weights obtained in S–V–O analyses must be interpreted in the context of the particular item sets being analyzed. This is an odd complaint. The derivation of the relevant-bias hypothesis, discussions of how weights should be interpreted, and the use actually made of weights in empirical work all show that S–V–O weights must be interpreted in the context of the item sets being analyzed. Moreover, with the single exception of Anderson (1977), every study conducted within the S–V–O framework has considered theoretical bases for making a priori predictions of the specific ways in which the weights given to biases will vary as a function of the particular content of the item sets being investigated. Finally, the results of most S–V–O studies demonstrate how the approach is used to compare the contributions of hypothesized cognitive biases across item sets of varying content in a systematic and psychologically meaningful way.

Anderson (1979) defines "operative weights" as "the weights that were operative in the actual judgment of each sentence" (p. 950). Because S–V–O analyses do not yield estimates of the operative weights, Anderson concludes that the S–V–O model is indeterminate, meaningless, untestable, and unable to deal with configural responses. It seems that Anderson's concern with operative weights prevents him from seeing any value whatsoever in studying and making predictions about other kinds of weights. He ignores the fact that the whole purpose of the S–V–O model is to conceptualize social cognition problems in the context of item sets of eight sentences; the S–V–O model was not designed to provide estimates of operative weights.

Although the problem of estimating operative weights is not the problem addressed by S–V–O analyses of $2 \times 2 \times 2$ item sets, estimation of such weights has been the focus of other research conducted by Gollob and his colleagues. For example, studies of impression formation by Gollob (1968b), Gollob and Fischer (1973), and Gollob and Rossman (1973) have estimated the values of operative weights that are given to various cognitive biases. In these studies, quantified indices of some of the S–V–O biases are used in place of the qualitative indices considered in the $2 \times 2 \times 2$ analyses. The basic ideas underlying the S–V–O biases have been useful in both types of studies.

In summary, Anderson (1977, 1979) does not attempt to evaluate what the S–V–O approach does, how it has been used, or even what claims have been made for it. His criticisms might be germane to other models constructed for other purposes; they surely are not relevant to the S–V–O model.

Goodness of Fit

Anderson (1979) emphasizes that he and Gollob are in agreement on several important methods that should be used in testing goodness of fit. Agreement about what methods *should be* used, however, is not relevant to the present debate. Anderson's original critique did not say Gollob's beliefs were in error; it said that the methods Gollob actually used to test goodness of fit were in error. The nonpostscript portion of the present article presents several specific examples that show Anderson's criticism is incorrect and misleading.[8]

[8] Since Anderson is well known for the stern admonitions he dispenses to those who fail to test goodness of fit, it is interesting to consider the fol-

Anderson (1977, 1979) sets up a classic "straw man" about goodness of fit by emphasizing that if all seven biases in the S–V–O model are assumed to be meaningful in every item set, the model cannot be invalidated. Although this statement is correct, it ignores the fact that only *restricted versions* of the model are discussed in S–V–O work. I am not aware of even one S–V–O study that has assumed that all seven biases will be relevant in every item set studied. It is obvious, and has been repeatedly stated, that the cognitive biases of the model must be psychologically meaningful on a priori grounds in order to be taken seriously. As far as I know, Anderson is the only person who talks as though anyone would expect all seven biases to be psychologically meaningful in every item set.

Contrary to Anderson's (1977, 1979) accusations, the present article has demonstrated that Gollob, Wyer, and their colleagues have consistently used appropriate methods to test the goodness of fit of several restricted versions of the S–V–O model. Anderson's (1979) sole response to this demonstration is to assert that restricted versions of the model are "merely plausible hunches" and have "no special theoretical status." Such polemic cannot take the place of careful model analysis and evaluation.

Conclusion

In summary, Anderson's (1979) reply presents as extreme a distortion of the S–V–O approach as did his original critique. I urge readers who are interested in more than the heat involved in this exchange to read Anderson's original critique, together with some examples of research conducted within the S–V–O framework. The research examples are important in showing how the S–V–O framework has been used and in demonstrating its usefulness in dealing with substantive issues in social cognition.

---

lowing points. In discussing goodness of fit, Anderson (1979) refers to Table 2 of his original article and says that the restricted S–V–O model implied by the relevant-bias hypothesis "failed badly in the representation of Design A." It is ironic that neither his original paper nor his reply tells the reader whether the restricted model showed statistically significant lack of fit. Nor does he mention that the restricted model accounted for 92% of the item set variation, or that the nonrelevant biases all had smaller weights than did any of the relevant biases.

Journal of Personality and Social Psychology
1979, Vol. 37, No. 6, 950-952

Indeterminate Theory: Reply to Gollob

Norman H. Anderson
University of California, San Diego

The central criticism of the balance triad model presented by Anderson concerned its theoretical indeterminacy: The same sentence can receive quite different theoretical representations, depending on arbitrary choice of other sentences. In other words, the balance triad model gives an indeterminate description of the cognitive processes involved in the judgment of any given sentence. Gollob's reply agrees on this point.

The Main Issue

The central criticism in my discussion of the balance triad model is conceded by Gollob to be correct (Anderson, 1977; Gollob, 1979). Since this central issue is becoming lost under a mass of technical detail, it deserves reemphasis in simple terms: The balance triad model yields a theoretical representation of the response to a single sentence that is arbitrary and indeterminate.

At face value, the balance triad model is simple. Seven balance tendencies or "cognitive biases" are hypothesized. They determine the judgment of each single sentence by a weighted sum model. The model conceptualizes "how the information carried by the various cognitive biases is combined by perceivers in making social inference judgments." Specifically, it assumes that the seven cognitive biases "are combined linearly to yield the judged probability of a sentence type" (Gollob, 1974, p. 292). Of primary concern, therefore, is the importance or operative weight of each cognitive bias.

Unfortunately, the model analysis yields indeterminate estimates of these weights. To estimate the weights requires the use of seven additional sentences to form a 2^3 item set. In principle, that is legitimate and proper. In this particular case, it has an illegitimate consequence: The estimated weights for the given sentence can have quite different values, depending on arbitrary choice of the additional sentences. Each sentence thus has many different theoretical representations; each representation tells a different theoretical story. Since the choice among them is arbitrary, the theoretical representation of the cognitive biases is indeterminate.

This indeterminacy was the central criticism of my article. This was the point of the "thought experiment." This point was verified in the actual experiment. The same sentence did indeed have two quite different theoretical representations. Each representation created a quite different theoretical picture of the underlying cognitive processes.

In his reply, Gollob (1979) agrees with this criticism: The model "assumes that the judgment of a given sentence is a linear combination of the various psychological biases," but the model analysis "does not yield direct estimates of the weights for that linear combination" (p. 940).

We are thus faced with a peculiar theoretical situation in which there are two different sets of weights. One set refers to the basic cognitive model and may be called *operative weights*; these are the weights that were operative in the actual judgment of each sentence. The other set may be called *estimated weights*; these are the weights that are obtained when the same basic model is

This work was supported by National Science Foundation Grant BNS 75-21235. Support was also received from grants from the National Institute of Mental Health to the Center for Human Information Processing, University of California, San Diego.

Requests for reprints should be sent to Norman H. Anderson, Psychology C-009, University of California, La Jolla, California 92093.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3706-0950$00.75

actually applied to these same judgments. But these estimated weights are not the operative weights. This is a strange kind of model analysis. What can it mean?

Gollob argues that the model analysis actually measures relative importance of the cognitive biases, not for any one sentence but for an item set of eight sentences. But this argument begs the question, for it rests on mere assumption that the model is correct. If the model is not correct for any single sentence, averaging over a set of sentences does not necessarily improve it.

The basic problem, as I have pointed out in other discussions of balance theory, is the problem of configurality. In my experiment (Anderson, 1977), for example, *harms* and *hinders* were interpreted in terms of configural relations to the other parts of the sentence. Such configurality is an important problem, but one to which the balance triad model is insensitive. No doubt Gollob is correct in saying that configurality poses no special problem to the balance triad model. That is not because the model solves the problem, however, but because it buries it by averaging over the varied configuralities involved in eight different sentences.

As it stands, therefore, the balance triad model should perhaps not be called a model at all. Instead, it illustrates the Procrustean, curve-fitting properties of analysis of variance.

In my original article (Anderson, 1977), I concluded that

> The balance triad model of Gollob (1974b), Insko et al. (1974), and Wyer (1975b) represents a notable attempt to bring needed precision into balance theory. If successful, this model could fractionate and measure each of several coacting balance tendencies. In Gollob's applications especially, this attempt went beyond the affective relations of traditional balance theory to consider semantic relations and a more informational analysis.
>
> Despite the present criticisms, the work of Gollob, Insko, and Wyer is important in representing thoughtful and serious attempts to break through the limitations of traditional balance theory to a new approach. There is no doubt about the interest and potential value of this approach. There is, of course, the question of its validity. (p. 154)

The question of validity still remains, more serious than before.

Goodness of Fit

Since the central criticism of the balance triad model was correct, technical details no longer seem to have much relevance. Although there is much to question in Gollob's comments on technical details, they require only minor qualifications of my original discussion. Accordingly, I wish to comment only on Gollob's implication that I have not correctly represented his position on goodness of fit.

On goodness of fit, I explicitly noted that "as Gollob (1974b) has carefully pointed out, the balance triad model can always fit the data perfectly" and that Wyer and especially Gollob "have given serious attention to this problem" of obtaining a test of goodness of fit (Anderson, 1977, p. 145). I also explicitly noted that "restricted versions of the complete balance triad model can readily be tested" (p. 146), again in agreement with Gollob's position. And in general, I completely agree with Gollob that common sense is a much-needed ingredient in model analysis (e.g., Anderson, 1976, in press).

Of course, the basic problem in testing the balance triad model still remains. The restricted versions of the model that can be tested have no special theoretical status. They are merely plausible hunches. If the test rejects the null hypothesis that some weight is zero, that only means that the hunch was wrong and that the weight is not zero after all.

Just this situation arose in the comparison of Designs A and C in Table 2 of my article. The two designs yielded two quite different theoretical representations of their four common sentences. The relevant-bias hypothesis succeeded in the representation of Design C but failed badly in the representation of Design A. I did not take the latter failure as a criticism of the theory, however, because the general model places no restriction on the weights. Gollob evidently concurs; referring to these two theoretical representations in his reply, he states that "both can be correct." In other words, failure of the relevant-bias hypothesis is not considered as a failure of the general model.

Indeed, Gollob (1974) explicitly asserts that "when no restrictions are placed in the weights . . . the model is useful as a method

of decomposing the data from an item set into psychologically interesting components" (p. 293). But when no restrictions are placed on the weights, the model is admittedly untestable. To say that it yields a meaningful decomposition is to assume its truth regardless of testability. In short, the general model disallows only its own disproof.

One further comment on goodness of fit should be added in light of Gollob's reply. It now is clear that the estimated weights obtained from the model analysis are not the weights in the operative model. Such model analyses would seem to have no relevance whatever to testing the operative model.

A Question of Understanding

In his reply, Gollob states that he never intended to make and never did make the theoretical interpretation to which my central criticism was directed. His present position, certainly, is quite different. I apologize for any deficiencies in my understanding, but I did spend considerable time on the article trying to be correct and fair. Since Gollob finds fault with my understanding, I feel it appropriate to add three remarks in self-defense.

First, I sent the original and the revised preprint of my article to Gollob and to other interested parties, together with covering letters asking for comments and criticism. Not one of the replies that I received suggested that my central criticism rested on a misunderstanding of the balance triad model.

Second, it was an editorial requirement that I verify the thought experiment with an actual experiment. But the sole purpose of the thought experiment was to illustrate the theoretical indeterminacy. Neither it nor the actual experiment had any other point or purpose. If Gollob's present position had been clear then, both the thought experiment and the actual experiment would have been unnecessary. Since this point was not raised in the editorial process, it seems proper to say that I was not alone in my understanding of Gollob's position.

Finally, my understanding of Gollob's (1974) theoretical position still seems to me to agree with his statements in that article. It is true, to be sure, that the numerical calculations to estimate the weights require use of a 2^3 item set of sentences. The issue is not how the weights are calculated, however, but what they mean. Gollob (1974) repeatedly refers to the weights of the cognitive biases in inferences or judgments to single sentences; these are the operative weights. No warning is given about any other set of weights. No warning is given that the estimated weights obtained from the model analysis are not the operative weights for any sentence.

Quite the contrary. The basic model (Gollob, 1974, Equation 1, p. 292) explicitly represents the judgment of the given single sentence as a weighted sum of cognitive biases; the same paragraph asserts a few lines later that the model analysis provides "least squares estimates of the weights." If Gollob (1974) considered that two different sets of weights were involved, he should have said so then. Since Gollob (1974) does not even allude to two different sets of weights, the complaints in Gollob (1979) seem ill-taken.

References

Anderson, N. H. How functional measurement can yield validated interval scales of mental quantities. *Journal of Applied Psychology,* 1976, *61,* 677–692.

Anderson, N. H. Some problems in using analysis of variance in balance theory. *Journal of Personality and Social Psychology,* 1977, *35,* 140–158.

Anderson, N. H. *Information integration theory. A case history in experimental science* (Vol. 1). New York: Academic Press, in press.

Gollob, H. F. The subject–verb–object approach to social cognition. *Psychological Review,* 1974, *81,* 286–321.

Gollob, H. F. A reply to Norman H. Anderson's critique of the subject–verb–object approach to social cognition. *Journal of Personality and Social Psychology,* 1979, *37,* 931–949.

Received August 31, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 6, 953-969

Negative Emotional Biasing of Unexplained Arousal

Christina Maslach
University of California, Berkeley

Schachter and Singer's two-factor model of emotion argued that a state of unexplained arousal could as easily be labeled "euphoria" as "anger," depending on the cognitions available in the immediate situation. A theoretical critique of this model suggests that the individual's search for a cognitive explanation of an arousal state is (a) more extensive than originally hypothesized and (b) biased towards negative emotional labels. Because of methodological weaknesses in the original study, a new paradigm, utilizing amnesia for hypnotically induced arousal, was developed to test the contrasting hypotheses. Subjects either did or did not experience unexplained arousal in the presence of a confederate who was displaying either happy or angry emotions. The results of this modified replication, as well as of an exact replication by Marshall and Zimbardo, fail to support the Schachter and Singer model and demonstrate the negative bias in the process of identifying unexplained internal states.

The study of emotion has been slowed by a lack of consensus about the scientific conceptualization of this phenomenon that we each know so well at a personal level. While some researchers define emotion in terms of motives or behavioral responses, others focus on the sensations of bodily changes, subjective experience, or cognitive processes. Within the more cognitive models of emotion, the most influential work has been the classic experiment by Schachter and Singer (1962). Two dimensions of emotion were independently manipulated in the study: physiological arousal and cognitive cues. Some physiologically aroused subjects were correctly informed as to the source of their arousal, while others were either uninformed or given an erroneous explanation. Crosscutting this manipulation of arousal was the manipulation of the emotional cues presented in the immediate situation. Subjects interacted with a confederate who was acting either euphoric or angry. The results appear to show that aroused subjects without a prior appropriate explanation used the available situational cues to label their state in emotional terms. If they were with a euphoric person, they labeled their own arousal as "euphoria," but if they were with an angry person, the same physiological arousal was labeled as "anger." Unaroused subjects or those with an appropriate explanation for their arousal were reported to be uninfluenced by the emotional cues provided by the confederate.

This experiment represented an important conceptual change from previous theoretical positions, all of which postulated a direct causal relationship between physiological arousal and cognitions about emotional states. In contrast, Schachter and Singer proposed that the factors of physiological arousal and emotional cognition could function independently and that they interacted with each other to produce a true emotion. In addition to this interactionist model, Schachter and Singer made a second signifi-

This research was financially supported by Office of Naval Research Contract NOOO 14-67-A-0112-0041 to Philip Zimbardo. I am very grateful to the following people who served as confederates, observers, and interviewers: Virginia Dupraw, Nancy Green, Jocelyn Gunnar, Craig Haney, John Mancolati, Neil Morse, Nancy Perry, and John Rhead. I also want to thank Daryl Bem, Phoebe Ellsworth, Carroll Izard, Gary Marshall, and Philip Zimbardo for their very helpful comments on an earlier draft of this article.

Requests for reprints should be sent to Christina Maslach, Department of Psychology, 3210 Tolman Hall, University of California, Berkeley, California 94720.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3706-0953$00.75

cant contribution by demonstrating the importance of situational determinants of emotion. The theoretical impact of this study cannot be overestimated, foreshadowing as it did the rise of cognitive analysis in social psychology. This general line of thinking played an important role in the current development of attribution theory, as a result of its emphasis on the role of cognition in defining reality. Furthermore, it has inspired several new lines of theory and research about a variety of internal feeling states, including fear, hunger, passion, and pain (see London & Nisbett, 1974; Mandler, 1975; Schachter, 1971; Walster, 1971).

However, despite its ingenious paradigm and its theoretical contributions, the Schachter and Singer (1962) study cannot be accepted uncritically. Although the conclusion is provocative, a careful inspection of the findings suggests that it is not, in fact, supported by the evidence. The initial data analysis reveals no significant difference in self-reported emotion between any of the unexplained arousal groups and the placebo control groups. Although the authors point to the predicted pattern of means within each emotion condition, the absolute size of the effects is minimal, since the difference in emotion between any of the groups amounts to less than one single scale unit on a 9-point scale. The theory also requires absolute qualitative differences in emotion as well as relative differences in emotion ratings between the experimental groups of the euphoria and anger conditions. However, the means of all conditions reflect a feeling of "slight happiness," regardless of the emotional cues provided by the euphoric and angry confederates.

The behavioral data are also weak. The comparison between unexplained arousal groups and placebo controls reaches statistical significance only in the anger condition, not in the euphoria one. Even though this pattern of "anger units" is in line with the authors' theory, the size of this response certainly is not. The low mean score for subjects with unexplained arousal could represent merely one instance of verbal agreement with the angry confederate's position over nearly 20 minutes of interaction. Somewhat stronger behavioral differences emerged only after a sizeable number of subjects were discarded in an internal analysis.

One possible reason for the lack of empirical support may be that there are methodological flaws in the design of the study (see Maslach, 1979, for a review of these problems; see also the critiques by Lazarus, 1968; Leventhal, 1974; and Plutchik & Ax, 1967). Given the theoretical significance of the Schachter and Singer study, one would expect that more methodologically rigorous replications would have already been done. Surprisingly, only one attempt has been made. Marshall and Zimbardo (1979) tested 80 male subjects in a partial replication of the Schachter and Singer paradigm for the euphoria condition. However, they found no empirical support for Schachter and Singer's conclusions. Subjects who had received the standard epinephrine injection, along with an "inadequate" explanation of its side effects, did not differ from placebo controls in either emotional affect or behavior. Furthermore, when the arousal was made stronger and more salient, subjects reported a *negative* emotional state, rather than the positive "euphoria" predicted by Schachter and Singer.

Alternative Views

The lack of empirical support for the conclusions advanced by Schachter and Singer (1962) forces us to consider alternative hypotheses. One possibility is that the Schachter and Singer theory is incorrect. The qualitative differentiation of emotions may have some basis other than situational cues. Moreover, an undifferentiated autonomic arousal may not be the physiological basis for emotion. The Schachter and Singer model may also be incorrect in its assumption that autonomic arousal is a neutral activator steered by available cognitions. Marshall and Zimbardo's findings suggest that attending to such arousal may have a negative quality that will overshadow positive emotional cognitions. If so, this would argue for the importance of physiological feedback in determining the quality (and not just the quantity) of the emotion.

An alternative interpretation is that the Schachter and Singer theory is correct but that the particular paradigm did not provide an adequate test of it. Schachter and Singer present their model as an explanation of emotion in everyday situations, in which the physiological and cognitive factors are completely entwined. However, they test their model by separating the two factors and creating a situation of unexplained autonomic arousal. Such a situation is rather rare and unusual in everyday life, and it may be erroneous to generalize from this special case to more common emotional experiences. Because of its novelty, the experience of unexplained arousal may not be affectively neutral, but may instead produce a bias in the search for an emotional cognition.

Negative biasing. The idea that unexplained arousal is an affectively neutral state seems intuitively correct only for someone who is experiencing such arousal for the first time (such as a child), and who must look to other people and/or the situation to make sense of this strange and puzzling experience. The concept does not apply as well to people who have had numerous experiences pairing arousal with emotional labels and with appropriate arousal instigators. If they were to feel aroused without knowing why, it would not only be an unusual experience but a disturbing or even frightening one, since this uncertainty would be felt as a loss of personal control over their own internal states. Indeed, the concept of unexplained arousal is very close to the clinical definition of "free-floating anxiety," which is always characterized by negative emotional affect. Therefore, unexplained arousal may generate a *biased,* rather than an unbiased, scanning of evidence and alternative hypotheses. That is, the search for an appropriate cognitive label for the arousal could be biased toward available negative instances. A negative bias also seems more probable because noticeable physical arousal is a far more common concomitant of negative emotional experiences than of positive ones (with the possible exception of sexual excitement).

An expanded search. This alternative view still accepts the Schachter and Singer notion of a cognitive search for an emotional explanation. However, their experimental test of the search process was a narrow one that did not consider the full range of information that a person might have used. They only assessed the impact of the confederate's mood on the subject and ignored other information in the immediate situation or in the subject's life experiences. For example, they did not consider the possibility that a person will search his or her memory for past events or prior comparable situations that would provide an appropriate explanation for the current state of arousal. Subjects might have tried to recall similar situations in which they had felt aroused (such as being in a testing situation or visiting a doctor). They might have then used these memories to derive an appropriate emotional cognition ("I must be getting upset about the tests they are going to do, since I always worry about failing" or "I've never liked getting stuck with needles by doctors, so that's why this injection bothers me"). To the extent that subjects' explanations were based on their past history, on attributions about their health, or on nonsocial aspects of the environment, it might explain the relatively weak impact of the confederate's mood or the substantial within-condition variance that attenuated treatment effects.

Although Schachter and Singer (1962) postulate that the relevant emotional cognitions arise from the immediate *situation,* they operationalize this concept in terms of a *person* who serves as a social comparison referent for the subjects. They predict that it is the emotion displayed by this comparison person that will be the "appropriate cognition" which the subjects use to label their own arousal. However, such a prediction must rest on the assumptions that subjects will view the other person as (a) being someone similar and comparable to them, (b) responding to a set of situational demands that are also impinging on the subjects, and (c) behaving in an appropriate way to these demands.

The first assumption may be justified in the experiment, since the comparison person was of the same sex and age as the subject and had a similar educational background. There is less evidence to support the other

two assumptions. An examination of the experimental procedure reveals that the comparison person was not always responding to the same situation that confronted the subject. For example, there is nothing inherent in the messy experimental room that should cause people to experience euphoria. Given this lack of situational inducements,[1] the confederate's playful behavior says more about his particular personality than it does about the appropriate way to respond to a situation. This *dispositional* interpretation is made especially salient by the confederate's own statement that "this is one of my good days" (Schachter & Singer, 1962, p. 384). From the subject's point of view, whatever is causing the confederate to feel euphoric is something unique to him and not something that they share in common. In contrast, the anger condition does provide some relevant situational demands, since the confederate's anger is directed at a long and insulting questionnaire that the subject is also completing.

Finally, there is insufficient evidence that subjects would view the confederate's behavior as appropriate to the situation. In both experimental conditions, the confederate behaved in unusual ways (agitated, manic, and assertive), especially compared to the norm of quiet compliance that is generally characteristic of research subjects. Spontaneously playing with paper airplanes and hula hoops is not a typical thing to do in a scientific laboratory, nor is abruptly walking out of the experiment. To the extent that subjects perceive the confederate as acting in idiosyncratic, unusual, or inappropriate ways, he loses his social comparison reference power, and subjects would not be likely to label their own arousal in terms of their cognitive interpretation of his mood.

Current Study

The current research was designed to further investigate the process of a cognitive search for the causal antecedents of unexplained arousal. It involved a modified replication in which procedural elements were altered to correct some of the original methodological problems. These improvements included the use of a hypnotic induction to produce an arousal state, rather than a drug injection. Previous research has demonstrated that marked physiological changes can be produced hypnotically (see the review by Sarbin & Slagle, 1972). In addition, the use of hypnosis allowed for a better degree of experimental control over the onset and the maintenance of the physiological arousal. The lack of an explanation was produced by a posthypnotic suggestion for amnesia of the true cause of the arousal. Subjects experienced physiological arousal but did not know why the arousal was occurring and did not remember that it was due to the hypnotic suggestion. Posthypnotic amnesia is a phenomenon that has been demonstrated in numerous studies (see Cooper's review, 1972) and is believed to involve a temporary disruption of the normal retrieval process (Evans & Kihlstrom, 1973). Cues that normally aid the process of recall are not utilized as they would be in normal waking memory.

Second, the behaviors of the two emotion conditions were made directly comparable—the nature, level, and timing of the activity were equivalent in both conditions, and only the confederate's obvious emotional state (happy or angry) was varied. The reasons given for the confederate's emotion were ones that the subject could readily share, and the confederate's behavior appeared believable and appropriate in the laboratory setting. Finally, several measures were added to the study, both to assess the validity of

---

[1] In a personal communication (1976), Bibb Latané (who was the confederate in the euphoria condition) pointed out that the confederate exerted a great deal of pressure on the subject to join him in the "euphoric" behavior and that this constituted a relevant situational inducement. However, this raises even more problems in interpreting the data, since the subject's behavior would then reflect social conformity rather than an experienced emotion. One could also argue that once subjects had been induced to imitate the confederate, they would make judgments of their emotional state on the basis of observing their own behavior (Bem, 1972; Walters & Amoroso, 1967). According to this self-perception interpretation, the "appropriate cognition" is not the mood of the confederate but the behavior of oneself (which has been induced by the confederate).

the experimental manipulations and to reflect more precisely any differences in emotional behavior.

Hypotheses. The proposed modification and extension of the concept of cognitive search leads to some different predictions from those made by Schachter and Singer (1962). These competing hypotheses center on the questions of (a) whether the search is limited to, or goes beyond, the immediate situation and (b) whether the search is unbiased or biased. On the first question, Schachter and Singer imply that subjects will rely primarily on the information provided by the confederate to label their own unexplained arousal. In contrast, the current study predicts that subjects will also utilize information from both their past and present life. Thus, their emotional response will be more independent of the confederate's mood and less constrained by the immediate situation. On the second question, Schachter and Singer argue that subjects with inadequately explained arousal will engage in an unbiased search for a cognitive label. In contrast, the current study proposes that regardless of the confederate's display of affect, the cognitive search will be biased towards labels for negative emotions, since a state of unexplained arousal is initially anxiety provoking for adult subjects. For those not experiencing arousal or those given an appropriate explanation for their arousal, both theories agree on the absence of a search for cognitive labels of emotion—and thus an absence of an experience of emotion.

Method

bjects

Forty-eight undergraduate subjects (25 males and females) [2] were drawn from the introductory ychology course subject pool at Stanford Univer- y and paid for their participation in this experi- nt. All of the subjects had received scores of high pnotic susceptibility on the Harvard Group Scale Hypnotic Susceptibility (Shor & Orne, 1962).[3] irty-six of these subjects were randomly assigned training in hypnosis prior to the experiment, while subjects were randomly assigned to control ups that did not receive such training. The pnotic training, which averaged about 10 hours subject, was conducted in small groups and utilized a relaxed and permissive approach (see Zimbardo, Maslach, & Marshall, 1972). The training program sampled a wide variety of hypnotic experiences and included as criterion tests (a) the successful completion of posthypnotic suggestions with accompanying amnesia and (b) increased tolerance of ischemic pain (a measure developed by Lenox, 1970). Throughout the experimental session, trained and untrained subjects were treated identically except for a brief hypnotic induction for the trained subjects.

Procedure

The study consisted of two experiments that followed the same procedural format. The testing session was divided into two parts. The purpose of Part 1 was to establish a verbally conditioned arousal response to a specific cue and to demonstrate that this arousal involved substantial physiological changes. In part 2, this arousal was either elicited or not in the presence of a same-sex confederate who was behaving happily or angrily.

Part 1: Physiological arousal. Subjects were told that the study involved several types of individual assessment, including personality characteristics, physiological responsiveness, and learning performance. The subject was seated in a comfortable chair in an acoustic chamber, and the experimenter wiped clean the recording areas with alcohol. Beckman silver–silver chloride skin electrodes (17 mm in diameter) filled with Beckman Electrolyte Gel were attached by adhesive discs to the thenar eminence of the subject's nondominant palm and the volar surface of the two forearms. For the remainder of the period in the acoustic chamber, the subject's heart rate and galvanic skin response (GSR), as measured by skin resistance, were continuously recorded on a four-channel Offner Type R dynagraph with a #9857 cardiotachometer coupler and a #9842 galvanic skin response coupler. Standardized, prerecorded instructions were presented to the subjects over an intercom. The 5-min hypnotic induction procedure (which was also prerecorded) was eliminated for the unhypnotized control groups, who were merely told to relax for an equivalent period of time while baseline recordings were being made.

All subjects then heard the following verbal con-

---

[2] There were no significant differences between the responses of male and female subjects, so their data were combined.

[3] The selection of high-scoring subjects was done for practical reasons, since they can complete the hypnotic training and pass the criterion tests successfully in a shorter period of time than low-scoring subjects. There is no evidence that they differ in substantial ways from the general population as a whole, although they tend to exhibit a greater imagination (Hilgard, 1970).

ditioning instructions:

> In this session, the following reactions will occur whenever you see the word *start*. When you see the word *start*, your heart will beat faster, your breathing will increase, there will be a sinking feeling in your stomach, and your hands will get moist. You will feel all of these sensations as soon as you see the word *start*, and they will last until I say to you, "That's all for now." When I say, "That's all for now," you will return to your normal state and feel relaxed and good. However, when you see the word *start* and experience these reactions, you will not know why you are feeling the way you are, or remember my telling you anything about it.

The four symptoms that were cited in these instructions had all been reported by a pretest sample of volunteer subjects who had received an injection of epinephrine. Also, the relatively quick onset of the hypnotically induced arousal parallels the rapid "rush" that is typical of many people's response to epinephrine.

Subjects then described their current mood state, using the Nowlis Mood-Adjective Checklist (1963). A series of stimulus lights were then individually illuminated. The first of three neutral cues was a red light, the second was a white light, the third was labeled "stop," and the cue word "start" was given as the fourth signal. Following this arousal cue, subjects filled out a second Mood-Adjective Checklist, describing their feelings at that moment. Next, the termination cue was given for the disappearance of any experienced symptoms and for the return to a normal physiological state and a relaxed psychological state. There was a repetition of the posthypnotic suggestion, and then the hypnotic subjects were brought out of their hypnotic state.

Part 2: Emotional arousal. The subject returned to the first room and joined a waiting confederate (ostensibly another subject). In front of each of them was a memory drum, a learning test sheet, and a folder containing bogus test materials, such as Thematic Apperception Test (TAT) pictures, the Stroop color-word test, and a puzzle. The experimenters observed the behavior of the subject and the confederate through a one-way mirror and communicated via an intercom system.

A learning task was introduced in order to have a nonobvious way of presenting the arousal cue without either the confederate or observers being aware of the manipulation. Two 15-word lists were serially presented on the memory drum with a 4-sec exposure for each word and a 1-min recall test administered immediately after each list. The final word on the second list was either the arousal cue word ("start") or a neutral word ("speedy"). Subjects in the aroused and unhypnotized conditions saw the arousal cue word, whereas subjects in the unaroused conditions saw the neutral one. Next, the subject and confederate were told to look through the materials in the folder and were then left alone for 4 min. During this period, the confederate went through a prearranged series of either "angry" or "happy" behaviors.

The confederate's behavior followed the same general pattern in both the happy and angry conditions, although it differed in affect and the reasons given to justify that affect. During the first minute, the confederate worked with the folder materials and made a few comments about them, while in the second minute he or she talked about the experiment and school. During the third minute, the confederate directed some questions at the subject, and during the fourth minute, he or she finished examining the materials and "accidentally" dropped them on the floor. In both conditions, the confederate's affective level was mild at first (annoyance or positive interest) but became more intense over time (visible anger or exuberant laughter). Both confederates explained their emotional reaction in terms of impending tests and term papers. The angry confederate resented the time away from study caused by this uninteresting and slow-paced experiment. The happy confederate enjoyed the temporary distraction from these academic demands that the experiment made possible.

At the end of this emotional modeling period, the pair was separated and the subject was asked to complete two questionnaires, one of which focused on the subject's present physiological state and experienced emotion, while the other assessed reactions to the confederate. Next, the experimenter took the subject to an adjacent room to be interviewed by a clinical psychologist. This person was not affiliated with the experimenters and did not know what had taken place in the experiment. The clinician's task was to learn what the subject was feeling at the time, discover the cause of any expressed affect, and (if the subject was feeling at all distressed) make him or her feel better. The interviews were open-ended, were about 30 min in length, and were tape recorded for subsequent analysis. The purpose of these interviews was to assess whether the amnesia induction was powerful enough to resist skilled probing by someone unconnected with the hypnosis experience.

After the interview was completed, the experimenter entered the room and removed the hypnotic suggestions for the arousal response and the amnesia. Subjects were then questioned further about their reactions to the study. All subjects promised not to talk about the experiment until it was completed. Experimental silence was maintained, and all subjects later participated in an elaborate debriefing session conducted in a series of informal small group sessions. There was no evidence of any emotional carry-over beyond the immediate experimental setting.

Summary of Design

The first experiment consisted of a 2×3 factorial design. There were three levels of arousal—

aroused (hypnotized + arousal cue), unaroused (hypnotized + no arousal cue), and unhypnotized (not hypnotized + arousal cue). There were two levels of confederate emotion, happy and angry. Subjects who were trained in hypnosis were randomly assigned to the four hypnosis groups, and untrained subjects were randomly assigned to the two unhypnotized groups, all with six subjects per cell.

Because the results of this first experiment did not replicate those found by Schachter and Singer (1962) for the "euphoria" treatment, two additional experimental conditions were tested to discover the basis for these differing outcomes. In both conditions, subjects were given the arousal treatment (hypnotized + arousal cue) and were exposed only to the display of happy emotion by the confederate. One of these additional groups (informed arousal) remained aware of the basis for their arousal, since they were not given any suggestion for posthypnotic amnesia. This group parallels that of the informed treatment in the Schachter and Singer (1962) study. The other group (simplified arousal) was given only two physiological symptoms (heart beating faster, respiration increasing), in order to create a more purely neutral arousal. It was conceivable that the other two symptoms (sinking feeling in the stomach, moist hands), although found in reactions to epinephrine, too strongly suggested a negative arousal state. As before, subjects trained in hypnosis were randomly assigned to these two conditions, with six subjects per group.

Results

Physiological Arousal

There was a striking difference in physiological response between the hypnotized and unhypnotized conditions (see Table 1). For each subject, the heart rate record was divided into 5-sec intervals, and both the highest and mean heart rates for each interval were calculated. There were six such intervals following each stimulus, and their individual scores were combined into a total mean score for each stimulus. A repeated measures analysis of variance for the highest heart rate scores showed that there was a significant main effect of visual stimuli, such that the highest heart rate following the arousal cue was significantly higher, $F(1, 34) = 32.88$, $p < .001$,[4] than the highest heart rate following the neutral stimuli (which did not differ from each other). However, there was a significant interaction between stimuli and hypnosis condition, $F(1, 34) = 5.35$, $p < .05$. The F tests for the two hypnosis conditions reveal that this main effect was due entirely to the hypnotized subjects, $F(1, 34) = 37.17$, $p < .001$. The unhypnotized subjects did not show a significant increase in highest heart rate in response to the arousal cue ($F = 2.19$, *ns*). An analysis of mean heart rate revealed exactly the same pattern of results.

Table 1
Conditioned Physiological Arousal

Condition	n	Cue: Neutral	Cue: Arousal	M increase
		Heart rate[a]		
Hypnotized	24			
Mean		80.8	88.0	7.2
High		84.6	94.1	9.5
Unhypnotized	12			
Mean		76.6	78.8	2.2
High		80.6	83.8	3.2
		Galvanic skin response[b]		
Hypnotized	24	2.6	4.4	1.8
Unhypnotized	12	2.1	2.4	.3

[a] Beats per minute.
[b] Number per period.

A similar pattern of arousal was reflected in the total number of GSRs following each stimulus. There was a significantly greater number of GSRs following the arousal cue than following the neutral stimuli, $F(1, 34) = 20.87$, $p < .001$. However, there was also a significant interaction between stimuli and hypnosis condition, $F(1, 34) = 6.04$, $p < .025$, and the F tests for the two hypnosis conditions show that the main effect was due entirely to the hypnotized subjects, $F(1, 34) = 26.51$, $p < .001$, and not to the unhypnotized group ($F = .40$, *ns*). This difference in arousal response is even more striking given the fact that both hypnotized and unhypnotized subjects showed a more rapid recognition of the cue word (as evidenced by a shorter latency and higher amplitude of the first GSR following the cue).

Two Mood-Adjective Checklists were completed by each subject, one before the

[4] All reported p values are two-tailed.

arousal stimulus and one after. For purposes of analysis, the adjectives on this list were divided a priori into four groups—positive (e.g., affectionate, overjoyed); negative (e.g., angry, clutched-up); active (e.g., energetic, vigorous); and passive (e.g., calm, quiet). Changes in mood responses were assessed by computing separate mean change scores for each category. The changes scores for the hypnotized subjects indicate the marked shift in emotional state that resulted from their arousal. Their mood became significantly more negative, $t(23) = 7.22$, $p < .001$; less positive, $t(23) = 7.68$, $p < .001$; and less passive, $t(23) = 7.68$, $p < .001$. In contrast, the unhypnotized subjects showed no changes on the negative, active, or passive mood indexes, although there was a tendency for them to report less positive feelings, $t(11) = 2.07$, $p < .10$. Overall, the hypnotized subjects showed a more extreme shift on a composite of these four indexes of mood than did the unhypnotized subjects, $F(1, 34) = 17.14$, $p < .001$.

Because hypnotic subjects responded physiologically to the arousal cue in the acoustic sound chamber, it was assumed that they would respond similarly when they again saw this stimulus during the learning task. Unfortunately, we could not directly assess the validity of this assumption because telemetry equipment was not available to us at the time of the study. Nevertheless, two indirect measures of arousal do provide support for our reasoning. One was the subjects' recall performance on the learning task. Since the second of the two lists either did or did not contain the arousal cue, it was predicted that differences in subsequent arousal states might reduce cue utilization and affect the recall process (cf. Easterbrook, 1959). Although there was no between-groups difference in overall recall accuracy on either list (about 60% accuracy), subjects in the arousal groups made significantly fewer errors of commission ($M = .7$) than did the unaroused and unhypnotized groups ($M = 1.7$) on the second learning list, $t(34) = 2.54$, $p < .02$. It appears that subjects performing under the distracting influence of arousal wrote down only the words they had stored in readily accessible memory and then stopped, without making guesses about the remaining words.

The second indirect measure of arousal was part of the final questionnaire. Subjects were given a list of eight physiological symptoms and asked to indicate which (if any) they were now experiencing. The percentage of subjects who checked each symptom was calculated for each of the experimental conditions. From 92% to 100% of the subjects in the aroused conditions reported experiencing each of the four suggested symptoms, whereas only 26% to 52% of the unaroused and unhypnotized subjects reported having even a single one of the symptoms. The between-groups differences between these proportions for each of the four symptoms computed separately are all beyond the .02 level of significance.

Assessment of the success of the posthypnotic suggestion was provided by open-ended questions on the final questionnaire, as well as by the in-depth postexperimental interviews conducted by the clinical psychologist. Both sources of data clearly support the conclusion that these trained hypnotic subjects who were given an amnesia suggestion were unaware of the true cause of their experimentally induced arousal.

Perception of Confederate

A check on the confederate manipulation was derived from the mean score of three bipolar scale items of seven alternatives about the confederate's emotional behavior (sad–happy, angry–peaceful, friendly–unfriendly). These data were, without exception, in opposite directions for subjects exposed to the happy versus the angry confederate. Subjects in the happy condition evaluated the confederate's mood as significantly more positive ($M = +1.2$) than the neutral midpoint of zero, $t(17) = 4.94$, $p < .001$. In contrast, the treatment mean for subjects in the angry condition ($M = -1.5$) was significantly more negative than zero, $t(17) = 7.45$, $p < .001$. The difference between the two confederate conditions is highly significant, $F(1, 30) = 83.16$, $p < .001$.

Table 2
Expressed and Experienced Emotion

Condition[a]	*M* no. of sociable behaviors: Verbal, Confederate: Happy	Verbal, Confederate: Angry	Nonverbal, Confederate: Happy	Nonverbal, Confederate: Angry	Self-reported emotion (happy minus angry)[b], Confederate: Happy	Self-reported emotion (happy minus angry)[b], Confederate: Angry
Aroused	4.4	.7	10.5	.8	−1.0	−1.2
Unaroused	4.3	.4	8.2	.4	+1.8	+2.5
Unhypnotized	2.5	.9	6.0	1.6	+.8	+.4

[a] $n = 12$ for each condition (6 per cell).
[b] Mean scores are based on the Schachter and Singer (1962) index, which ranges from +4 (most happy) to −4 (most angry).

Overt Emotional Behavior

Measures of subjects' behavior were coded by two judges (interrater reliability was at least .80) who observed each subject during the 4-min emotional modeling period. For each of the 4 min, the judges coded the subject's verbal and nonverbal responses. Some of these responses were sociable ones (i.e., smiling, nodding, agreeing with the confederate), while others were not (i.e., disagreeing, ignoring, frowning). The overt behavior of the subjects varied as a function of the confederate's mood, but seemed unaffected by experienced arousal. As shown in Table 2, subjects in the presence of a happy confederate exhibited a significantly higher amount of sociable behaviors than did subjects who were with the angry confederate. This was true for both verbal behaviors, $F(1, 30) = 20.97$, $p < .001$, and nonverbal behaviors, $F(1, 30) = 44.80$, $p < .001$.

Reported Emotional Experience

Two scale items tapping extent of experienced anger and happiness were adopted from Schachter and Singer (1962). To the items, "How irritated, angry, or annoyed would you say you feel at present?" and "How good or happy would you say you feel at present?" subjects responded on 5-point scales ranging from 0 (not at all) to 4 (extremely). Analyses were made of the difference scores (happy–angry), as well as of the happy and the angry ratings separately. A broader portrait of the emotional state of each subject comes from eight additional bipolar scales (e.g., happy–sad, angry–peaceful, confident–apprehensive) with seven response alternatives. These responses were analyzed separately and then combined into a single index of emotional response ranging from extremely negative to extremely positive.

Although the confederate's mood had a demonstrable effect on the subjects' overt behavior, it did not exert the expected influence on their reported emotional state. For the Schachter and Singer index, there was a significant main effect of arousal, $F(2, 30) = 5.35$, $p < .025$. A comparison of means (shown in Table 2) revealed that this was due to the more negative scores of the aroused condition ($p < .005$).[5] Analyses of each of the happy and the angry ratings separately revealed the same significant difference in emotional reactions between the aroused subjects and the other subjects without arousal.

The mean ratings on the eight-scale index of emotional response are shown in Figure 1. Again, there was a significant main effect of arousal condition, $F(2, 30) = 15.94$, $p < .001$, and a comparison of means revealed that the aroused groups reported a significantly more negative emotional state ($p < .001$) than both the unaroused and the unhypnotized groups (which did not differ from

[5] All comparisons between means following a finding of a significant F ratio were performed using Tukey's honestly significant difference test.

each other). This same pattern of results was found for each individual scale. On the overall index, the aroused conditions reported negative emotions that were significantly different from the neutral midpoint of zero, $t(11) = 6.47$, $p < .001$. While the unhypnotized groups were neutral in their emotions (their mean score did not differ significantly from zero), the unaroused subjects rated themselves as feeling a positive emotional state, $t(11) = 2.32$, $p < .05$. Apparently, the unaroused subjects were responding appropriately to the posthypnotic suggestion that they would feel relaxed and good when they came out of hypnosis. The positive feelings of the unaroused group offer a dramatic comparison level for the negative reaction reported by the aroused subjects (despite the same posthypnotic suggestion to feel good and identical prior to hypnotic training). This positive response is particularly noticeable in the unaroused angry condition, although it did not differ statistically from the unaroused happy group. This finding may be due to a contrast effect (the unaroused subjects may have felt particularly good in comparison to the negative reactions of the angry confederate) or to a self-persuasion effect (many of the unaroused subjects made reference to their positive feelings in an attempt to cheer up the angry confederate).

Attribution of Causality

On the final questionnaire, subjects were asked if they knew why they felt as they did and, if so, to state the reason. These responses were scored by judges who were blind to both the purpose of the experiment and the condition of the subject. The aroused subjects always reported reasons for a *negative* emotional state, regardless of the mood expressed by the confederate. Some subjects said they were nervous about performing well on the experimental tests, others said they were feeling tense about upcoming final exams, and still others labeled their arousal as irritation generated by the confederate's "talkativeness." In only a single instance did a subject explicitly state that the confederate's mood determined his own: "The other guy was really hassled—it rubbed off."

The possibility that the posthypnotic suggestion had the effect of blocking any search for an explanation of the experienced arousal, rather than merely preventing the "correct" explanation, does not appear tenable. The number of aroused subjects (67%) who thought they knew the cause of their current feelings was as high as that of the unaroused and unhypnotized groups. However, at no time did the aroused subjects state that the

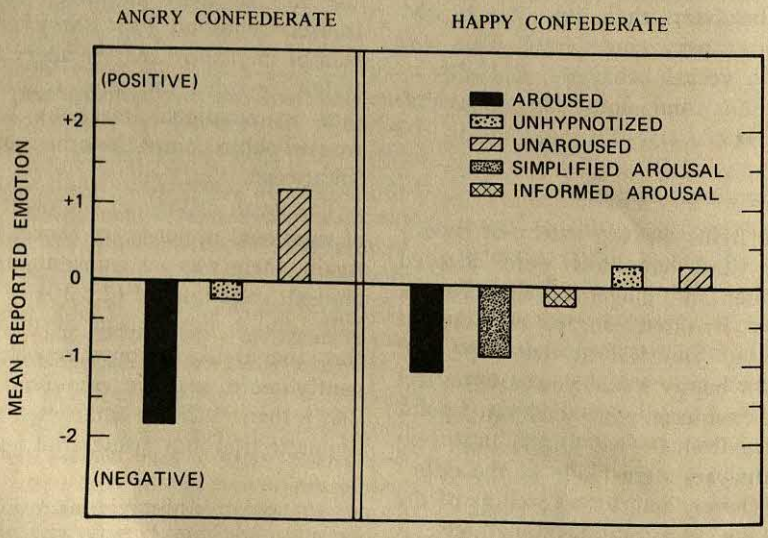


Figure 1. Self-reported emotion (mean self-ratings on an eight-scale index of emotion).

reason was related to the experimentally induced cause of posthypnotic suggestions.

Comparison of Arousal Control Groups

Both of the arousal control groups (informed arousal and simplified arousal) displayed the same pattern of physiological arousal as the happy aroused subjects in the first experiment. They showed the same increase in heart rate, similar changes on the Mood-Adjective Checklist, the same pattern of recall errors, but not a similar change in GSR.[6] On the self-report of physiological symptoms, all of the informed aroused subjects reported experiencing the four suggested arousal symptoms. The simplified arousal subjects reported increased breathing and heart rate, as expected. Surprisingly, however, all of them also reported sweaty palms, and two thirds reported a sinking feeling in the stomach. Even though they had not received any suggestion for these last two symptoms, they generated them spontaneously as correlates of changes in respiration and heart rate. The perception of the confederate by the two control groups was the same as that of the happy aroused group, and they showed the same pattern of sociable behaviors (both verbal and nonverbal) in their interaction with the confederate.

Overall, the two arousal control groups were similar to the happy aroused group in terms of physiological arousal, veridical perception of the confederate, and overt behavior. However, they differed in experienced emotional state. The informed arousal subjects reported a neutral set of feelings (which did not differ from the zero midpoint), as opposed to the negative feelings expressed by the happy aroused group, $t(10) = 1.90$, $p < .10$. As expected, they all stated that the reason for their experienced arousal was the hypnotic suggestion. In contrast, the simplified arousal subjects reported the same negative emotion as the happy aroused group (on each of the self-report measures) and gave the same type of causal explanations. As can be seen in Figure 1, their mean rating on the emotion index was significantly more negative than the neutral midpoint of zero, $t(5) = 4.15$, $p < .01$. In other words, despite a posthypnotic suggestion of only the two most basic physiological concomitants of arousal, these subjects who were unaware of the "appropriate" explanation for their arousal did *not* label their emotion as "happy" in accordance with the available cognitions provided by the confederate's behavior. Rather, they experienced a negative emotional state.

Discussion and Conclusions

The current study was designed as a modified replication and extension of Schachter and Singer's (1962) two-factor theory of emotion. The necessary conditions of strong physiological arousal, lack of an explanation for that arousal, and varied emotional–situational cognitions were all achieved in an experimental setting that controlled for many of the methodological problems of the original study. However, the results reveal a remarkably consistent and coherent pattern of emotional response (especially noteworthy given the small sample size) that is at variance with the interaction pattern predicted by Schachter and Singer. In all cases, subjects with unexplained arousal reported negative emotions, irrespective of the confederate's mood.[7] This finding stands in opposi-

[6] The only inconsistency in the comparison of both the simplified arousal and informed arousal groups with the aroused condition tested initially is the failure of the former to display the same expected changes in GSR. While this may reflect a lower level of physiological arousal, this interpretation is weakened by the fact that these subjects did show the same pattern and degree of change in heart rate and reported awareness of physiological symptoms. A more probable explanation is that since these two groups were run several months after the original conditions, the source of invalidity might have been instrument variation, through unmonitored changes in the electrodes or GSR coupler.

[7] It might be argued that the minimal impact of the confederate on the subject's emotional state was due to the fact that subjects first experienced the physiological arousal while alone (in the acoustic chamber), prior to being aroused for a second time in the presence of the confederate. However, the same pattern of consistently negative affect occurred during the pretesting for the experiment, when subjects received the hypnotic arousal induc-

tion to the position that unexplained physiological arousal is a neutral, energizing variable that does not contribute directly to the qualitative labeling of the emotion. Whereas Schachter and Singer postulated that the lack of explanation for arousal would motivate people to search in an unbiased way for an appropriate cognition, the present study suggests that it not only motivates but biases the search, since it tints perception of that arousal with negative affect. The fact that Marshall and Zimbardo (1979) also found self-reports of negative affect among subjects who lacked an explanation for their strong physiological arousal (and focused their attention on that discordant state) lends additional support to this alternative view.[8]

What might account for this negative emotional bias? One possibility is that since there are more terms referring to negative emotions than to positive ones (e.g., Izard, 1971), people might report negative emotions more often on a purely statistical basis alone. Second, if noticeable physical arousal is indeed a more frequent correlate of negative emotional experiences in everyday life, then the negative affect reported by subjects in the arousal conditions could be regarded as a logical learned response based on prior experiences. Third, as proposed earlier, it may be that unexplained arousal is always characterized by negative affect because it is akin to a state of free-floating anxiety.

The results of the current study suggest that a negative emotional bias will have implications for the person's search for causal information. Subjects who reported negative feelings almost always provided an explanation for them, such as "I'm upset because I don't do well on tests" or "I'm annoyed because this guy keeps joking all the time." In contrast, subjects reporting positive or neutral emotions rarely indicated their cause and gave, instead, such nonexplanatory statements as "I don't know, it's just the way I normally feel."[9] In line with the proposition by Jones and Davis (1965), it seems as if negative experiences are more likely to motivate a search for causal information. People want to know why they feel upset, frightened, or angry, perhaps because they want to control (and thus reduce) the future occurrence of such experiences. On the other hand, they are less motivated to know why they feel happy, pleased, or content, perhaps because they consider such feelings to be their normal baseline condition as opposed to a "deviant" response that demands explanation.

Social Influences on Emotion

An important conclusion to be derived from the current research is that experienced and expressed emotions are not perfectly correlated. Although subjects with unexplained arousal displayed sociable behaviors while

---

tion in the presence of the comparison person. Furthermore, in subsequent research (Zimbardo & Maslach, Note 1), subjects received the hypnotic arousal induction in the presence of at least one other comparison person, and again there was the same pattern of negative affect. The fact that the data from these two studies parallel the results in the current experiment suggests that the procedure surrounding the initial induction (i.e., subject alone vs. subject with confederate) was not a crucial element in the subject's search for an emotional explanation.

[8] It is important to note that Marshall and Zimbardo's (1979) research provides a connecting link between the current study (which found a significant negative emotional bias) and the Schachter and Singer (1962) study (which found no systematic biases from the baseline of the placebo control condition). In the conditions that were an exact replication of Schachter and Singer's experiment, Marshall and Zimbardo "replicated" the original finding of no difference between epinephrine and placebo subjects. Some of the physiological data suggested that the induced arousal was not very strong or salient for many of the subjects, and this may explain why their responses did not differ from the unaroused placebo controls. However, when Marshall and Zimbardo did make the physiological arousal more salient for the subjects, they reported a negative emotional state. The hypnotic manipulation used in the current study also made the arousal state salient for the subjects, and their consistently negative emotional responses replicate those found by Marshall and Zimbardo in their "salient arousal" conditions.

[9] In a personal communication (1976), Gerald Blum reported that subjects in his research on affective arousal (Blum, 1972) showed a similar pattern of responses, giving more causal statements about their negative feelings and fewer about their positive feelings.

in the presence of the happy confederate, their self-reported mood state was not positive. Their comments imply that their overt behavior was more influenced by social contingencies (i.e., norms of social appropriateness) than was their more private feeling state. Such deliberate "management" of one's emotional expression is a fairly common experience for people—as when, for example, one laughs and acts happily with friends even though feeling depressed or hurt by some remark. To assume that one's behavior and self-reported mood are equivalent indexes of emotion is to ignore functional differences between these two response classes.

This issue is relevant for interpreting the results of Schachter's follow-up study on the effects of variations in unexplained arousal on amusement (Schachter & Wheeler, 1962). In this experiment, subjects received an injection of either epinephrine, a placebo, or chlorpromazine (a depressant drug) and were led to believe that there would be no side effects. They then watched an excerpt from a slapstick movie. Whether or not subjects labeled their arousal state in terms of the available cognition of "amusement" was measured by self-reported evaluations of the film and observation of their amusement behavior. Even with large numbers of subjects, there were no significant differences between the three conditions on the self-report measures. The epinephrine subjects and the unaroused placebo controls did not differ on the overall behavioral index (although there was a significant difference on the number of laughs, as opposed to smiles and grins). The major finding was the depressed behavioral response of the chlorpromazine subjects, which may have been a direct somatic effect of the tranquilizer. This lack of correspondence between amusement behaviors and self-reported evaluations of amusement, the lack of variation across treatments in self-report, and the minimal differences between the epinephrine subjects and the placebo controls obviate any serious challenge of the Schachter and Wheeler (1962) study to the results or line of reasoning of the present study.

Another conclusion to be drawn from the current research is that people utilize far more sources of information in generating their emotional explanations than was initially suggested by Schachter and Singer's (1962) analysis. The aroused subjects in the current study perceived correctly the emotion being expressed by the confederate, but they did not necessarily use this emotion to label or explain their own state. Their emotional explanations appeared to be a complex function of their past experience, current life situation, and the immediate situational circumstances, rather than just the mood of the confederate. Not only does this finding extend the notion of "cognitive search" but it also points to the necessity of reconceptualizing the type of information that is presented by the confederate.

Schachter and Singer's model is based on a theory of social comparison (Festinger, 1954) that argues that self-evaluation is often accomplished by comparing oneself with other people. This is particularly true when one has doubts about the appropriateness or correctness of one's feelings, beliefs, or behavior. According to this viewpoint, what is provided by the confederate (the social comparison referent) is *normative* information. The confederate models the "appropriate" way to label the arousal, and the subject adopts a similar response in order to appear (both to self and others) as normal. While normative information indicates how one should feel and act, it does not necessarily indicate *why* one should feel that way. In other words, normative information is not always *causal* information as well. Schachter and Singer argue that the lack of causal information about one's aroused state is what motivates the search for an emotional cognition. However, in postulating that this cognition is obtained via social comparison processes, they are suggesting that the ultimate solution to the search is normative information. If only normative information were needed, then the confederate's reaction would provide a sufficient basis for the subject to evaluate his or her emotion. However, if the subject wants information as to *why* the arousal response is occurring, then the confederate's overt reaction becomes evidence only for the fact that an unknown cause is generating a strong reaction in others as well.

Therefore, the overt reaction is not itself the cause. The subject will look beyond the confederate's reaction for such causal information and may arrive at an emotional cognition that differs from the emotion being expressed by the confederate.[10]

As mentioned earlier, a state of unexplained physiological arousal may not be an appropriate paradigm for studying common emotional experiences. However, the negative emotional bias associated with unexplained arousal suggests that it may be a useful paradigm for studying emotional pathology, particularly if it is viewed as just one of a class of anomalous personal and social experiences. Such anomalous reactions are anxiety provoking because they represent a threat to self-control, and they will usually generate a search for suitable causal linkages. A rational, information-processing model might assume that people engage in a scientific (unbiased) search, in which they seek out and consider all possible evidence in an objective way. However, it may be more correct to assume that they are biased towards certain types of explanations as a result of their past history of reinforcement for particular classes of explanation or ways of thinking about new experiences. We are currently conducting exploratory research on this biasing process, again using hypnosis as a methodological tool to create unexplained arousal.

Furthermore, following from an earlier argument, the study of the process of searching for causal explanations would benefit from an experimental analysis that focused on children's reactions to novel or anomalous events (e.g., first headache or nightmare). This line of work suggests that emotional pathology may not be an impairment of one's cognitive abilities, but a normal cognitive process used to explain an unusual state or event. It extends and expands Maher's (1974) hypothesis about the perceptual genesis of delusional thinking and may provide a unique social psychological perspective on the dynamics of abnormal behavior.

[10] In a personal communication (1977), Stanley Schachter argued that the discrepancy in results between the current research and the Schachter and Singer (1962) study may be due to differences in the timing of the confederate's behavior. In their study, an attempt was made to have the confederate's routine precede or occur simultaneously with the subject's experienced arousal. In the current research, the confederate's behavior followed the onset of arousal by about one minute (the duration of the recall task). According to Schachter, subjects in the current experiment would not have used the confederate's emotion to label their own, since it occurred after the onset of arousal and thus does not qualify as a prior cause. However, this argument has several flaws. First of all, it continues to confuse whether normative or causal information is being provided by the confederate. If it is normative information (as Schachter and Singer's theory implies), then the confederate's behavior is a clue to a possibly shared prior cause but is *not* the cause itself. The confederate's behavior need only occur at about the same time as the subject's arousal to suggest a common, proximal causal event; it is not necessarily required to occur prior to it. Second, Schachter's argument ignores the fact that Marshall and Zimbardo (1979) used the same timing procedure as in the Schachter and Singer study and obtained results supportive of the current research. Third, even if Schachter's argument were correct, it would not explain the highly consistent negative bias of the aroused subjects. If they were indeed forced to interpret their arousal in terms of events preceding the confederate's behavior, then we would expect to find far more variability in the affective quality of their emotional explanations than was actually the case.

Reference Note

1. Zimbardo, P. G., & Maslach, C. *Biased searches for causal explanations of experienced discontinuities.* Unpublished manuscript, 1979.

References

Bem, D. J. Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6). New York: Academic Press, 1972.

Blum, G. S. Hypnotic programming techniques in psychological experiments. In E. Fromm & R. E. Shor (Eds.), *Hypnosis: Research developments and perspectives.* Chicago: Aldine-Atherton, 1972.

Cooper, L. M. Hypnotic amnesia. In E. Fromm & R. E. Shor (Eds.), *Hypnosis: Research developments and perspectives.* Chicago: Aldine-Atherton, 1972.

Easterbrook, J. A. The effect of emotion on cue utilization and the organization of behavior. *Psychological Review,* 1959, *66,* 183–201.

Evans, F. J., & Kihlstrom, J. F. Posthypnotic amnesia as disrupted retrieval. *Journal of Abnormal Psychology,* 1973, *82,* 317–323.

Festinger, L. A theory of social comparison processes. *Human Relations,* 1954, *7,* 114–140.

Hilgard, J. *Personality and hypnosis: A study of imaginative involvement*. Chicago: University of Chicago Press, 1970.

Izard, C. E. *The face of emotion*. New York: Appleton-Century-Crofts, 1971.

Jones, E. E., & Davis, K. E. From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2). New York: Academic Press, 1965.

Lazarus, R. S. Emotions and adaptation: Conceptual and empirical relations. In W. J. Arnold (Ed.), *Nebraska Symposium on Motivation* (Vol. 16). Lincoln: University of Nebraska Press, 1968.

Lenox, J. R. Effect of hypnotic analgesia on verbal report and cardiovascular responses to ischemic pain. *Journal of Abnormal Psychology*, 1970, *75*, 199–206.

Leventhal, H. Emotions: A basic problem for social psychology. In C. Nemeth (Ed.), *Social psychology: Classic and contemporary integrations*. Chicago: Rand McNally, 1974.

London, H., & Nisbett, R. E. (Eds.). *Thought and feeling: Cognitive alteration of feeling states*. Chicago: Aldine, 1974.

Maher, B. Delusional thinking and cognitive disorder. In H. London & R. E. Nisbett (Eds.), *Thought and feeling: Cognitive alteration of feeling states*. Chicago: Aldine, 1974.

Mandler, G. *Mind and emotion*. New York: Wiley, 1975.

Marshall, G., & Zimbardo, P. G. Affective consequences of inadequately explained physiological arousal. *Journal of Personality and Social Psychology*, 1979, *37*, 970–988.

Maslach, C. The emotional consequences of arousal without reason. In C. E. Izard (Ed.), *Emotions in personality and psychopathology*. New York: Plenum Press, 1979.

Nowlis, V. The concept of mood. In S. M. Farber & R. H. L. Wilson (Eds.), *Conflict and creativity*. New York: McGraw-Hill, 1963.

Plutchik, R., & Ax, A. F. A critique of "determinants of emotional state" by Schachter and Singer (1962). *Psychophysiology*, 1967, *4*, 79–82.

Sarbin, T. R., & Slagle, R. W. Hypnosis and psychophysiological outcomes. In E. Fromm & R. E. Shor (Eds.), *Hypnosis: Research developments and perspectives*. Chicago: Aldine-Atherton, 1972.

Schachter, S. *Emotion, obesity, and crime*. New York: Academic Press, 1971.

Schachter, S., & Singer, J. E. Cognitive, social, and physiological determinants of emotional state. *Psychological Review*, 1962, *69*, 379–399.

Schachter, S., & Singer, J. E. Comments on the Maslach and Marshall–Zimbardo experiments. *Journal of Personality and Social Psychology*, 1979, *37*, 989–995.

Schachter, S., & Wheeler, L. Epinephrine, chlorpromazine, and amusement. *Journal of Abnormal and Social Psychology*, 1962, *65*, 121–128.

Shor, R. E., & Orne, E. C. *The Harvard Group Scale of Hypnotic Susceptibility, Form A*. Palo Alto, Calif.: Consulting Psychologists Press, 1962.

Walster, E. Passionate love. In B. I. Murstein (Ed.), *Theories of attraction and love*. New York: Springer Press, 1971.

Walters, R. H., & Amoroso, D. M. Cognitive and emotional determinants of the occurrence of imitative behaviour. *British Journal of Social and Clinical Psychology*, 1967, *6*, 174–185.

Zimbardo, P. G., Maslach, C., & Marshall, G. Hypnosis and the psychology of cognitive and behavioral control. In E. Fromm & R. E. Shor (Eds.), *Hypnosis: Research developments and perspectives*. Chicago: Aldine-Atherton, 1972.

Received November 18, 1977 ■

Postscript

In their attempt to dismiss the relevance of the data generated by the present paradigm for understanding the role of unexplained arousal in emotion, Schachter and Singer (1979) focus solely on the issue of timing. The points I will address briefly in this rejoinder are designed to (a) clarify the precise timing sequence in the present study; (b) question the generalizability of a model of emotion that depends on such a restricted temporal synchrony between arousal and social cognitions; (c) highlight the importance of the subject's awareness of a discontinuous arousal state; and (d) reaffirm the distinction between two alternative processes a person may engage in when "bodily symptoms search for an explanation"—the normative search and the causal search.

The Timing Sequence and Awareness

In my study, the arousal cue was followed by a 1-min recall test and a 6-sec experimenter statement that "the next part of the study will take a few minutes to set up. In the meantime, you might want to look through the folder materials." The very first minute of the confederate's routine was *not* affectively neutral, since it included smiles, humming, and verbal pleasantries in the happy condition and frowns and verbal expressions of annoyance in the anger condition. Thus, between 1 min, 6 sec and no more than 1 min, 30 sec elapsed from arousal cue to situational manipulation of "happy–angry" social cognition. I submit that this temporal sequence is adequate to allow the confederate's behavior to serve as a reasonable social context for the subject's unexplained arousal.

Indeed, if it is not, then neither is the timing in the original Schachter–Singer (1962) anger condition. The stooge's display of affect in that study might well have occurred *after* most subjects started experiencing their epinephrine-induced reactions. About 1 min is reported to have elapsed before subject and stooge were brought together. Then they had to take some additional minutes to complete the first of five pages of questionnaires, which "start off innocently by requesting face sheet information. . . . The stooge . . . paces his own answers so that at all times subject and stooge are working on the same question" (p. 385). This procedure allows loose subject-controlled temporal variations where Schachter and Singer demand tight experimenter control. Since the onset latency of epinephrine symptoms varies across subjects, some "anger" subjects would have been aroused several minutes before the first anger cognitions were available, some at about the same time, and some slightly after. The timing sequence is less well regulated than in my study and violates Schachter and Singer's condition that social cognitions must precede unexplained arousal in order to be utilized as cues to emotional state. Furthermore, if precise timing is as critical as Schachter and Singer state in their comments on my research, then it must be individually determined for each subject and not left to "on the average" variations in approximate time before symptom onset and approximate time before stooge's display of affect.

Although I address the timing issue directly in Footnotes 7 and 10, Schachter and Singer (1979) have deliberately chosen to dismiss it with their cavalier "footnotes notwithstanding" (p. 990). In rebuttal to their description of the procedures experienced by subjects in the present study and to their comments in their Footnote 1, I assert that the pretest I conducted did *not* involve the critical timing issue raised by Schachter and Singer. Subjects were induced to experience unexplained arousal in the presence of another subject who was already expressing anger or happiness. Similarly, in the unpublished study by Zimbardo and Maslach (Note 1), the timing criticism is again *not* an issue, and the consequence of experiencing unexplained arousal is still negative.

If, as noted earlier, 1-minute differences are so critical, then the generalizability of this model of emotion is severely restricted. It is perception of one's arousal state that motivates a search for a causal explanation. The psychologically relevant variable is the timing of the subject's *awareness* of his of her arousal and available cognitions, and not the time when the drug was injected, when an arousal cue was given, or even when measurable physiological symptoms first appeared. We know that awareness varies considerably across subjects in terms of individual differences in self-monitoring and sensitivity to autonomic discontinuities. The proximal social cognitions that immediately precede this perception of arousal may be the most salient candidates for a causal explanation, but they are not the only ones. Schachter and Singer (1962) manipulated one class of attributes for the experienced arousal, namely a physically present other person. As befits social psychologists, they chose social cognitions as the reasonable context for unexplained arousal. But the subjects could have chosen otherwise, from events in their past or anticipated future, from nonsocial cognitions in the current physical environment, or from health and biological sources. The immediacy of the social cognition increases its probability of being used as the basis for transforming physiological arousal into a psychological emotional state—but it is not a sufficient condition for that to occur. Furthermore, there is no reason to assume that the search process is usually completed in a matter of minutes. Provisional explanations may be generated, tried out, rejected for others, changed by new feedback, and so on. Yes, the confederate in the Schachter–Singer paradigm will be of little effect if he starts after the subject has already chosen a mood label (as Footnote 1 of their 1979 Comments states), but that choice may require some time before it is final, especially in novel situations.

The Normative–Causal Distinction

Schachter and Singer's (1979) Footnote 1 credits me with a "new" distinction between the normative and causal aspects of the cognitive search to account for unexplained arousal. (I will accept that as a compliment and not interpret "new" to mean untenable because not yet empirically validated.) More than the timing issue is involved in recognizing the way these different psychological processes operate when an individual tries to make sense of a bodily state of insufficiently explained arousal.

The normative search entails seeking cognitions that one's reactions are situationally appropriate and socially acceptable. The person wants to know how to behave in an unfamiliar setting or to discover what range of responses is expected. In the process of looking for this "how," the person often comes up with a suitable "what" as well. "What" is the label, or socially agreed upon term, that normalizes an otherwise idiosyncratic and ambiguous reaction.

The causal search entails seeking cognitions that provide information about likely causes and consequences of one's perceived arousal. This search for "why" starts with an effect to be explained in terms of reasonable causal antecedents. Such an analysis need not necessarily involve any social cognitions (although these, of course, are central to any normative search process).

The search for a causal explanation, then, is an attempt to establish effect–cause linkages that help make sense of or bring rationality out of uncertainty. On the other hand, the search for an emotion label and for the "right" way to react may induce emotional contagion merely as a compliance phenomenon.

Schachter has operated primarily out of a social comparison framework since his early work on affiliation and anxiety. Because of this theoretical orientation, he either emphasizes the normative search or fails to acknowledge adequately the causal alternative. When subjects in novel or uncertain situations seek to discover how to behave and what to feel, the confederate's affective display is of little relevance to explaining *why* the subject is suddenly aroused. It serves only a labeling function ("I don't know why, but I am feeling angry, or acting happy," etc.). Or it serves to normalize strange reactions ("I don't know why I'm reacting like this, but at least I'm not the only one").

If the search is normative, then the timing of subject arousal and confederate's behavior should be of little consequence as long as both occur in the same setting (so that the emotional state is perceived as shared), the other person is an acceptable social comparison referent, and there is no other readily available label for the arousal. Timing should be of importance in the causal search only to the extent that the relative contiguity of the confederate's response narrows the search to common elements of the situation that might have caused both it and the subject's own internal reaction.

In conclusion, I am willing to accept the substance of Schachter and Singer's characterization of what my study "proves," when stated in the following form: When subjects are aroused for no apparent reason, they report themselves to be in a "lousy mood" regardless of whether a confederate has been acting happy or angry in their presence and regardless of whether they themselves have displayed happy or angry behaviors. They do so more than do unaroused controls or those who are aroused and know the cause. This is indeed the major finding of my study: that strong unexplained arousal *per se* is typically perceived as a negative state by adults and not as an undifferentiated, affectively neutral state, as the Schachter–Singer model proposes. At this point, I assume further that *both* the search for a causal explanation and the normative search for an appropriate label will be biased in negative directions as a result of prior experiences in which perception of strong arousal is more often associated with aversive events than with pleasant ones.

Journal of Personality and Social Psychology
1979, Vol. 37, No. 6, 970-988

Affective Consequences of Inadequately Explained Physiological Arousal

Gary D. Marshall and Philip G. Zimbardo
Stanford Medical School
Stanford University

This research reexamined Schachter and Singer's reported demonstration of cognitively influenced physiological determinants in the experience of emotion. In the present study, reports of affective state by placebo-injected and epinephrine-injected subjects, who were misinformed about possible somatic effects and exposed to a euphoric confederate, did not differ. Closer inspection of the Schachter and Singer results also indicates a failure of inadequately explained epinephrine-produced physiological arousal to enhance affective susceptibility to the social environment. Additional evidence provided by the present study reveals a tendency for inadequately explained epinephrine-induced physiological arousal to produce negatively toned reports of affect. The results for additional experimental conditions indicate that (a) manipulations to heighten the experience of physiological arousal produced greater negative affect; (b) similarly, expectation of epinephrine side effects produced negative affect; and (c) the affective tone of the social environment had minimal effect on arousal subjects. These negative affective reactions occurred in a social context where a confederate was perceived to be happy.

The theorizing and research on emotions by Schachter and Singer (1962) must be counted among the primary forces responsible for redirecting the field of social psychology toward a more cognitive orientation (Mandler, 1966, 1975). Their classic study strongly emphasized the important role played by cognitive influences in determining the role of physiological arousal in affective experience.

The Schachter and Singer study examined the contribution of both cognitive and physiological determinants by using a distinctively social psychological research strategy in which both factors were manipulated independently. Physiological arousal was controlled by providing injections of either epinephrine or a placebo. Cognitive factors regarding the arousal were controlled by providing accurate or inaccurate information regarding symptoms. In addition, the social situation was experimentally manipulated by means of a confederate who behaved either euphorically or angrily.

The experience of emotion in this situation was reportedly shown to depend on the interaction of (a) the presence of physiological arousal (via epinephrine injection), (b) the lack of an adequate explanation for this arousal, and (c) the particular affective information provided by the social environment.

This article is based on the first author's doctoral dissertation. The second author served as chairperson of the dissertation committee. The research was supported by Office of Naval Research Contract N000 14-67-A-0112-0041 to Philip Zimbardo. We want to thank Jan Hubbard, Deborah Cowley, David O'Brien, and Jeff Allen for working as research assistants, and Howard Buckley for serving as the physician in the experiment. We also want to thank James McClenahan for making the Student Health facilities available for the research and Carl Thoresen for making available the telemetry equipment. We are grateful to Daryl Bem, Edward Smith, and Karl Pribram for their helpful advice as members of the dissertation committee. We would like to thank Christina Maslach for her helpful comments on this manuscript.

Requests for reprints should be sent to Gary D. Marshall, Laboratory for the Study of Behavioral Medicine, Stanford Medical School, 730 Welch Road, Suite 234, Stanford, California 94304.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3706-0970$00.75

It was intriguing to discover the reported ease with which the same condition of physiological arousal could be steered in as seemingly opposite affective directions as happiness or anger as a result of these cognitive manipulations. The Schachter and Singer report thus constituted a strong argument for a common physiological substrate for different emotions (Fehr & Stern, 1970). It also suggested the existence of an "emotional plasticity" that had apparent therapeutic significance (Nisbett & Schachter, 1966; Ross, Rodin, & Zimbardo, 1969; Storms & Nisbett, 1970, among others).

Despite its importance, there have been no replications of the Schachter and Singer (1962) study reported in the literature. However, several critiques of the study have appeared (Averill & Opton, 1968; Lazarus, 1967; Leventhal, 1974; Plutchik & Ax, 1967; Shapiro & Schwartz, 1970; Stein, 1967). These critiques have questioned the adequacy of some of Schachter and Singer's basic conclusions, and, given the widespread citation of the study, the importance of a direct attempt at replication is obvious.

The present research design was limited to the "euphoric social cognition" condition, because of constraints imposed by the Stanford University Medical School Human Subjects Committee.[1] Furthermore, we chose to concentrate on the information condition that produced the most euphoric outcome for the physiologically aroused subjects in the Schachter and Singer study—the one in which subjects were actively *misinformed* regarding the somatic effects of the epinephrine injection. This was the condition in which the greatest emotional plasticity was reportedly demonstrated. Schachter and Singer actually found only weak treatment outcomes on affect for the physiological arousal manipulation. In the present study, a number of procedural improvements and operational refinements were introduced to provide a better test of the affective consequences of inadequately explained arousal. These changes included forewarning subjects that they were to receive an injection, introducing tasks to support the cover story, obtaining pretest measures in order to directly assess *change* in affect as a result of the manipulations, and continuous physiological monitoring, including additional self-report measures (see Marshall, 1976, for further details).

To further explore the arousal–affect relationship, four additional comparison groups were tested. These were introduced to assess the effects of heightened physiological arousal, the influence of the confederate's behavior, and the effects of information about the likely somatic effects of the injection. The fundamental comparison for assessing inadequately explained physiological arousal, however, is that between the basic placebo subjects and the basic epinephrine-injected subjects (who were misinformed about what to expect from the injection). Both groups were exposed to a euphoric confederate.

Method

Overview

The basic design of the study involved manipulation of physiological arousal via injections of epinephrine or placebo, misinformation about the drug's anticipated somatic effects, and exposure to a confederate who behaved euphorically. In the additional conditions, the following factors were systematically altered: the method of determining epinephrine dosage, the focus of the subject's attention, the confederate's behavior, and the type of somatic effects expected. Heart rate was continuously monitored by telemetry. Pre- and posttreatment self-reports of affect were collected. The subject's behavior while in the presence of the confederate was both rated directly by observers and videotaped for subsequent scoring.

As volunteers for an alleged study of the effects of a special vitamin on vision, a pair of subjects (one a confederate) went through the different

[1] The "angry" condition was disallowed because the committee believed it unethical to induce anger in unsuspecting subjects (assuming the validity of the Schachter and Singer conclusions). The concern of the committee about possible adverse reactions to epinephrine injections in normal young adult males was based on an unreferenced comment in a general pharmacology reference text. In over 100 administrations of epinephrine of the dosages reported in this article, there was not a single adverse reaction. The precautions we took of screening for potential overreactions, availability of a cardiologist, and continuous monitoring of physiological functioning proved to be sufficient safeguards and should be used in further investigations of the psychological effects of this drug in normal populations.

phases of experimental procedures together. First, following an explanatory overview of the session, they completed a series of questionnaires that provided an evaluation of their preinjection physical and affective state. Then they took specious "pre-vitamin-injection" vision tests and had a baseline heart rate measurement while performing visual tasks. Next a physician administered the injection (alleged to at least subtly affect visual activity), with accompanying information regarding expected side effects. During a 15-minute waiting period, supposedly before the drug would begin to take effect, the confederate engaged in a prearranged sequence of euphoric activities. At the end of the waiting period, measures of self-reported affect, experienced symptoms, and an evaluation of the confederate were obtained. The specious "drug condition" vision tests were then readministered in keeping with the purported rationale of the study. The physician and the experimenter then examined and interviewed each subject, who was dismissed only when his pulse and blood pressure were normal and there were no manifest signs of arousal. The duration of each session was about 90 minutes. Finally, a separate and thorough debriefing session was held for all subjects at the completion of the study.

Subjects

The subjects were 85 male undergraduates attending Stanford University. They were initially recruited by means of a notice asking for males over 18 years of age to join the "Stanford Subject Pool for Drug Research." The medical records of these subjects (maintained at the student health center) were examined by a physician. Subjects were excluded from the subject pool if they met any of several criteria that were contraindicative of a normal-range reaction to epinephrine or that might indicate a prior history of epinephrine use (e.g., asthma, a respiratory condition, high blood pressure) or if currently using medication. Fifteen volunteers were thus judged ineligible to participate.

Subjects were individually contacted by phone and invited to participate in a study "being conducted by a member of the ophthalmology department concerning the influences of an injection of a vitamin supplement on visual activity." A standard medical interview was conducted over the phone, both as a further screening device and to lend credence to the cover story (several additional volunteers were eliminated through this second phase of screening.) The eligible volunteers were then randomly preassigned to one of six experimental conditions, which are described subsequently.

Procedure

In keeping with the purported medical nature of the study, the experiment was performed in the student health center, where appropriate signs directed the incoming subjects to the "vision testing area." The subject and a male experimental confederate of comparable age went through all of the procedures together. At the beginning of the session, the experimenter presented an overview of the experiment to them. He repeated the description of the study as an ophthalmology experiment dealing with vitamins and visual activity and described the tandem procedures for the repeated vision tests and "physiological monitoring." Subjects were told that the physician would administer the vitamin injection following the completion of the initial series of tests. The injection was described as having no major side effects, although the experimenter indicated that some minor, transitory effects had been reported by some previous subjects. Misinformation was given regarding these somatic effects; instead of the usual epinephrine sequelae, subjects were led to expect dryness of the throat, slight headache, and some slight coolness in toes and fingers. Subjects were told that they would have to wait about 15 minutes after the injection before the vitamin would be affecting the appropriate visual structures, and then they would proceed with the second series of eye tests, physiological recordings, and questionnaires.

After the experimenter delivered this overview, the first set of questionnaires was administered. These contained self-ratings concerning vision, physical condition, and current affective state. When the forms were completed, the participants proceeded to the "baseline testing" phase of the experiment. The vision tests consisted of reading a standard eye chart with and without a pair of field-inverting goggles and taking the Speed of Color Discrimination Test (Messick, 1964). The "preliminary physiological monitoring" period, which included several visual tasks, initiated the continuous recording of the subject's heart rate as he proceeded through the rest of the experiment. Silver disk surface electrodes were attached to the center of the subject's upper and lower sternum. These were then connected to a small radio transmitter (Narco Bio-Systems, FM1100-E) placed on the back of the subject's shoulder. The transmitter was described to the subject as a "relay box" that was connected to the polygraph (E & M Physiograph) immediately behind him. In actuality, the signal from the transmitter was picked up by two FM receivers (Narco Bio-Systems, FM1100-6). One was connected to the polygraph in the vicinity of the subject; the other was connected to a polygraph (Beckman Type R dynagraph with a 9806-AC coupler and a 9857 cardiotachometer coupler) concealed from view in another room.

After the subject performed several "eye movement" tasks, a 1-minute resting baseline measure of heart rate was obtained. As the subject's "initial physiological monitoring" period was being completed, the physician entered the room and pretended to administer an injection to the confederate, who had just completed his baseline vision testing. The experimenter conspicuously disconnected the "relay box" from the adjacent polygraph and informed the subject that no further physiological recording would be done until the second part of the experiment.

actuality, the subject's heart rate information was being continuously telemetered by the radio transmitter to the receiver and polygraph in the other room.

Manipulation of physiological arousal. After the physician finished with the confederate, he entered the physiological recording area where the subject was waiting. As the physician prepared to give the injection, he repeated both its earlier description as a "multicomponent vitamin supplement called D-27" and the earlier statement regarding its likely somatic effects. The injection was then administered subcutaneously to the subject's upper right arm. Subjects in the epinephrine-basic group received an injection of .5cc epinephrine (a 1:1000 solution of Parke-Davis Adrenalin). The placebo-basic group received a .5-cc injection of bacteriostatic water. Only the physician (a cardiologist) was aware of the contents of specific injections. The experimenter then escorted the subject to the adjacent waiting area, where the confederate was already seated at a table. The rationale for the 15-minute waiting period was restated. After instructing the two participants simply to make themselves comfortable, the experimenter excused himself and left them alone in the room.

Manipulation of social environment. The confederate enacted a carefully rehearsed sequence of behaviors that was closely patterned on the behavior of the "euphoric" accomplice in the Schachter and Singer (1962) study. The confederate had been instructed to be friendly and responsive to the subject but to maintain a primarily independent role.

The programmed "euphoric" script of the confederate, both verbal and nonverbal, followed this approximate chronology (which began within a minute after the subject's injection):

1. *Minutes 1 through 3.* Looks around, examines materials on the table, sketches or doodles, hums.

2. *Minutes 4 through 5.* Crumples paper into a ball, tosses it at wastepaper can; repeats with a couple more sheets of paper; gets up, retrieves misses, tries different types of shots from different positions; tosses one to the subject, inviting him to try a shot.

3. *Minutes 6 through 8.* Sits down again, picks up a piece of paper, makes a paper airplane and flies it; makes another one, flies it; retrieves plane from the floor and flies it in the direction of subject; repeats flying plane.

4. *Minute 9.* Notices and tries a swivel-type foot exerciser on the floor.

5. *Minutes 10 through 12.* Picks up paper airplane, sits down and plays with materials on the table; makes a slingshot with a rubber band; tears off a piece of the airplane to use as "ammunition" and shoots it at a clock on the far wall; while retrieving shot, notices old, empty folders on a piece of equipment and sets them up as a target; makes more "ammunition" and continues to shoot at the target.

6. *Minutes 13 through 15.* While picking up "ammunition" from floor, notices hula hoops against the wall behind some exercise equipment (in the adjoining physical therapy room), picks one up and tries a couple of times to rotate it freely around his hips, and then spins it across the room toward the subject; continues spinning the hoop on the floor until the sound of a door opening indicates the experimenter's imminent return; replaces hoop and returns to his seat.

When the experimenter reentered, he stated that as soon as the second set of questionnaires was completed, the participants would proceed with the physiological monitoring and vision tests "under vitamin conditions." These questionnaires were the same as the first series but also included both a checklist and ratings of a variety of experienced physical symptoms. An additional form asked both participants for a rating of their experimental partner. The purpose of this "partner evaluation" (really a check on the perceived mood of the confederate and also of suspicion as to his role) was rationalized as necessary to match subjects for a future phase of the study. When the questionnaires had been completed, the initial vision tests and tasks were repeated. Following this, the subject was brought to another room where the physician checked his pulse rate and blood pressure. The experimenter gave a brief exposition about the experiment, paid the subject, and requested that he not discuss the experiment. A complete debriefing session was postponed for several weeks until all of the subjects had been tested. At these subsequent sessions subjects filled out a modified form of the Autonomic Perception Questionnaire (Mandler, Mandler, & Uviller, 1958) and were then completely debriefed.

Experimental Conditions

Sixteen subjects were randomly assigned to each of the two basic experimental conditions. The epinephrine-basic group received the epinephrine injection, while the placebo-basic group received the placebo injection. In all other respects the procedure outlined above was identical for both conditions. Four other groups (with 12 subjects in each) were utilized to provide essential comparative data. In the epinephrine – neutral confederate condition, the confederate did not behave euphorically, but instead he was merely passively pleasant. In the placebo – arousal symptoms group, the information regarding anticipated consequences of the injection was changed to that appropriate for epinephrine. The final two groups represented an attempt to strengthen the manipulation of physiological arousal. In the epinephrine – increased arousal group, the dosage of the epinephrine injection was individually adjusted for the subject's body weight. In the epinephrine – increased salience group, subjects not only received the adjusted dosage but also engaged in activities that were designed to increase the salience of their arousal.

Epinephrine – neutral confederate group. In this condition every aspect of the procedure, with the exception of the confederate's behavior, was identical to that for the subjects in the epinephrine-basic group. The confederate, instead of acting in a eu-

phoric manner, offered no dramatic cues as to his emotional state. When the subject and confederate were brought together in the waiting room after the injection, the confederate casually removed a paperback book from his back pocket and proceeded to read it for the entire waiting period. He responded in a friendly, pleasant manner to any questions that were asked, but then returned to his reading.

The primary reason for this neutral confederate condition was to provide a comparison level for assessing the extent of affect induction elicited by the euphoric confederate in epinephrine-injected subjects. The neutral confederate condition permits a direct evaluation of the impact of varied social environments, holding inadequately explained physiological arousal constant.

Placebo – arousal symptoms group. In this condition, a group of placebo subjects was told that the symptoms expected from the injection included those actually appropriate to an injection of epinephrine. This information was provided during the overview at the beginning of the session and again just prior to the injection. Subjects in the placebo – arousal symptoms group were told, "You may notice that your hand will start to shake, your heart will start to pound, and your face may get warm and flushed." With the exception of this change in symptom expectation, the procedures for this group were identical to those for the placebo-basic group.

The primary reason for this experimental condition was to further examine the effects of the symptom-expectancy manipulation. This is the particular information Schachter and Singer (1962) employed to provide some subjects in epinephrine-injection conditions with an "adequate" explanation for their ensuing physiological arousal. The arousal symptoms group illustrates the effects of providing arousal-appropriate information, independent of the subsequent induced state of physiological arousal. In particular, we wanted to test the possibility that telling subjects to expect their heart to pound and so forth might have independent, direct effects on affect in addition to simply providing a possible explanation for any subsequent experience of physiological arousal.

Heightened arousal groups. In these conditions, instead of receiving a "fixed" dosage of epinephrine, subjects received dosages in proportion to their body weight. The average body weight of the subjects in the epinephrine-basic group was approximately 72 kg; thus, the standard injection of .5 cc represents an average injection of about .007 cc/kg. Heavier-weight subjects in the constant dosage condition, of course, received relatively smaller dosages than lighter ones.

The 12 subjects randomly assigned to the epinephrine – increased arousal treatment received an injection of .01 cc of epinephrine per kilogram of body weight. The individual dosages ranged from .55 cc to .82 cc ($M = .72$ cc; $SD = .08$). With the exception of this change in dose determination, the procedures and manipulations for this group were identical to those for the epinephrine-basic group.

This method of selecting dose levels controls for any differential effects resulting from individual differences in body weight. The average of the actual doses administered to this group is equivalent to 1.44 times the fixed dosage administered in the other epinephrine conditions and in the Schachter and Singer study. As such, this method also represents a somewhat stronger level of the physiological arousal manipulation for the group as a whole.

The 12 subjects randomly assigned to the epinephrine – increased salience condition were also administered the body-weight-adjusted drug dosage of .01 cc/kg. Individual dosages ranged from .64 cc to .93 cc ($M = .77$ cc; $SD = .08$). In addition, for these subjects the postinjection waiting period, while euphoric in tone, was designed to increase the salience of the induced physiological arousal. The subject and confederate spent the first 5 minutes of the waiting period writing a story about a Thematic Apperception Test (TAT) picture (to increase the salience of hand tremor). The remaining 10 minutes were spent sitting with their eyes closed (to focus attention on increased palpitations and changes in respiration). The confederate enacted his "euphoric" role both verbally and nonverbally. He joked, hummed, and gave various indications of being happy.

This change in task demands during exposure to a "euphoric" confederate, combined with the somewhat stronger arousal manipulation resulting from the weight-adjusted drug dosages, provided conditions under which the awareness of physiological arousal should have been enhanced. The critical components were still present: arousal that was inadequately explained and a generally euphoric confederate.

Dependent Measures

Physiological, behavioral, and self-report responses were measured.

Physiological measures. Each subject's heart rate was monitored continuously using a cardiotachometer. The three measurement periods were (a) baseline—the 1-minute relaxation period during the initial physiological monitoring; (b) preinjection—the 1-minute period immediately preceding the injection; and (c) waiting period—the period beginning at 1 minute after the injection and lasting until the end of the waiting period. For scoring purposes, the rate indicated by the cardiotachometer record was read at 10-sec intervals for each of the appropriate periods. Means where then calculated for the appropriate epochs.

Behavioral measures. During the waiting period, the subject was both observed and videotaped through a one-way mirror. Each discrete occurrence of smiling or of laughter was recorded on an event recorder (Esterline-Angus). The total frequency of each of these behaviors was determined for each subject. A composite index of positive emotional expression was calculated by weighting the number of smiles by one and the number of laughs by two and then summing these scores. The videotape of each

subject's activities was scored by an observer who was blind to the experimental condition. The type, frequency, and duration of the subject's various activities, as well as their temporal occurrence with respect to the confederate's actions, were noted. Our scoring procedure was based on a weighting of type of activity, following Schachter and Singer (1962, p. 391). Thus, 5 points were given for hula hooping, 4 for shooting with the slingshot, 3 for paper airplanes, 2 for paper basketballs, 1 for doodling, and 0 for doing nothing. Scores for amount of "imitation" (follows confederate's lead) and "initiation" (innovates new activities or encourages further participation of confederate) were also determined. To assess the reliability of these various measures, 10 of the videotapes were selected at random and rated by a second independent observer. The scores for both judges of the total frequency for each activity for each of the 10 subjects differed by only 1 unit or less 97% of the time and were in complete agreement 85% of the time. The imitation and initiation scores were in exact agreement 90% and 70% of the time, respectively; and for both categories the scores were within 1 point of each other 100% of the time.

Self-report responses. Questionnaires were used to assess the following: (a) the subject's description of his affective state, (b) his perception of physiological arousal, and (c) his evaluation of the confederate. The measures concerning affective state were administered both at the beginning of the experiment and following the waiting period. The reports of physical state and the confederate evaluations were obtained only after the waiting period. In order to disguise the actual focus of the questionnaires, several "dummy" questions were included (such as ratings of hunger, thirst, and degree of alertness, as well as ratings of different aspects of visual activity). The questionnaires also contained several open-ended questions encouraging the subject to comment on various aspects of his experience.

Two of the affect self-report items provided the major dependent measure for purposes of replication of the previous findings by Schachter and Singer (1962): "How irritated, angry, or annoyed would you say you feel at present?" and "How good or happy would you say you feel at present?" Subjects checked one of five response alternatives: 0 (not at all), 1 (a little), 2 (quite), 3 (very), and 4 (extremely). An additional series of 11 items sampled a broader affective domain. These were in the form of bipolar adjective pairs, with the subject rating himself on a 7-point scale ranging from −3 for one adjective to +3 for the other. Anxious–calm, elated–depressed, and angry–peaceful are examples of these 11 item pairs.

The questionnaires also contained two sections pertaining to the subject's perception of various physical symptoms. The experience of palpitations and the feeling of tremor were rated on the following scale: 0 (not at all), 1 (a slight amount), 2 (a moderate amount), and 3 (an intense amount). The subject also completed a checklist noting the presence or absence of a variety of symptoms, such as heart beating faster, breathing faster, numbness, and watery eyes.

The subject rated the confederate on a scale from −3 to +3 for each of nine adjective pairs (e.g., friendly–unfriendly, relaxed–agitated). One pair, happy–sad, permitted a direct check on the adequacy of the confederate's portrayal of a "euphoric" mood.

An additional questionnaire was administered at the debriefing session. This questionnaire was developed from the Autonomic Perception Questionnaire and was used to measure the extent to which a subject experienced specific physical symptoms with different states of affect. The standard deviations of symptom ratings were used to derive an affect symptom–discrimination index. Subjects with a low index score did not report experiencing symptoms that varied with the different emotions (e.g., if a subject rated himself as "perspiring a great deal" regardless of the specific emotion, that symptom would have a standard deviation of zero). Subjects with a high index score, on the other hand, did report experiencing different symptoms with different emotions.

Results

Five subjects met an a priori exclusion criterion based on skepticism of the validity of the confederate's role (three subjects from placebo and two from epinephrine groups). Exclusion resulted when a subject noted himself as "suspicious" on the adjective checklist and questioned the credibility of the confederate on the open-ended questionnaire items. These subjects were replaced with additional subjects from the available subject pool.

This study was conducted in two phases: First, the basic epinephrine and placebo groups, both paired with a euphoric confederate, were tested; a week later, testing of the four additional groups began. All subjects came from the same population, were tested within a comparable time period, and were randomly assigned to experimental conditions. For purposes of data analysis, the six groups are treated as if they had been tested contemporaneously.

Checks on the Manipulations

The manipulations appear to have been effective in creating the conditions necessary to evaluate the hypotheses under consideration. Heart rate increases as a function of the injection were found only for the epinephrine – heightened arousal groups, in both

Table 1
Reported Experience of Physiological Symptoms

Measure	Placebo groups		Epinephrine groups			
	Basic	Arousal symptoms	Basic	Neutral confederate	Increased arousal	Increased salience
Mean degree of tremor	.1	.2	1.6	1.5	1.6	1.7
Mean degree of palpitations	.1	.2	.9	.9	1.4	1.6
Mean number of epinephrine symptoms	.5	.8	2.8	2.4	3.4	3.4

of which the dosage of epinephrine had been individually adjusted for body weight. The injections of epinephrine, however, clearly did lead subjects to perceive more symptoms of physiological arousal than were perceived by placebo subjects. The confederate was perceived as being happy and euphoric in all experimental conditions where he behaved that way and as significantly less so where he enacted an affectively neutral role. Finally, the misinformation regarding the symptoms associated with the injection was reflected in the failure of a majority of the subjects to mention the injection as a possible explanation for their feeling state.

Physiological arousal. Because of apparatus problems, only 68% of the sample had complete physiological records. Examination of the change in heart rate from the period immediately preceding the injection to the average heart rate during the 15-minute waiting period revealed the following results. There were no significant changes in heart rate for the two epinephrine conditions given the standard, fixed dosage of .5 cc of epinephrine: $M = -2.3$ beats per minute (bpm) for the epinephrine-basic group, and $M = +.9$ bpm for the epinephrine – neutral confederate group. There were, however, increases in heart rate for the two heightened arousal conditions. The mean increase for the epinephrine – increased arousal group was +6.7 bpm, $t(7) = 3.12$, $p < .02$, while the mean increase for the epinephrine – increased salience group was +6.0 bpm, $t(8) = 2.26$, $p < .10$. The placebo-basic group did not show a significant change in heart rate ($M = -3.0$ bpm), while the placebo – arousal symptoms group had a significant decrease in heart rate ($M = -6.7$ bpm), $t(8) = 2.54$, $p < .05$.

Perceived arousal symptoms. Table 1 presents the mean ratings of palpitation and of tremor. Four of the items on the symptom checklist ("moist, sweaty palms," "heart is beating faster," "stomach feels upset," and "breathing is faster") were considered as indicative of epinephrine – induced arousal, and their mean frequency is also reported in Table 1. Subjects in the two placebo groups rarely reported tremor and palpitations, and only occasionally did they report experiencing one of the epinephrine symptoms. In contrast, the subjects in the four epinephrine conditions on the average reported experiencing slight to moderate amounts of tremor and palpitations, as well as at least several of the epinephrine symptoms. Thus, with respect to reports of both the degree of tremor and palpitation and also to reports of the number of symptoms experienced, the injection of epinephrine did definitely produce discernible symptoms of arousal that were not found with the placebo injection.

Confederate's affective state. The success of the "social cognition" manipulation was determined from the ratings of the confederate completed by each subject after the waiting period. The means of the ratings in each of the five groups paired with a euphoric confederate indicate that he was indeed perceived as being happy. On a scale from +3 (happy) to −3 (sad), the confederate was given mean ratings that ranged from +1.5 to +2.2 for the five conditions having a euphoric confederate. When tested against a neutral score of zero, each of these mean ratings is highly significant ($p < .001$). This perception of the confederate as being happy was even found in the ratings of the more restrained confederate in the epinephrine – in-

creased salience condition ($M = +1.7$). In contrast, the epinephrine group with the affectively "neutral" confederate gave him a mean rating of +.7 (not significantly different from the neutral midpoint of zero).

Explanation of symptoms. The subjects were misinformed regarding the possible symptoms associated with the injection. Following Schachter and Singer's (1962) procedure, subjects were subsequently categorized as "self-informed" or "not self-informed" about the source of their physiological arousal. Subjects who mentioned something about the injection (e.g., "suspect that it's the vitamin influence"; "the shot"), when asked to explain their feeling state on the final questionnaire, were classified as "self-informed." Those who made no reference to the injection (e.g., "test situation"; "the guy I was waiting with") were classified as "not self-informed." There were indeed subjects in each group who met the "self-informed" criterion. In the epinephrine conditions, this varied from 25% of the basic group and 17% of the increased arousal group to 42% of the neutral confederate group and 50% of the increased salience group, both of whom were exposed to a relatively inactive and less distracting confederate. In the placebo conditions, 12% of the subjects in the basic group and 8% of the arousal symptoms group also met the criterion. The reports in these latter groups suggest that at least some of the references to the injection in the epinephrine groups may have occurred independently of induced physiological arousal. Although the misinformation regarding likely side effects did not necessarily forestall cognitions linking the injection with subsequent feeling states, the majority of subjects did not offer such an explanation in spite of the obvious contiguity of the injection. This supports the apparent face validity of the subjects' not having a readily available "adequate" explanation for any ensuing physiological arousal.

Experimental Comparisons

The epinephrine-basic subjects did *not* express more positive affect (on either the behavioral or the self-report measures) than did the placebo-basic group. In fact, what few differences did exist between the two conditions were in the opposite direction of more positive affect for the placebo group. Heightening the arousal did not increase reports or displays of positive affect, but instead produced signs of more negative affect. For subjects with inadequately explained physiological arousal, it seemed to make little difference if they were in the presence of an affectively positive or neutral confederate. Placebo subjects given information to expect symptoms appropriate to epinephrine evidenced a negative affective bias prior to the actual injection.

Preinjection. With the exception of the information regarding possible somatic effects, the experimental procedures were identical for all subjects prior to the injection. Subjects in the placebo – arousal symptoms group were told to expect epinephrine-appropriate symptoms; subjects in the other groups were told to expect only minor, transitory symptoms (unrelated to predominant epinephrine effects). An index of general affect was derived from the set of 11 bipolar affect-adjective scales by combining the ratings on the items such that a higher score represented more positive affect in general. The placebo – arousal symptoms group was the only one that did not have a positive index score significantly different from zero (all other ps at least $<.02$). The a priori comparison of the mean for the placebo – arousal symptoms group ($M = 2.5$) with the average of the scores for the other groups ($M = 7.4$) indicates that these subjects had a significantly lower or less positive index score ($MS_w = 55.59$), $t(74) = 2.12$, $p < .05$, after being informed about possible symptoms.

The effect of the symptom-information manipulation was also assessed by the a priori comparison of the change in heart rate for the placebo – arousal symptoms group in anticipation of the injection ($M = 14.8$ bpm) with the average of the change in heart rate for the other five groups ($M = 6.4$ bpm). This comparison indicates that subjects in the arousal symptoms condition had a significantly greater increase in heart rate in anticipation of their injection than did the other subjects who had been told to expect only minor, transitory symptoms ($MS_w = 46.70$), $t(47) = 3.40$, $p < .01$. This and the general affect index demonstrate that the two de-

Table 2
Self-Report Ratings of Experienced Affect

Measure	Experimental condition					
	Placebo–basic (n = 16)	Epi–basic (n = 16)	Epi–increased arousal (n = 12)	Epi–increased salience (n = 12)	Epi–neutral confederate (n = 12)	Placebo–arousal symptoms (n = 12)
Schachter–Singer index post score[a]	1.5	1.4	.9	1.1	1.2	1.5
"Adjusted" pre–post change score[b]	+.40*	+.28	+.03	−.07	−.25	+.39
General affect index post score[a]	10.4	7.4	1.8	.5	8.2	8.3
"Adjusted" pre–post change score[b]	+3.64*	+.19	−4.31*	−6.41**	−1.00	+4.20*

Note. Epi = epinephrine.
[a] The higher the score, the more positive the reported affect.
[b] A positive score indicates a positive change in affect over time; a negative score indicates a negative change. Scores were adjusted by regression for any differences in initial values.
* $p < .05$.
** $p < .01$.

scriptions of side effects were not equivalent and that the symptom-information manipulation produced effects *prior* to the injection.

Self-report of affect. One of the self-report measures of experienced affect was identical to that employed by Schachter and Singer (1962): Subjects rated how "good or happy" and how "irritated, angry, or annoyed" they felt. Following Schachter and Singer's procedure, the "angry" score was subtracted from the "happy" score to yield a single index of affect. As can be seen in Table 2, all of the six conditions had a positive postscore on this affect index. This measure was administered both before and after the experimental manipulations in order to assess changes in reports of affect. An analysis of these changes ("adjusted" by regression on the initial scores) revealed *no* significant changes for any of the epinephrine conditions. However, the placebo-basic group showed a positive change, $t(15) = 2.20$, $p < .05$, and the placebo–arousal symptoms group also tended to do so, $t(11) = 1.86$, $p < .10$.

The other self-report measure of experienced affect was the general affect index derived from the set of 11 bipolar affect-adjective scales. Table 2 presents the mean scores on this index for the different groups. A positive general index score (significantly different from the neutral midpoint of zero) was obtained by both the epinephrine-basic group, $t(15) = 3.73$, $p < .01$, and the placebo-basic group, $t(15) = 6.40$, $p < .001$. Similarly positive scores were also obtained by the placebo–arousal symptoms group, $t(11) = 7.13$, $p < .001$, and the epinephrine–neutral confederate group, $t(11) = 2.92$, $p < .02$. In contrast, the two epinephrine–heightened arousal groups had scores that did not differ from zero.

The general index score of the epinephrine-basic group was slightly *less* positive than that of the placebo-basic group (contrary to the Schachter and Singer prediction), and that of the epinephrine–increased arousal group was significantly less positive ($p < .01$). In the epinephrine conditions, there was no significant difference between the basic and neutral confederate groups, while the scores of both of the heightened arousal groups were significantly less positive than that for the basic group (p at least $< .05$).

An examination of the pre- to postexperimental changes in general index scores ("adjusted" by regression) reveals an interesting pattern (see Table 2). In the epinephrine conditions, both the basic and neutral confederate groups showed no significant change on this affect index. However, both of the heightened

arousal groups had significantly *negative* changes on this index: for the increased arousal group, $t(11) = 2.39$, $p < .05$; for the increased salience group, $t(11) = 3.56$, $p < .01$. In contrast to all of the epinephrine groups, the two placebo conditions showed a significantly *positive* change in this general index: for the basic group, $t(15) = 2.34$, $p < .05$; for the arousal symptoms group, $t(11) = 2.28$, $p < .05$.

A comparison of the epinephrine-basic group with the placebo-basic group reveals *no* significant difference between their general index change scores. The epinephrine–increased arousal condition had significantly more *negative* change scores than the placebo-basic group ($p < .01$). In the epinephrine conditions, the change scores for the increased salience group were significantly more negative than that for the basic group ($p < .01$), and those for the increased arousal group also tended to be more negative ($p < .10$). There was no significant difference between the epinephrine-basic group and the neutral confederate group.

Emotional behavior. The analysis of the subjects' euphoric activities is restricted to the four groups in which the subjects spent the experimental waiting period with an active, euphoric confederate who went through the standardized Schachter–Singer (1962) routine. One behavioral measure of affect involved nonverbal expressions of positive affect. The numbers of observed smiles and laughs were combined to form an index of positive emotional expression. Subjects paired with the neutral confederate expressed very little positive affect ($M = 1.3$), while all the other subjects who were paired with a euphoric confederate displayed considerably more positive emotional reactions. The placebo groups had the highest scores on this index ($M = 28.1$ for the basic group; $M = 22.5$ for the arousal symptoms group). The mean score for the epinephrine-basic group ($M = 21.7$), although lower, did not differ significantly from the placebo-basic condition. However, the scores for the epinephrine–increased arousal group ($M = 18.2$) were significantly lower ($p < .05$) than those of the placebo-basic group.

Table 3
Behavioral Measures of Euphoric Activity

Measure	Experimental condition: Placebo–basic ($n = 15$)	Epi–basic ($n = 11$)	Epi–increased arousal ($n = 12$)	Placebo–arousal symptoms ($n = 11$)
Euphoric activity	35.0	25.6	14.5	19.8
Imitation	1.4	1.2	1.0	1.3
Initiation	.8	.9	.5	.5

Note. Epi = epinephrine. Higher numbers reflect greater degrees of activity. The videotapes for some subjects were not of scorable quality.

Similar to Schachter and Singer's procedure, the number of times each of the activities was performed by a subject was multiplied by the weight assigned to the presumed euphoric quality of that activity. All of these scores were then combined into an overall euphoric activity index (see Table 3). There was a considerable range of scores within each of the groups (the smallest intragroup range was 53.0), with some subjects remaining relatively inactive and others performing a great deal of euphoric behavior. The placebo-basic group had the highest score on this index—as opposed to either of the epinephrine groups. While their scores did not differ significantly from those of the epinephrine-basic group, they tended to be significantly higher than the activity scores for the epinephrine–increased arousal group ($p < .10$). The difference between the two epinephrine conditions did not approach significance. The number of different activities (i.e., the range of euphoric behavior) in which the subject engaged was also examined. Each subject was classified as either (a) participating in none or only one activity or (b) participating in more than one activity. The percentage of subjects in the second category was 67% for the placebo-basic group, 55% for the placebo–arousal symptoms group, 45% for the epinephrine-basic group, and 33% for the epinephrine–increased arousal group. The difference between the placebo-basic and epinephrine-basic groups was not significant, while the epinephrine–increased arousal

group had a significantly lower percentage of subjects exhibiting a range of euphoric behaviors than the placebo-basic group ($z = 1.76, p < .05$).

Similar to the Schachter and Singer (1962) procedure, the number of confederate activities that a subject performed after the confederate had done so was used to constitute an imitation score, and the number of innovative euphoric behaviors performed by the subject was used as his initiation score. As can be seen in Table 3, all of the experimental groups both imitated and initiated euphoric behavior. There were no significant differences between groups on either of these measures.

Additional analyses. As noted earlier, if subjects cited receiving an injection as a possible cause for their feeling state, they were categorized as "self-informed." Being self-informed, according to Schachter and Singer, should render subjects functionally unsusceptible to the affect manipulations. Consequently, their reactions should differ from cohorts not so self-informed. To test the significance of this variable on the pattern of results in the present study, t tests were performed between self-informed and not self-informed subjects on each of the dependent measures. There was no trend nor even tendency for differences between them on any of the self-report or behavioral measures of affect (all p values $> .10$). In other words, subjects responded similarly to the affect manipulations, regardless of their attributions regarding the injection.

The subjects' responses to the physiological arousal manipulation may have been influenced by their perceived association of particular symptoms with different states of affect. Correlations between the affect self-report scores and three scores derived from the subsequently administered Autonomic Perception Questionnaire (APQ) were computed. The APQ scores included the affect symptom-discrimination index (i.e., the extent to which the subject reported experiencing different symptoms with different emotions) and the standard deviation for the ratings on the specific item regarding experiencing "increased heart beat." The final APQ score was the mean on an item assessing how "bothered" the subject was by these "bodily reactions." For the placebo subjects, there were no significant correlations between the APQ scores and the self-reports of affect. However, for comparable subjects who had received an injection of epinephrine, there was a consistent pattern of significant ($p < .05$) negative correlations (ranging from $-.67$ to $-.49$) between affect reported in the experimental setting and the affect symptom-discrimination index and also the other APQ scores. The subjects who reported more negative responses to their experimentally induced, inadequately explained physiological arousal were also those who were more likely to report doing the following: (a) discriminating between different states of affect in terms of physical symptoms, (b) experiencing increased heart beat differentially with various emotions, and (c) being bothered by these bodily reactions. Thus, symptom-sensitive subjects showed a negative affective bias in response to symptoms associated with the epinephrine-induced physiological arousal.

Discussion and Conclusions

There is no evidence in the data generated by this study to warrant acceptance of the Schachter and Singer (1962) reported demonstration of the interaction of cognitive and physiological determinants of emotion. In particular, there was no instance in which subjects with inadequately explained epinephrine-produced arousal were significantly more susceptible than placebo controls to the induction of affect by exposure to a confederate who modeled euphoric behavior. If, as Schachter (1971) has suggested, the disguised injection of epinephrine actually did create "a bodily state 'in search of' an appropriate cognition" (p. 33), the majority of these subjects did not discover that cognition in the mood displayed by the euphoric confederate. Whatever arousal treatment differences did exist were likely to suggest more positive affect in this situation for placebo subjects than for epinephrine subjects.

The particular physiological substrate induced by exogenous epinephrine seems, from the consistent pattern of our data, to have

produced a relatively negative affective bias, even given the presence of a euphoric confederate. This finding became more evident under conditions of increased arousal and increased salience of the arousal. The presence of a euphoric confederate exerted some functional control over the "emotional" behaviors of the subjects, but had minimal impact on their self-reported affect. "It's like being at a party," one epinephrine subject later reported, "when everyone is clearly having a good time and you have a headache or feel depressed for some unknown reason. You don't want to rain on their parade; you might even laugh and try to join in, but you don't really feel happy." Nowhere was there evidence of inadequately explained epinephrine-produced arousal enhancing affective susceptibility to situational influences.

Reappraisal of the Schachter and Singer Conclusions

As previously mentioned, Schachter and Singer (1962) similarly found no significant differences between their placebo controls and epinephrine aroused groups (see their Table 2, p. 390). However, they considered the direction of the differences to provide "tentative but consistent" support for their predictions regarding cognitively steered physiological arousal. Their actual conclusions concerning the role of physiological arousal, however, are primarily based on several somewhat questionable post hoc internal analyses, such as classifying placebo subjects with no change in pulse as being "aroused" (see Marshall, 1976, for a fuller critique). By these means, they were able to obtain some "predicted" effects within groups on the behavioral measures. They were still unable, however, to demonstrate similar effects on the self-report measures. In the present study, procedural improvements were introduced to strengthen the possibility of finding any between-group arousal treatment effects. Despite this, there were still no significant differences obtained between the arousal treatment conditions. Opposite to what Schachter and Singer found, this time the direction of differences that did exist was in the favor of more positive affect in the presence of a euphoric confederate for the placebo group. Furthermore, this pattern of results was supported by the data from additional comparison groups. These data suggest a negative affective bias for epinephrine-produced arousal. Thus, it seems most accurate to conclude that inadequately explained epinephrine-produced physiological arousal does not enhance general affective susceptibility to environmental influences.

This lack of "emotional plasticity" is also suggested by other results in the Schachter and Singer study. Interestingly, Schachter and Singer do not compare the effects of the different social environments presented in their study. When such a comparison is done, it is apparent that the difference in reported affect between epinephrine-ignorant (without an adequate explanation) subjects exposed to an angry confederate and those exposed to a euphoric confederate is rather small (.4 scale units) compared to the difference found for epinephrine-informed (with an adequate explanation) subjects (.9 scale units). This marginal situational effect for inadequately explained physiological arousal is similar to the minimal influence found for the affective tone of the social environment in the present study. Thus, there is additional evidence to support the conclusion that inadequately explained epinephrine-produced physiological arousal does not enhance general affective susceptibility to environmental influences.

Although they only employed postexperimental measures, Schachter and Singer interpreted their within-confederate condition results in terms of *changes* in affect for their unexplained arousal groups. In the present study, the use of pre- and postexperimental measures permitted a direct assessment of actual change. No significant change in affect was found for subjects who were comparable to Schachter and Singer's unexplained arousal group. This suggests the possibility that the group differences they obtained were actually due to changes by their accurately informed epinephrine subjects. As already pointed out, this condition was the one to show the greatest situational effect. Additional evidence suggesting that the accurate information condition was the one to have the changes in affect is provided by the re-

sults in the present study for the placebo subjects who were told to expect epinephrine-type symptoms (the "accurate" information manipulation).

Placebo subjects who were told to expect the somatic effects associated with epinephrine reported relatively more negative affect prior to the injection and had a relatively greater increase in heart rate in anticipation of the injection. This information about possible effects of the injection seems to have produced a negative affective response by itself. This response appears to have dissipated over the course of the experiment, presumably because the anticipated effects did not occur (since in fact they received an injection of a placebo). If these subjects had indeed received an epinephrine injection, then it is probable their initial negative reaction would have persisted over the course of the contact with the confederate. That condition is, of course, the epinephrine – informed treatment in the Schachter and Singer study. Therefore, it seems likely that their "accurate information" manipulation, instead of precluding any affective change, actually produced one. If the differences found by Schachter and Singer are actually due to changes for their informed condition instead of their unexplained condition, then their posited motivated search to account for inadequately explained arousal is not demonstrated by their results.

Schachter and Singer (1962) have clearly made valuable theoretical contributions to our understanding of emotion. However, there does not appear to be an adequate evidential basis for the conclusions they have drawn. The conclusions suggested by the present study also seem to question the body of literature on the cognitive alteration of feeling states (see London & Nisbett, 1974), which was largely based on the Schachter and Singer work. We recognize, of course, the power of cognitions to control a range of human reactions (Zimbardo, 1969), as is also evidenced by the results of the current placebo group with arousal expectations. We equally believe, however, that there are definite limitations to the power of cognitive control. We argue for a much greater complexity in the interaction of physiological substrate and cognitive cues in determining emotions. Immediately available social cognitions are but one of many sources of information used to give arousal an emotional quality. When other determinants of various classes of responses are not explicitly acknowledged in research designs but still exert an influence, their presence is manifested in excessive error variance. Thus, many of the studies that followed in the wake of the Schachter and Singer study have failed to find predicted effects across both behavioral and experiential variables (see Bem, 1974). Or, as in the case of the extention of Schachter's theorizing to the cognitive manipulation of pain (Nisbett & Schachter, 1966), the statistically significant effects are conceptually insignificant. In that study, for example, no differences were found on any of three measures between subjects who attributed their high degree of fear to a pill or to an impending shock. The only reliable result is the least interesting one, namely, low-fear subjects who accurately attributed their arousal to the shock had a lower tolerance threshold than all other groups (none of which differed among themselves).

Inadequately Explained Arousal

The idea of dissociating physiological from cognitive elements of emotion had been suggested to Schachter and Singer by the earlier work of Marañon (1924). Marañon, it will be recalled, reported two types of reaction to epinephrine injections: "true" emotions and "cold" or "as if" emotions. In the former, not only were there the same somatic elements as in the mock emotions but also the "psychic element." It is interesting to note that the true emotions, though less frequent, were typically of a negative affective tone. A true emotion appeared spontaneously in only some cases; "in other cases, to bring it about the investigator must suggest a memory of great affective energy or force" (p. 307).

Later research (Cantril & Hunt, 1932; Landis & Hunt, 1932) confirmed Marañon's finding. In these and subsequent studies, when cases of "true" emotion were judged to be present, the emotion was generally de-

scribed as "fear," "anguish," or "anxiety." Reviews of this literature offer various explanations. Breggin (1964) suggests that the appearance of these negative emotions may have been due to the superimposed stress of the experimental procedures. A more interesting possibility is that these true negative emotions "arose from the particular (learned) associations the somatic sensations may have" (Rothballer, 1959, p. 511). These associations, in turn, are likely to have arisen from prior experiences with endogenous epinephrine. In the present study, the results indicate that subjects who reported distinguishing between emotions in terms of somatic sensations also reported more negative affect in response to their epinephrine-induced physiological arousal. We know from the reviews of the extensive research on endogenous epinephrine that "increased epinephrine excretion seems to occur in states of anxiety or in threatening situations of uncertain or unpredictable nature in which active coping behavior may be required but has not been achieved" (Schildkraut & Kety, 1967, p. 23).

The physiological substrate associated with at least a certain degree of epinephrine also seems to be somewhat uniquely associated with negative affect. Parenthetically, it is interesting to note that the incidence of subjects being "self-informed" in the Schachter and Singer study varied with the social environment. It was greater with the apparently dissimilar euphoric confederate (32%) than with the clearly negative angry confederate (13%); here the incidence was almost identical with the rate obtained for placebo subjects in the present study (12%). This finding suggests that the confederate's bizarre angry behavior provided a sufficiently plausible explanation for the subject's own negatively toned experience of arousal symptoms. It is likely, of course, that there are certain critical thresholds that must be reached before epinephrine, whether endogenous or exogenous, is associated with negative affect. In the present study, the procedures of both increasing the arousal (by taking individual differences in body weight into account) and additionally increasing its salience consistently produced indications of more negative affect. The variance in the literature in the reports of affect associated with epinephrine may be partially accounted for by individual differences in pharmacokinetics, such as rates of absorption, distribution, and metabolism (Lasagna & Laties, 1959; Smith & Rawlins, 1973), which would contribute to determining whether a threshold had been reached.

The general picture that emerges from the present study, the companion study by Maslach (1979), and a reexamination of the Schachter and Singer (1962) study indicates that epinephrine-related physiological arousal does not provide "emotional plasticity," but rather shows a consistent association with negative affect. The specific expression or manifestation of this negative affective bias may be influenced by cognitive manipulations (Lazarus, 1967), but we believe that the range of outcomes would still be limited to a negative affective domain. It is somewhat reassuring, especially considering their possible adaptive significance, that our true emotions may be more rationally determined and less susceptible to transient or whimsical situational determinants than has been suggested by Schachter and Singer. Perhaps we social psychologists should better appreciate our biological "hardware" or, to paraphrase a frequent media message, "It's not so easy to fool Mother Nature."

References

Averill, J. R., & Opton, E. M. Psychophysiological assessment: Rationale and problems. In P. McReynolds (Ed.), *Advances in psychological assessment* (Vol. 1). Palo Alto, Calif.: Science and Behavior Books, 1968.

Bem, D. J. Cognitive alteration of feeling states: A discussion. In H. London & R. E. Nisbett (Eds.), *Thought and feeling: Cognitive alteration of feeling states*. Chicago: Aldine, 1974.

Breggin, P. R. The psychophysiology of anxiety. *Journal of Nervous and Mental Diseases*, 1964, *139*, 558–568.

Cantril, H., & Hunt, W. A. Emotional effects produced by the injection of adrenalin. *American Journal of Psychology*, 1932, *44*, 300–307.

Erdmann, G., & Janke, W. Interaction between physiological and cognitive determinants of emotions: Experimental studies on Schachter's theory of emotions. *Biological Psychology*, 1978, *6*, 61–74.

Fehr, R. S., & Stern, J. A. Peripheral physiological variables and emotion: The James-Lange theory

revisited. *Psychological Bulletin,* 1970, *74,* 411–424.

Kemper, T. D. *A social interaction theory of emotions.* New York: Wiley, 1978.

Landis, C., & Hunt, W. A. Adrenalin and emotion. *Psychological Review,* 1932, *39,* 467–485.

Lasagna, L., & Laties, B. G. Problems involved in the study of drug-modified behavior in normal humans. In J. O. Cole & R. W. Gerard (Eds.), *Psychopharmacology: Problems in evaluation.* Washington, D.C.: National Research Council, 1959.

Lazarus, R. S. Cognition and personality factors underlying threat and coping. In M. H. Appley & R. Trumbull (Ed.), *Psychological stress: Issues in research.* New York: Appleton-Century-Crofts, 1967.

Leventhal, H. Emotions: A basic problem for social psychology. In C. Nemeth (Ed.), *Social psychology: Classic and contemporary integrations.* Chicago: Rand McNally, 1974.

London, H., & Nisbett, R. E. (Eds.). *Thought and feeling: Cognitive alteration of feeling states.* Chicago: Aldine, 1974.

Mandler, G. Emotion. In *New directions in psychology: I.* New York: Holt, Rinehart & Winston, 1966.

Mandler, G. *Mind and emotion.* New York: Wiley, 1975.

Mandler, G., Mandler, J. M., & Uviller, E. T. Autonomic feedback: The perception of autonomic activity. *Journal of Abnormal and Social Psychology,* 1958, *56,* 367–373.

Marañon, G. Contribution à l'étude de l'action émotive de l'adrenaline. *Revue Française d'Endocrinologie,* 1924, *2,* 301–325.

Marshall, G. *Affective consequences of "inadequately explained" physiological arousal.* Unpublished doctoral dissertation, Stanford University, 1976.

Maslach, C. Negative emotional biasing of unexplained arousal. *Journal of Personality and Social Psychology,* 1979, *37,* 953–969.

Messick, S. *Speed of Color Discrimination Test.* Princeton, N.J.: Educational Testing Service, 1964.

Nisbett, R. E., & Schachter, S. Cognitive manipulation of pain. *Journal of Experimental Social Psychology,* 1966, *2,* 227–236.

Plutchik, R., & Ax, A. F. A critique of "Determinants of emotional state" by Schachter and Singer (1962). *Psychophysiology,* 1967, *4,* 79–82.

Ross, L., Rodin, J., & Zimbardo, P. G. Toward an attribution therapy: The reduction of fear through induced cognitive–emotional misattribution. *Journal of Personality and Social Psychology,* 1969, *12,* 279–288.

Rothballer, A. B. The effects of catecholamines on the central nervous system. Symposium on Catecholamines. *Pharmacological Reviews,* 1959, *11,* 494–547.

Schachter, S. *Emotion, obesity and crime.* New York: Academic Press, 1971.

Schachter, S., & Singer, J. Cognitive, social, and physiological determinants of emotional state. *Psychological Review,* 1962, *69,* 379–399.

Schachter, S., & Singer, J. Comments on the Maslach and Marshall–Zimbardo experiments. *Journal of personality and Social Psychology,* 1979, *37,* 989–995.

Schachter, S., & Wheeler, L. Epinephrine, chlorpromazine, and amusement. *Journal of Abnormal and Social Psychology,* 1962, *65,* 121–128.

Schildkraut, J. J., & Kety, S. S. Biogenic amines and emotion. *Science,* 1967, *156,* 21–30.

Shapiro, D., & Schwartz, G. E. Psychophysiological contributions to social psychology. *Annual Review of Psychology,* 1970, *21,* 87–112.

Smith, S. E., & Rawlins, M. D. *Variability in human drug response.* London: Butterworth, 1973.

Stein, M. Some psychophysiological considerations of the relationship between the autonomic nervous system and behavior. In D. C. Glass (Ed.), *Neurophysiology and emotion.* New York: Rockefeller University Press and Russell Sage Foundation, 1967.

Storms, M. D., & Nisbett, R. E. Insomnia and the attribution process. *Journal of Personality and social Psychology,* 1970, *16,* 319–38.

Wilde, Oscar. *Complete writings of Oscar Wilde* (Vol. 2). New York: John W. Luce, 1908.

Zimbardo, P. G. *The cognitive control of motivation.* Glenview, Ill.: Scott, Foresman, 1969.

Received March 7, 1978 ■

Postscript

[The] analyses sometimes suffer the imprint of someone convinced that he's right and willing to force his data to prove it. (Schachter, 1971, p. 141)

One function of replication studies is the independent assessment of the validity of conclusions drawn by investigators who sometimes force their data to fit their preexisting belief systems, thereby violating basic canons of scientific evidence. We began our research fully expecting to replicate Schachter and Singer's (1962) conclusion that inadequately explained, epinephrine-induced physiological arousal was readily manipulable into the disparate feeling states of euphoria and anger by varying available social cognitions. Our interest was to extend their analysis to the area of emotional pathology. After extensive testing, experimental variations, and procedural refinements designed to yield the predicted finding, we were forced to acknowledge the failure to replicate this now classic conclusion.

Obviously, there are many ways not to find an effect and few to demonstrate it. We now believe that both the basic paradigm and the methodology employed in the Schachter and Singer study and in our own may not be adequate to uncover the "emotional plasticity" predicted by their theory. But in addition, our data and that of Maslach lead us to propose the hypothesis that regardless of the experimental conditions utilized, awareness of one's strong arousal without adequate explanation is most likely to be interpreted as a negative mood state. Thus it may be that the extent to which states of unexplained physiological arousal are manipulable into disparate psychological mood states is limited to conditions of *mild* arousal (as Schachter and Singer's, 1979, current comments suggest, pp. 992–993. Some support for this constraint on the latitude of emotional plasticity comes from Nisbett and Schachter's (1966) inability to manipulate the emotion of subjects in a high-fear state, obtaining the misattribution effect only for subjects in a low-fear condition. Such clarification of the range or areas of applicability of a theory is another contribution of replication studies.

However, when Schachter and Singer's rejoinder now declares that *they* too "never found a difference," our failure to replicate becomes a replication of the null hypothesis and to some extent belittles our efforts. How could we—and, we presume, most other psychologists—have been led to misunderstand that the conclusions of the Schachter and Singer (1962) study were merely "tentative," based on nonsignificant treatment differences and internal analyses of questionable status? We were all led there by bold assertions to the contrary, by the rhetoric of conviction that Schachter and Singer chose to describe what they "really " found. "Conformation of Data to Theoretical Expectations" (p. 393) is their section heading, which belies any equivocation. It is important in this exchange to show briefly that both in their original article and in their present comments on our study and Maslach's, Schachter and Singer sometimes use their considerable literary talent to misdirect the reader's attention toward inadequately justified conclusions such that what was "never found" becomes one of the most cited findings in psychology textbooks.

Let us consider one instance of this misdirection. A one-sentence (cautiously worded) disclaimer of their findings can indeed be found: "However, the fact that we were forced, to some extent, to rely on internal analyses in order to partial out the effects of experimental artifacts inevitably makes our conclusions somewhat tentative" (Schachter & Singer, 1962, p. 396). This lean statement is sandwiched between an opening sentence that proclaims, "The pattern of data, then, falls neatly in line with theoretical expectations," and the following forceful conclusion:

> It has been suggested, first, that given a state of physiological arousal for which an individual has no explanation, he will label this state in terms of the cognitions available to him. This implies, of course, that by manipulating the cognitions of an individual in such a state we can manipulate his feelings in diverse directions. Experimental results support this proposition for following the injection of epinephrine, those subjects who had no explanation for the bodily state thus produced, gave behavioral and self-report indications that they had been readily manipulable into the disparate feeling states of euphoria and anger.
>
> From this first proposition, it must follow that given a state of physiological arousal for which the individual has a completely satisfactory explanation, he will not label this state in terms of the alternative cognitions available. Experimental evidence strongly supports this expectation. (pp. 395–396; reiterated verbatim in Schachter, 1971, p. 23)

We would like to propose that these conclusions should be recoded by psychologists as follows: There is *no* evidential basis for the empirical conclusion that epinephrine-aroused subjects can be readily induced to experience (or label that arousal as) euphoria or anger by manipulating available social cognitions (through a confederate's behavior).

Schachter and Singer's (1979) rejoinder cites the results of a study by Erdmann and Janke (1978) whose "results are consistent with ours and present none of the interpretative problems of the Marshall and Zimbardo and Maslach studies" (p. 992). However, they did not state what those "interpretative problems" are, nor did they mention several critical points about the Erdmann and Janke study. Unlike epinephrine, the drug used in the study (ephedrine) has strong central effects; the study failed to find any expected physiological differences, and the final conclusion advanced by the authors was the following:

> Regarding Schachter's theory the results and the conclusions of the present experiment imply that the relationship between emotional and physiological arousal might be less tight and/or less general than the theory proposes. (Erdmann & Janke, 1978, p. 73)

Dosage Levels, Source Awareness, and Plasticity

Since we found no differences between placebo controls and epinephrine-injected subjects, the drug dosage was increased (by tailoring level to body weight, a standard procedure in drug studies) to generate sufficient arousal differences for demonstrating possible emotionality effects of the manipulated social cognitions. Schachter (1971) led us to believe that emotionality was a direct, not a curvilinear, function of arousal by his reference to "the proposition that the degree of emotionality is directly related to the degree of physiological arousal" (p. 35). In their rejoinder, Schachter and Singer (1979) say that it is so obvious that the function is U-shaped as to be unworthy of mention. What may be worthy of mention are the parameters of that function, the specification of which should have been systematically investigated in humans. Having found weak effects on the subjects with their standard dose of epinephrine and the negative affect induced by our increased dosage (with inadequate explanation) does not justify their conclusion that our dose level put subjects on the other side of the curve. Clearly one cannot fit a U function with only two points. Moreover, it does not seem evident that the increase from .007 cc/kg (the equivalent of Schachter and Singer's .5 cc) to our .01 cc/kg would represent a vast difference on an inverted-U curve leading to performance reversal. We know of no empirical literature to support such an inference. The magnitude of the physiological reactions and self-reports of our increased arousal subjects were not qualitatively different from standard arousal subjects. (As can be seen in the data in our Table 1, the mean palpitation index was .9 for the standard dose subjects and only slightly higher, 1.4, for the increased arousal group.) In the animal literature cited by Schachter (1971), no differences were found for a 2:1 dosage ratio (p. 38), and to get dramatic debilitating effects the dosage ratio was pushed to 20:1.

Switching from rats to Schachter and Singer's (1979) phenomenological reports of their drug reactions, not one of our subjects ever reacted with anything resembling their near-to-death sensitivity. Of course, we and our graduate assistants pretested all drug doses on ourselves before administering them to any subject. The effects were remarkably variable. Marshall was hardly bothered by the .01-cc/kg dose, Zimbardo's experience was more marked, and the reactions of four others were arrayed from mild to strong.

When one's immediate physical reaction to a drug is intense, then the direct consequence is the perception that something is "very wrong" and the feeling is "lousy." In short, such a condition is not amenable to manipulation into positive moods by social cognitions—according to the current comments by Schachter and Singer (1979, p. 991). They state further that adrenaline alone produces a negative state when animals are in disturbing situations (pp. 991–992). Following their lines of argument, we would restate our position as follows: In unfamiliar settings (like those faced by the subjects in these experiments), the awareness of strong physical arousal symptoms in the absence of an adequate explanation creates a disturbing situation that is likely to be appraised in negative affect terms. Strong unexplained arousal states, then, should exhibit a negative bias, which is likely to be resistant to the blandishments of external social cognitions that are positive. That is a central conclusion of our research.

All of the evidence advanced by Schachter and Singer (1979) regarding the drug studies done in the 1920s and '30s (p. 994) is irrelevant both to our point and the necessary conditions for testing their two-factor theory of emotion. Subjects must find themselves aroused without adequate explanation in order for there to be a "bodily state in search of an appropriate cognition." In the studies by Marañon, Cantril and Hunt, and Landis and Hunt, cited by Schachter and Singer (1979), the subjects were fully aware of their arousal; they had a totally adequate explanation of its source in the injection they received. The "cold" or "as if" emotions shown by the majority of these subjects are not pertinent to our analysis, because we are interested in understanding the psychological experience of affect when particular somatic elements are present as part of an emotion, and not simply in a person's description of his or her somatic state.

It must be noted that *repeated* exposure to drug-induced arousal with a known physical source was typical for many of the Cantril and Hunt subjects. Such a condition is qualitatively different from those we are concerned with and have studied in the analysis of the consequences of unexplained arousal. Moreover, the "genuine-emotion subject" selected by Schachter and Singer (1979, p. 994) to illustrate that reactions to adrenaline can be pleasant, also reported "cold emotions" at another episode and also reported strongly negative consequences of arousal: "Extreme fear was present, but no con-

tent for it at the time. . . . I was intensely afraid" (p. 305).

Unfortunately, the bias of researchers willing to force the evidence to prove their point also is evident if one reads the original Landis and Hunt paper and Schachter and Singer's (1979) summary of it. Hysterical laughter does not tell us anything about the internal subjective state experienced by the institutionalized "lunatics" used as subjects. Landis and Hunt (1932) make quite explicit the difference between "outward and manifest expressive emotional reactions." Despite the overt hysteria, all cases labeled "emotion-subjective" were *negative*. This is exactly what Schachter and Singer would have predicted if they were using these data to support their earlier point about the negative effects of large doses of adrenaline. The Landis and Hunt subjects received much larger doses than our increased arousal subjects. Their reactions were quite variable but intense. One subject (classified "dementia praecox") reported feeling "pretty nervous" at 1.0 cc and laughed hysterically at 1.5 cc, while another in a manic phase had a crying spell at 1.0 cc and hysterical laughter at 1.5 cc.

This evidence again underscores the need for Schachter and Singer and all who work in the area of emotion to acknowledge the difference between public expressions of emotional reactions and the subjective experience of emotion. The two are rarely correlated in any of the studies on "emotional plasticity" or misattribution. The overt behavior may not reflect underlying affect at all, but rather conformity to situational contingencies. In our research and that of Maslach, many subjects faced with the antics of a euphoric confederate gave positive overt reactions, even laughed, but that behavior was independent of their self-reported affect, which was usually negative.

This same inconsistency shows up in Schachter and Singer's (1962) original data (the internal analyses used to salvage the study do not even present self-reported affect data) and is more glaring in Schachter and Wheeler (1962). The significant finding that is emphasized is that placebo subjects gave less overt laughter to a slapstick movie than did epinephrine subjects. However, on an overall amusement index, placebo subjects did *not* differ significantly from the epinephrine subjects (see their Table 3, p. 126), despite Schachter and Singer's (1979) fanciful relishing the contrary in their Footnote 3. Of greater import is the inability of Schachter and Wheeler to find any difference on self-reported reactions to the film. The means in their Table 4 are virtually identical for epinephrine, placebo, and chlorpromazine subjects: "funny," 4.1, 4.0, and 3.8, respectively; "enjoy it," 4.0, 3.9, and 3.8, respectively; "recommend it," 2.0, 1.9, and 1.8, respectively. All subjects reported that they felt the film was "mildly funny," they "enjoyed it a little," and would "recommend it moderately." This is hardly a pattern of data one would predict from Schachter and Singer's theoretical propositions.

Experimental Instructions

In response to an injection of epinephrine, "as far as the subject is concerned, the major subjective symptoms are palpitation, tremor, and sometimes a feeling of flushing and accelerated breathing" (Schachter & Singer, 1962, p. 382). Rothballer (1959), an expert on the effects of the drug, finds that tremor is "almost invariably" noted; palpitation is "the next most common finding" (p. 510). Thus, to mislead subjects about their anticipated reactions, symptoms *other* than these primary ones were mentioned as the side effects of their injection by both Schachter and Singer and us.

In the original Schachter and Singer (1962) study, the symptoms suggested were headache, numbness, and itching. In our replication we also mentioned the irrelevant symptom of headaches, but changed numbness to coolness and itching to dryness of mouth. Our reason for doing so was intentionally to include some minor, occasionally reported symptoms associated with epinephrine in order to increase the validity of the general instructions. We felt that if subjects experienced some weak effects that were attributed to the drug, they would be more likely to believe in the validity of the physician's instructions. This should have given them *less* reason to suspect that their major reactions (tremor and palpitations) were drug-related, since the credible communicator did not include them in his description of anticipated side effects.

In Schachter and Singer's (1962) study, the epinephrine–misinformed subjects tended to report *fewer* of the false symptoms, headaches, numbness, itching, than some or all of the other groups. In rejecting the symptoms specified by the physician, these subjects may have also doubted her competence. In a thorough critique of this feature of Schachter and Singer's work, Kemper (1978) suggests that the epinephrine–misinformed subjects were likely to have perceived the whole study as "unscientific" because of the totally wrong symptom expectations they

were given. Our instructions, thus, were an attempt to make the misinformation treatment function more appropriately in operationally evaluating the two-factor model of emotion.

Schachter and Singer (1979) argue that in doing so we did not properly misinform our subjects, and that is the reason we failed to find even a "shred of support for [their] original findings." They remind us that Schachter and Wheeler (1962) found a significant "dryness of mouth" effect. Examination of those data reveal that on a 5-point scale, the mean reported degree of dryness for epinephrine subjects was a scant .7, while placebo subjects were different by only .4 units. We believe that the dominant symptoms experienced by our subjects were unaccounted for and that the conceptual essence of the misinformation procedure is satisfied.

We would argue further that there are more substantial procedural differences between our study and Schachter and Singer's that readers should note. They dropped epinephrine subjects who reported "virtually no" palpitations or changes in heart rate before making their between-conditions analysis. This questionable subject selection excludes those who may have been physically inactive (activity alone will accelerate heart functions). These are the very subjects likely to have been focusing on their internal arousal rather than hula hooping and thus responding more negatively to their unexplained arousal (as we found). Another key difference may be that their subjects did not know prior to their appearance that they were to be injected. Ours consented to being injected before finding themselves in a fait accompli situation.

There are other differences that might be dredged up to account for our failure to replicate Schachter and Singer's (1962) findings—but sight is lost of the fact that we replicated their null effects only too well, despite many surface differences in procedures. On the Schachter and Singer affect index, the mean of their placebo group is 1.6, while ours is 1.5. The comparison between the two epinephrine–misinformed groups shows their mean is 1.9; ours is 1.4. Although they never report standard deviations, these small differences (.5 scale units on a 9-point scale) are well within sampling error range. Thus, we found virtually the same degree of inconsequential treatment difference as they did originally.

"The most pressing and crucial need with regard to Schachter and Singer results is a replication," states Kemper (1978, p. 174), who goes on to comment that "the credence with which the findings have been greeted is somewhat surprising in light of the absence of confirmatory results from another experimenter *in which exactly the same findings were sought*" (italics in original).

We indeed sought to replicate what we thought were the important results found by Schachter and Singer and could not. We tried to salvage something of value from an unsatisfying failure to replicate this classic by adding treatments that extended and tested some boundaries of their theory of emotion.

All the available data from our study, Maslach's, and Schachter and Singer's should be open to the close scrutiny of readers who by now must be wary of possibly being misinformed by the personal biases of the researchers involved in this argument. We hope some important clarifications have emerged from this exchange and apologize for occasionally replying in kind to personal attacks on our good sense and sensibility. In the future we will heed Oscar Wilde's (1908) admonition: "Arguments are to be avoided; they are always vulgar and often convincing" (p. 52).

Journal of Personality and Social Psychology
1979, Vol. 37, No. 6, 989-995

Comments on the Maslach and Marshall–Zimbardo Experiments

Stanley Schachter
Columbia University

Jerome E. Singer
Uniformed Services University of the Health Sciences, Bethesda, Maryland

This article addresses interpretive questions raised in the immediately preceding articles by Christina Maslach and by Gary D. Marshall and Philip G. Zimbardo.

First man: How's your wife?
Second man: Compared to what?
(H. Youngman)

The Maslach Study

On page 384 of our write-up (Schachter & Singer, 1962) of the epinephrine–emotion experiment we wrote, "Immediately [6] after the subject had been injected, the physician left the room and the experimenter returned with a stooge." Footnote 6 read:

> It was, of course, imperative that the sequence with the stooge begin before the subject felt his first symptoms for otherwise the subject would be virtually forced to interpret his feelings in terms of events preceding the stooge's entrance. Pretests had indicated that, for most subjects, epinephrine-caused symptoms began within 3–5 minutes after injection. A deliberate attempt was made then to bring in the stooge within 1 minute after the subject's injection.

The matter of timing is obviously crucial in all of these experiments. If the subject experiences palpitations, accelerated breathing, and so on *before* the emotion-inducing manipulation, there is obviously little possibility that he can "attribute" or "interpret" his physiological state in terms of the experimental manipulation and he must interpret it in terms of events preceding the manipulation (e.g., he is nervous about being in the experiment, excited about something that happened to him, getting sick, etc.).

In the crucial phase (Part 2) of Maslach's (1979) experiment, the sequence of events is as follows:

1. For roughly a 1-minute period the subject is exposed to a 15-word list on a memory drum.
2. There is a 1-minute recall test.
3. For the next minute, the subject is exposed to another 15-word list. The final word on this list is either the neutral word or the arousal word *start*.
4. There is a 1-minute recall test.
5. "Next, the subject and confederate were told to look through the materials in the folder and were then left alone" (p. 958). On the assumption that there was at least some brief description of what they were supposed to do with the "materials in the folder," this could have taken anywhere from a few seconds to a few minutes.
6. Only after they start looking through the folder does the confederate go into her act, and from the description, it seems that the first minute of this routine is essentially neutral, with little affective overtone.

Maslach's posthypnotic arousal suggestion was the following: "When you see the word *start*, your heart will beat faster, your breathing will increase, there will be a sinking feeling in your stomach, and your hands will get moist. You will feel all of these sensations as soon as you see the word *start*" (p. 958).

The word *start* appeared on the second word list, which was followed by a 1-minute recall test. Since Maslach uses performance on this recall test as proof that the posthypnotic suggestion has taken, it must follow that her subjects were in a state of arousal

Requests for reprints should be sent either to Stanley Schachter, Department of Psychology, Columbia University, New York, New York 10027 or to Jerome E. Singer, Medical Psychology, Uniformed Services University of the Health Sciences, Bethesda, Maryland 20014.

In the public domain

for some minutes before the confederate even began his or her act.

In addition, we note that this is not the first time during this experimental session that subjects were thrown into this state of marked physiological arousal with no explanation possible from the preplotted script of the confederate. To check on the effectiveness of the hypnotic induction of arousal, the first part of the experiment consisted of a private session with the subject during which he or she was hypnotized, the arousal suggestion was planted, the subject was exposed to the cue word *start*, and his or her physiological responses were measured. In short, twice during the experimental hour, a subject experienced, suddenly and with no explanation (if we credit the effectiveness of the amnesic suggestion), palpitations, speeded-up breathing, a sinking feeling in his or her stomach, and sweaty hands. The first time this happened the subject had not yet met the confederate; the second time it happened, the subject was working on a memory drum task and the confederate had not yet said a word.

Because of these matters of timing, Maslach's experiment and results have little to do with our experiment or with our hypotheses. The study does, however, unequivocally prove that when for no apparent reason, a subject's heart starts pounding, his or her breathing accelerates, palms turn sweaty, and stomach sinks, his or her mood is not likely to be affected by the irrelevant behavior of a confederate that follows after the appearance of this bodily state. The study also proves that subjects who, again for no apparent reason, find themselves in this bodily condition are likely to describe themselves as in a lousier mood than do subjects to whom this does not happen.

Footnotes notwithstanding,[1] within her experimental context there is only one way for Maslach to prove her position, and that is to run a set of conditions in which the word *start* appears a minute or two *after* the confederate starts his or her happy or angry routine and where there has been no prior posthypnotic arousal. If she is right, she will find the same results in these conditions as in those already run. If we are right, she will find that subjects in the happy condition are happy.

The Marshall and Zimbardo Study: 1. Dosage Levels

As to the Marshall and Zimbardo (1979) study, we find ourselves both bemused and perplexed by their strategy of replication. The simple fact is that their experiment replicates ours all too well. We never found a difference between the conditions they chose to replicate, and neither do they. Having found what we did, they then ran an additional condition with markedly increased doses of epinephrine and found that this produced a state of negative affect. Then ignoring virtually everything else that we and others have done, they conclude on the basis of this, almost their sole positive finding, that "epinephrine-related physiological arousal does not provide 'emotional plasticity,' but rather shows a consistent association with negative affect" (p. 983). As Maslach (1979) summarizes these data, "When the arousal was made stronger and more salient, subjects reported a *negative* emotional state, rather than the positive 'euphoria' predicted by Schachter and Singer" (p. 954).

In fact, Schachter and Singer would have predicted nothing of the sort. Maslach, Mar-

---

[1] Apparently in response to comments made by us in previous communications, Maslach discusses these issues in her Footnotes 7 and 10. Footnote 7 rejects the prior-arousal possibility by citing, without data, a pretest that could contain the faulty-timing artifact, and by citing an unpublished manuscript (Zimbardo & Maslach, Note 1). Footnote 10 raises three arguments against the faulty-timing objection. The first raises a new distinction—between normative and causal information—and claims that for causal information the confederate's actions must occur merely at the same time as the subject's arousal, not necessarily before the arousal. But that misses our point: The confederate need not start before the arousal but will certainly be of little effect if he or she starts after the arousal has occurred and a label has already been chosen. The reason for starting early is to assure that the unexplained arousal does not occur in the absence of a reasonable social context. Arguments 2 and 3 presuppose uncritical acceptance of the Marshall and Zimbardo article, while Marshall and Zimbardo presuppose uncritical acceptance of the Maslach article.

shall, and Zimbardo do not mention the fact, but we have been explicitly concerned with the effects of dose on emotional behavior, and in Schachter (1971), we reported experiments and reviewed evidence that small doses of epinephrine facilitate emotionally mediated behaviors in animals, whereas large doses interfere with such behavior. As dose increases, animals move less, get sick, and at large enough doses, die. Obviously we have done no counterpart to these dose experiments with humans, but it hardly seemed necessary, for in searching for a feasible dosage of epinephrine in our original experiment, we ourselves tried doses as heavy as, and heavier than, that used by Marshall and Zimbardo in their increased arousal condition. We would not do it again and can only assume that Marshall and Zimbardo never tried their increased dose on themselves or, if they did, that they are made of considerably stronger stuff than are we. At this dose, we did not have palpitations—our hearts pounded; we did not have tremors—we shook. We might have been convinced by someone that we were about to die, but no amount of social psychological tomfoolery could have convinced us that we were euphoric, or angry, or excited, or indeed anything but that something was very wrong and that we felt lousy. The point has always seemed too obvious to dignify with the status of the formal hypothesis that for humans, too, there is an inverted-U relationship between adrenaline dosage and the effectiveness of an emotion-inducing manipulation.

Other Research Evidence

We concede that we read Marshall and Zimbardo's (1979) description of our results with wistful envy. When they report that they were intrigued "to discover the reported ease with which the same condition of physiological arousal could be steered in as seemingly opposite affective directions as happiness or anger as a result of these cognitive manipulations" (p. 971), they certainly are not describing our experiment. The elaborate stage setting, the large cast, and the attention to exquisite nuances of dosage and timing yielded results requiring two internal analyses and three supplementary studies to make our point. Our conclusions about emotion were hardly based on this experiment alone, and we believe that there was no write-up of this study in which we did not make explicit what we believed was wrong with the experiment and did not point out the gross inadequacies of the data and then go on to describe the further research that did support our propositions. It was not that we were particularly humble about our work; we simply did not want to seem to be nitwits by basing a theoretical structure on obviously inadequate data.[2]

Since Maslach, Marshall, and Zimbardo either disregard or, to our mind, misinterpret this supporting material, we shall review this literature. First, Schachter and Wheeler (1962) examined subjects' reactions to a slapstick movie. In terms of quiet reactions, such as smiling, there were no differences between subjects injected with adrenaline and those injected with placebo. In terms of broad reactions such as laughter, epinephrine subjects laughed significantly more than did placebo subjects. As far as raucous reactions are concerned, 16% of the epinephrine subjects belly laughed; not a single placebo subject did so ($p < .01$)—hardly the pattern of findings one anticipates from an agent that produces a bias toward "negatively toned reports of affect."[3]

On negative emotions such as fear, we did a number of animal studies, of which the most relevant was Singer's (1963) demonstration that in neutral situations, an animal injected with adrenaline is no more disturbed or fearful than a control animal. In frightening situations, however, the adrenaline animal is mark-

---

[2] Relevant to the issue raised on pages 984–986 of the Marshall–Zimbardo Postscript, we suggest reading all of page 396 of our original article rather than the few sentences that Marshall and Zimbardo have selected from this page. Readers can then decide whether we have based our conclusions on this single experiment or whether we base them on the body of research that was generated by this study.

[3] Maslach also describes this experiment in her article. Anyone who relished Hastorf and Cantril's (1954) observations on the Princeton–Dartmouth game of 1951 should enjoy comparing these two descriptions to Table 3 and the pertinent paragraphs on pages 125–126 of the original article (Schachter & Wheeler, 1962).

Table 1
The Effects of Ephedrine and Placebo on Mood in Various Experimental Contexts[a]

Condition	Setting		
	Happiness	Neutral	Anger
Ephedrine	33.2	27.2	24.3
Placebo	28.5	30.7	30.8

Note. Higher scores indicate more pleasant ratings.
[a] From Erdmann and Janke (1978).

edly more frightened than a placebo animal. This is a result that in various forms has been replicated several times, most recently by Haroutunian and Riccio (1977). In and of itself, then, adrenaline does not produce a negative state in animals. Only in disturbing situations does it do so.

As to replications and related studies conducted since our group was working on these studies in the early 1960s, we simply have not followed the literature systematically, and in preparing these comments we have not had sufficient time to institute a thorough library search. However, what we have learned relevant to the issue of "emotional plasticity" seems to us to support our original findings. Though only we and Marshall and Zimbardo seem to have used epinephrine, on the whole, studies employing other sympathomimetic agents report findings consistent with our original findings. Erdmann and Janke (1978) relied on a well-disguised oral administration of the sympathomimetic agent ephedrine to manipulate arousal. They had an eight-condition experiment—four placebo conditions and four (in our terms) ephedrine–ignorant conditions (i.e., because of the mode of administration, subjects could have no idea of why they felt as they did). Crosscutting these conditions, Erdmann and Janke manipulated four emotional contexts: neutrality, happiness, anger, and anxiety. They measured the joint effects of context and arousal on self-reports of mood using both rating scales and a standardized adjective checklist. For the conditions relevant to this article, their results were as shown in Table 1 for the mood scale derived from the adjective checklist.

The three placebo conditions do not differ. With ephedrine, however, the conditions diverge, and in the predicted directions. The ephedrine–happiness group is more euphoric than the ephedrine–neutral group, which in turn is happier than the ephedrine–anger group. These results are consistent with ours and present none of the interpretive problems of the Marshall and Zimbardo and Maslach studies. Given the Maslach–Marshall–Zimbardo hypothesis that sympathomimetic agents bias people toward negative affect, it is of interest to note that only in the anxiety conditions did ephedrine fail to have an effect—a finding that Erdmann and Janke attribute to the ceiling effect.

We note, too, that our interactionist view of emotion has become something of a model for studies of a variety of psychopharmacological agents. For at least two agents, marijuana and alcohol, there have been studies that clearly demonstrate that whether or not these agents produce positive, negative, or neutral affective states depends on the interaction of the physiological effects of these agents with cognitive, social, and situational factors (Pliner & Cappell, 1974; Rossi, Kuehnle, & Mendelson, 1978). Obviously, all of the work on the effects of set on the action of drug agents makes much the same point (Wikler, 1957).

The Marshall–Zimbardo Study: 2. Experimental Instructions

Given this background material, it hardly seems an exotic hypothesis to repeat our suggestion of almost 20 years ago that adrenaline, too, in mild doses, produces a state of autonomic arousal that of itself is neither pleasant nor unpleasant but to a considerable extent takes on its affective coloration from the situation. How then can we account for Marshall and Zimbardo's results in their basic epinephrine and placebo conditions? In these conditions, unlike the increased arousal condition, their dose was the same as ours, yet consistently on self-report and activity measures their epinephrine subjects were somewhat unhappier than placebo subjects. Although, given the inadequacies of our experimental design, we do not expect significant differences between these conditions, at the very least

our line of thought would lead us to expect that, as in our experiment, the epinephrine–basic subjects would be slightly more, not less, euphoric than placebo–basic subjects.

Since the two experiments differ in numerous, possibly consequential, ways, it is obviously guesswork to try to understand why Marshall and Zimbardo fail to find even this shred of support for our original findings, but we have a pretty good guess. Marshall and Zimbardo attempted to replicate what we called the epinephrine–misinformed condition. The original instructions in this condition were:

> I should tell you that some of our subjects have experienced side effects from the Suproxin. These side effects are transitory, that is, they will only last for about 15 or 20 minutes. What will probably happen is that your feet will feel numb, you will have an itching sensation over parts of your body, and you may get a slight headache. (Schachter & Singer, 1962, p. 383)

None [4] of these symptoms, of course, are consequences of an injection, and in our study these instructions were deliberately designed to provide the subject with an inappropriate explanation of his bodily feelings.

For no reason we can fathom,[5] Marshall and Zimbardo decided to change our symptoms, and their instructions were the following:

> I should mention that although there are usually no side effects to the injection, a few of our subjects have reported some minor transitory reactions that may have resulted from the injection. These have typically included a feeling of dryness in the back of the mouth or throat, or a feeling of coolness in the hands or feet, and a couple of people have reported a slight headache. (Marshall, 1976, p. 13)

Unfortunately, both dryness of mouth and coolness in the hands or feet are consequences of an injection of adrenaline, and rather than having the counterpart of our epinephrine–misinformed condition, Marshall and Zimbardo to some extent have the counterpart of our epinephrine–informed condition. In the Schachter and Wheeler (1962) study, one of the questions asked to assess the effects of the drug manipulations was, "Does your mouth feel dry?" In Table 1 of that article it can be seen that subjects injected with epinephrine reported significantly ($p < .01$) more dryness than subjects injected with placebo. On the question of coolness of limbs, it is well known that adrenaline causes marked peripheral vasoconstriction, and several studies (e.g., Bierman, 1941; Wenger et al., 1960) have demonstrated that epinephrine causes a marked decrease (as much as 5 °C) in the temperature of the limbs. An injection of epinephrine will produce both a feeling of dryness in the mouth and a feeling of coolness in the hands and feet.

The Marshall and Zimbardo basic epinephrine condition can hardly, then, be considered the counterpart of our epinephrine–misinformed condition. If anything, it is somewhat comparable to our epinephrine–informed condition, in which the subjects were given an accurate description of their symptoms. This condition was designed to test the hypothesis that when subjects have an appropriate explanation for their physiological symptoms, they will be unlikely to label their feelings in terms of the alternative cognitions available. This proved to be the case, since subjects in the euphoria–epinephrine–informed condition of our experiment were less happy than placebo subjects. We suggest that this finding is inadvertently replicated in the Marshall–Zimbardo experiment.

Obviously, we are unconvinced that either the Maslach or the Marshall–Zimbardo experiments demonstrate what is claimed for them. We believe that the various experimental problems we have identified invalidate the conclusions drawn from these studies. Oddly enough, however, the simplest and most direct test of the major conclusion drawn by these

---

[4] The pharmacological texts (Goodman & Gilman, 1975) list "headache" as one of the effects of an injection of epinephrine. At the doses we used, this proved not to be the case, as can be seen in Table 1 of the Schachter and Singer (1962) article. In our pretests, however, headaches did sometimes accompany heavier doses of adrenaline.

[5] Relevant to Marshall and Zimbardo's postscript justification for these changes in procedure, we note that we have read Marshall's PhD dissertation carefully and nowhere in the 224 pages of text that is heavily concerned with differences between his study and ours is there any justification for changing the symptom description from "numbness to coolness and itching to dryness of mouth" (Marshall & Zimbardo, 1979, p. 987). In fact, nowhere in the entire text is this change even mentioned. These are simply referred to again and again as "irrelevant symptoms."

authors would bear little resemblance to either their experiments or ours.

Evidence on Emotional Plasticity

Clearly, Marshall, Maslach, and Zimbardo feel that their experiments are more than just a failure to replicate; they signal a fundamental deficiency in our model. Emotions, they claim, are not really so plastic. While there is no fundamental disagreement between us concerning the need for both arousal and an appropriate cognition, they suggest that the ultimate label of the physiological state associated with adrenaline is always likely to be unpleasant. Marshall and Zimbardo (1979) summarize their own and Maslach's work in these words, "The general picture that emerges . . . indicates that epinephrine-related physiological arousal does not provide 'emotional plasticity,' but rather shows a consistent association with negative affect" (p. 983).[6]

The test of the hypothesis that such arousal is unpleasant requires neither our cumbersome set of studies nor the equally complicated Marshall, Maslach, and Zimbardo replications. All it requires it that a number of people be injected with adrenaline and asked to describe how they feel—precisely the studies done in the 1920s and '30s by Marañon (1924), Cantril and Hunt (1932), and Landis and Hunt (1932). These studies are cited by Marshall and Zimbardo, but our reading of them differs considerably from theirs. They restrict themselves only to Marañon's subjects, who report "true" emotions and neglect the larger group who reported "cold" or "as if" ones. But the point at issue is not whether exogenous adrenaline produces an emotion rated in Marañon's context as genuine, but whether it produces a feeling of unpleasantness. The "cold" emotions, as far as we can determine from Marshall, Maslach, and Zimbardo's theory, are as pertinent to this point as any. This is especially true since Marañon suggests that "real" emotions following the injection are most likely to occur in those people with a prior endocrine disturbance or a history of affective disorder.

Marañon (1924) does not systematically categorize the reactions of his subjects, but from the various quotations throughout his article, it is evident that his subjects ran the full gamut of reactions from the negative "I feel as if I were afraid" to the exultant "something which comes up from the guts as when one is awaiting a great happiness" (p. 309). Cantril and Hunt (1932) find no evidence of a uniformly negative affective reaction to adrenaline. In their Table 1, entitled "Summary of the Results of the Experiment," one out of the four cases of genuine emotion was a pleasant one, and in 11 cases of reported affect, 3 were pleasant, 3 were unpleasant, and 5 were indifferent. The following self-description, from a genuine-emotion subject, certainly does not appear to us as the product of a bias toward negativism:

> Excitement (have felt like this in large class when I think I have a good idea, when raising my hand). No elation, but a background of elation. Feel light, as I used to in a burst of creative activity (writing). A pleasant state, possibly by association with such activity. Not unfamiliar. Used to attempt to bring on such states deliberately in adolescence. . . . During times of tremendous intellectual experiences in the past, times when my vision seemed peculiarly clear and far reaching, I have experienced similar reactions, but not even then so strongly. A tremendous illusion of power, vitality. (p. 303)

Landis and Hunt (1932) report results that may be unrepresentative, since they used institutionalized lunatics as subjects, but here, too, the authors report considerable variety in emotional response. Following the injection, three of their subjects had weeping fits, and three of their subjects broke into giggles or

[6] It is difficult to determine if Marshall and Zimbardo feel that the state of arousal induced by epinephrine is of itself sufficient to produce negative affect or if it requires an inadequate explanation as well. The quotation above suggests that the arousal state alone is sufficient; in other places they seem to indicate that an "inadequate explanation" is also required. In either case, their experiment has nothing to do with the question, for they do not manipulate the adequacy of the explanation. Our own results suggest that contrary to what we believe are Marshall and Zimbardo's expectations, providing a subject with an adequate explanation of his or her arousal state will, in euphoria conditions, worsen his or her mood. Our epinephrine-informed subjects were deliberately told what they would feel and why. Such subjects reported themselves as significantly *less* euphoric than subjects who had no adequate explanation for their bodily state.

hysterical laughter—hardly the uniformly negative affective tone required by the Marshall, Maslach, and Zimbardo formulation. Landis and Hunt conclude, "The adrenaline syndrome may be the basis for either pleasant or unpleasant emotional experiences" (p. 484).

To add our own bit, the sensation was for us curious and exhilarating—a feeling of emotional déjà vu. We had felt this way often in the past, but in this laboratory context it made no sense to feel this way—a pure arousal state uncluttered by passion or fear, affection or hatred. It was neither pleasant nor unpleasant; it was, though, absorbing enough so that at mild doses, we were delighted to serve as our own subjects again and again.

In these days of ethical guidelines and human subjects committees, this may very well be the end of the matter, for it is unlikely that anyone will do experiments such as ours or Marshall and Zimbardo's for quite a while, if ever again. On the particular issue at stake, however, this is probably of little moment, for this is one issue on which the readers can serve as their own subjects. If they will do a thorough introspective job after convincing a physician to inject them with .5 cc of a 1:1000 solution of epinephrine, they can decide which of us is right—or would this, too, require the approval of a human subjects committee?

Reference Note

1. Zimbardo, P. G., & Maslach, C. *Biased searches for causal explanations of experienced discontinuities.* Unpublished manuscript, 1979.

References

Bierman, W. Temperature change in the muscles of the human leg. In *Temperature: Its measurement and control in science and industry.* New York: Reinhold Press, 1941.

Cantril, H., & Hunt, W. A. Emotional effects produced by the injection of adrenaline. *American Journal of Psychology,* 1932, *44,* 300–307.

Erdmann, G., & Janke, W. Interaction between physiological and cognitive determinants of emotions: Experimental studies on Schachter's theory of emotions. *Biological Psychology,* 1978, *6,* 61–74.

Goodman, L. S., & Gilman, A. (Eds.). *The pharmacological basis of therapeutics* (5th ed.). New York: Macmillan, 1975.

Haroutunian, V., & Riccio, D. C. Effect of arousal conditions during reinstatement treatment upon learned fear in young rats. *Developmental Psychobiology,* 1977, *10,* 25–32.

Hastorf, A. H., & Cantril, H. They saw a game: A case study. *Journal of Abnormal and Social Psychology,* 1954, *49,* 129–134.

Landis, C., & Hunt, W. A. Adrenaline and emotion. *Psychological Review,* 1932, *39,* 467–485.

Marañon, G. Contribution à l'étude de l'action émotive d'adrénaline. *Revue Française d'Endocrinologie,* 1924, *2,* 301–325.

Marshall, G. D. *Affective consequences of "inadequately explained" physiological arousal.* Unpublished doctoral dissertation, Stanford University, 1976.

Marshall, G. D., & Zimbardo, P. G. Affective consequences of inadequately explained physiological arousal. *Journal of Personality and Social Psychology,* 1979, *37,* 970–988.

Maslach, C. Negative emotional biasing of unexplained arousal. *Journal of Personality and Social Psychology,* 1979, *37,* 953–969.

Pliner, P., and Cappell, H. Modification of affective consequences of alcohol: A comparison of social and solitary drinking. *Journal of Abnormal Psychology,* 1974, *83,* 418–425.

Rossi, A. M., Kuehnle, J. C., & Mendelson, J. H. Marihuana and mood in human volunteers. *Pharmacology, Biochemistry and Behavior,* 1978, *8,* 447–453.

Schachter, S. *Emotion, obesity, and crime.* New York: Academic Press, 1971.

Schachter, S., & Singer, J. Cognitive, social, and physiological determinants of emotional state. *Psychological Review,* 1962, *69,* 379–399.

Schachter, S., & Wheeler, L. Epinephrine, chlorpromazine, and amusement. *Journal of Abnormal and Social Psychology,* 1962, *65,* 121–128.

Singer, J. Sympathetic activation, drugs, and fear. *Journal of Comparative and Physiological Psychology,* 1963, *56,* 612–615.

Wenger, M. A., Clemens, T. L., Darsie, M. L., Engel, B. T., Estess, F. M., & Sonnenschein, R. R. Autonomic response patterns during intravenous infusions of epinephrine and nor-epinephrine. *Psychosomatic Medicine,* 1960, *22,* 294–307.

Wikler, A. *The relation of psychiatry to pharmacology.* Baltimore, Md.: Williams & Wilkins, 1957.

Received November 20, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 6, 996-1016

Bem Sex Role Inventory: A Theoretical and Methodological Critique

Elazar J. Pedhazur
New York University

Toby J. Tetenbaum
Fordham University

The rationale underlying the construction of the Bem Sex Role Inventory (BSRI), and the psychometric properties of the resulting instrument are scrutinized. Following a brief discussion of different types of sex role research, the results of two studies are presented and discussed. In Study 1, 1,464 graduate students rated the desirability of the BSRI traits for one of three referents: man, woman, or adult in American society. In Study 2, 571 graduate students used the BSRI for self-ratings. It was found that, regardless of the referents used, the "masculine" traits were relatively high in desirability but some of the "feminine" traits were low in desirability. Discriminant function analyses revealed that discrimination among groups was primarily due to the differential ratings of the two traits Masculine and Feminine for the different referents. Results from factor analyses of the ratings of desirability and the self-ratings indicated that (a) Bem's classification of the BSRI traits into masculine, feminine, and neutral is not tenable; (b) the dimensions that underlie desirability ratings differ from those that underlie self-ratings; and (c) the dimensions of self-ratings of males differ from those of females. The article concludes with a discussion of the implications of the findings for the measurement of androgyny.

Although sex roles have been a subject of study for some time, interest in them has greatly intensified as a direct result of the activities and writings dealing with, or emanating from, the women's liberation movement. (For a review of sex role research, see Hochschild, 1973, and, more recently, Ruble, Frieze, & Parsons, 1976.) Investigations of masculinity, femininity, and androgyny have proliferated, emphasizing such aspects as historical perspective (Bullough, 1973; Hunter, 1976; Taylor, 1973), development (Block, 1973; Block, Van der Lippe, & Block, 1973; Maccoby & Jacklin, 1974; Rebecca, Hefner, & Oleshansky, 1976), psychological and sociological correlates (Maccoby & Jacklin, 1974; Parsons, Ruble, Hodges, & Small, 1976), and paradigms for change (Bernard, 1975, 1976; Lipman-Blumen, 1973). Definitions of masculinity (e.g., Brannon, 1976; Pleck, 1975, 1976), femininity (e.g., Sherman, 1971, 1976; Steinmann & Fox, 1966), and androgyny (e.g., Bem, 1974, 1975) similarly abound.

There is a wide variety of instruments purportedly designed to measure masculinity and femininity—for example, the *Fe* scale of the California Psychological Inventory (Gough, 1957); the *Mf* scale of the Minnesota Multiphasic Personality Inventory (Hathaway & McKinley, 1943); the GAMIN M scale (Guilford & Zimmerman, 1949); and the Stereotype Questionnaire (Rosenkrantz, Vogel, Bee, Broverman, & Broverman, 1968). In an extensive review of such instruments, Constanti-

This research was supported by a Fordham University Faculty Research Grant. The computation for this study was supported by the Courant Institute of Mathematical Sciences at New York University, which is supported by the Department of Energy under Contract EY-76-C-02-3077.

We wish to thank Sheldon Kastner and Marvin Sontag for commenting on an earlier version of this article and for discussing various aspects of the study. Appreciation is extended to Gunārs V. Lucāns and Bert Holland for their assistance in computer programming.

Requests for reprints should be sent to Toby Tetenbaum, Department of Foundations, Psychology and Services, Fordham University, Room 1012F, 113 West 60 Street, New York, New York 10023.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3706-0996$00.75

nople (1973) concluded that they were inadequate on theoretical as well as psychometric grounds. In addition to questioning definitions of masculinity and femininity as operationalized by existing measures, Constantinople's primary criticism was directed at the assumptions of unidimensionality and bipolarity that seem to underlie the construction and use of masculinity–femininity measures. She concluded that the available evidence indicates that masculinity and femininity are neither bipolar nor unidimensional.

Reflecting similar concerns about the treatment of masculinity and femininity as antithetical, Bem (1974) constructed the Bem Sex Role Inventory (BSRI) to measure the two attributes as orthogonal dimensions, and also to yield an androgyny score. In the present article the rationale underlying the construction of the BSRI and its psychometric properties are scrutinized. It should be noted that much of what is said about the BSRI applies also to various other sex role measures currently in use. The BSRI was chosen because of its great popularity, and because, from a perusal of research studies in which it was used and from statements about it in recent textbooks (e.g., Oskamp, 1977; Wrightsman, 1977), it appears that its validity has been accepted without much critical questioning.

Hochschild (1973) has identified four broad types of research in sex roles. One type, for example, deals with sex roles as reflected in actual or attributed sex differences in personality traits, interests, and the like. Another type focuses on appropriate behaviors for males and females. Regardless of the relative merits of the different types of sex role research, failure on the part of a researcher to use an unambiguous definition of the type of sex roles being investigated is bound to result in the construction or use of measures of dubious validity, as well as in ambiguities in the conceptualization and execution of the research. Moreover, explicit delineations and specifications are necessary within each type of sex role research. For example, is one investigating role expectations or role enactments (Sarbin & Allen, 1968)? Are sex differences in traits studied normatively or prescriptively? (For a detailed discussion, see Spence, Note 1.)

Bem's work was done in the context of trait stereotyping. Consequently, it is important to bear in mind the potential problems and ambiguities associated with research in stereotyping. For example, some researchers use the "everyday" or "commonsense" conception of stereotypes, which is usually considered to be narrow and pejorative. Other researchers are primarily interested in one or more of the following aspects of stereotypes: their complexity, function, accuracy, resistance to change, to name but a few (Fishman, 1956; Vinacke, 1957). Bem offers no clear statement about the aspect, or aspects, of stereotypes in which she is interested, except to say that her concern is with positive traits only.

Also, it has been demonstrated that the format and the conditions under which stereotypes are elicited affect the results. For example, Erlich and Rinehart (1965) provided respondents with either a trait checklist or an open-ended format for the purpose of describing different ethnic groups. It was found that respondents given the checklist used a larger number of traits and exhibited greater consensus than did those who responded to the open-ended format. A probably more important finding was that the trait list obtained with the open-ended format was considerably different from the one used in the checklist format. Dealing specifically with stereotyping of a woman, Clifton, McGrath, and Wick (1976) have demonstrated that different traits were attributed depending on the role ascribed to her, for example, housewife, bunny, athlete.

With the above comments in mind, we shall proceed to examine the method used in the construction of the BSRI. Basically, judges were asked to rate the desirability of approximately 400 personality traits either for a man or for a woman in American society. Tests of significance of the difference between the mean ratings of each item were performed. Twenty traits rated significantly more desirable for a man than for a woman were designated as masculine, and 20 traits rated significantly more desirable for a woman than for a man were designated as feminine. In addition, 20 traits whose mean ratings of d

sirability for a man and for a woman did not differ significantly were designated as neutral. Having thus selected the traits, Bem proposed that the BSRI be used as a self-rating instrument, in which the respondent is asked to indicate on a 7-point scale, ranging from "never or almost never true" to "always or almost always true," the degree to which each trait describes himself or herself. "Masculinity equals the mean self-rating for all endorsed [sic] masculine items and Femininity equals the mean self-rating for all endorsed [sic] femininity items" (Bem, 1974, p. 158). Two additional scores can be obtained from the responses to the BSRI: an Androgyny score, which is essentially the discrepancy between the Masculinity and the Femininity scores, and a Social Desirability score, which is the mean self-rating on the 20 "neutral" traits. (For further details, see Bem, 1974.)

Of the various problems and issues related to the construction and use of the BSRI, only several will be noted at this stage. Instead of defining the domains of masculinity and femininity and attempting to construct measures consistent with the definitions, Bem has chosen a strictly empirical approach. While such an approach may have limited utility when the primary concern is with criterion-related validity, its appropriateness is very doubtful when the focus is on construct validity (e.g., Cronbach & Meehl, 1955; Travers, 1951). Regardless of one's viewpoint on this issue, however, it is important to examine some of the possible consequences and to indicate the basic requirements of the method of scale construction that Bem has used.

It is clear that Bem was studying the desirability of traits, or personality characteristics, for males and females. But since the meaning of the term *desirable* in the instructions was not clarified, it was open to different interpretations by respondents (e.g., normatively, prescriptively; see Strahan, 1975, and Spence, Note 1).

In the discussion of the rationale for the construction of the BSRI, the distinction between traits and behaviors is occasionally blurred, as is evidenced by the statement that "because the BSRI was founded on the conception of the sex-typed person as someone who has internalized society's sex-typed standards of desirable *behavior* for men and women, these *personality characteristics* were selected as masculine or feminine on the basis of sex-typed desirability" (Bem, 1974, p. 155; italics added).

The fact that 400 *t* tests were treated as if they were independent of each other seems to require no elaboration. It is, however, important to note that when relying solely on tests of significance one runs the risk of overlooking the distinction between statistically significant and substantively meaningful findings.

The criterion employed for the designation of traits as masculine, feminine, or neutral seems to be based on the assumption that traits used to characterize one of the sexes are not, or should not be, used to characterize the other sex. No justification is offered for this assumption, which appears to be untenable on the basis of research on ethnic or racial stereotyping (see, for example, Erlich & Rinehart, 1965). One possible consequence of the trait selection procedure is that it may have resulted in the inclusion of traits that are not desirable but, rather, *less undesirable* for one of the sexes. In sum, the method of trait selection for the BSRI is not unlike the method occasionally encountered in other areas (e.g., achievement testing) where exclusive reliance on the statistical characteristics of the items (e.g., difficulty, discrimination) for their selection may lead a test constructor to neglect the most important property of the test: its validity.

It is necessary to identify the dimensions that underlie the relations among the ratings of trait desirability, and those that underlie the relations among ratings of traits when they are used for self-ratings. The demonstration that each of the proposed subscales is unidimensional is necessary for their valid use as summated rating scales, regardless of whether they were designed to be consistent with theoretical definitions of the constructs or whether they were arrived at empirically (Green, 1954). No evidence about the dimensions underlying the BSRI is provided by Bem.

To provide support for and amplify some

of the observations made above, two studies were undertaken: The first dealt with the ratings of desirability of the BSRI traits, and the second was concerned with the properties of the BSRI when it is used as a self-rating instrument. In an attempt to replicate Bem's work as closely as possible, her format and instructions were used despite some of the reservations expressed above.

Study 1: Desirability Ratings

The 60-item BSRI was administered to 1,464 graduate students of education attending one of three universities in New York City. The instructions were those used by Bem (1974) in the trait-selection phase of her study; that is, the subjects were asked to rate the desirability of each trait for a man or a woman in American society. In addition, the present study sought ratings of desirability for an adult in American society.

Descriptive Statistics

Means and standard deviations of desirability ratings are reported in Table 1. Also included in the table, for comparative purposes, are mean likableness ratings of traits common to the BSRI and to trait lists used in studies by Anderson (1968) and Bryson and Corey (1977).

Several points will be noted. The traits labeled by Bem as masculine are generally perceived as desirable when applied to a man. All the item means are greater than 5 on a 7-point scale, ranging from "not at all desirable" (1) to "extremely desirable" (7). The means for the same traits, while still on the desirable end of the continuum, are generally lower when the traits are applied to an adult, and lowest when they are applied to a woman. Note, however, that the standard deviations associated with the trait-desirability ratings for a woman are generally larger than those for a man or an adult, indicating less agreement among respondents when the traits are applied to a woman.

Overall, the mean ratings for traits designated as feminine tend to be lower than those designated as masculine, even when the former are applied to a woman. This is also noted when contrasting the mean ratings of the "masculine" and "feminine" traits for an adult. More important, however, is the finding that some of the "feminine" items are perceived as relatively undesirable, or negative. See, for example, the means for the traits Shy, Gullible, and Childlike. Not only are these traits rated low in desirability when they are applied to a man or an adult, but their desirability is also low even when the referent is a woman. These findings cast doubt about Bem's (1974, 1975, 1977) statements that the "masculine" and "feminine" traits are positive, or socially desirable. In fact, Bem indicates that she included 20 "neutral" traits, 10 positive and 10 negative ones, as a means of checking whether the responses to the "masculine" and "feminine" traits are affected by a social desirability response set. The negativity of some of the "feminine" traits is underscored when their mean ratings are compared with the mean ratings of "neutral" traits designated by Bem as being negative. For example, the mean ratings of the "feminine" traits Gullible and Childlike are lower than the mean ratings of the "neutral" traits Theatrical, Unpredictable, Jealous, and Secretive.

About half of the traits on the Masculinity and Femininity subscales of the BSRI appear also on the list of traits rated for likableness in a study by Anderson (1968), who says: "Sex-linked words were . . . deliberately omitted from the list" (p. 277). The overlap between the two lists is actually greater when one considers not only identical words but also close synonyms. Rosnow, Wainer, and Arms (1969) indicate that certain of the traits used by Anderson are sex linked, and others are not. Among the latter are Loyal and Shy, which Bem designated as feminine traits, and Ambitious, which according to Bem is a masculine trait. The situation becomes more complicated as one considers additional studies of trait ratings. For example, Bryson and Corey (1977) asked subjects to indicate, in addition to ratings of likability, the sex-relatedness of a set of traits. They found that for some traits designated as masculine or feminine there were neither sub-

Table 1
Mean Desirability Ratings and Standard Deviations for the 60 BSRI Traits

Trait	Man[a]	Woman[b]	Adult[c]	Anderson (1968)	Bryson and Corey (1977)		
					Male	Female	General
1. Self-reliant	6.40 (.82)	4.45 (1.56)	6.06 (1.03)	4.62			
2. Defends own beliefs	5.81 (1.12)	4.43 (1.53)	5.46 (1.29)	—			
3. Independent	6.31 (1.00)	4.19 (1.70)	5.91 (1.16)	4.55	7.06	7.12	7.42
4. Athletic	5.57 (1.27)	3.54 (1.47)	4.72 (1.31)	—	6.69	3.93	6.67
5. Assertive	5.88 (1.13)	3.71 (1.63)	5.13 (1.32)	—			
6. Strong personality	5.91 (1.08)	4.08 (1.69)	5.32 (1.26)	—			
7. Forceful	5.38 (1.43)	3.11 (1.53)	4.47 (1.48)	—			
8. Analytical	5.09 (1.29)	3.68 (1.50)	4.63 (1.34)	—			
9. Has leadership abilities	6.04 (1.04)	3.95 (1.64)	5.58 (1.08)	—			
10. Willing to take risks	5.42 (1.15)	4.06 (1.53)	4.93 (1.24)	—	5.14	5.58	5.21
11. Makes decisions easily	5.47 (1.26)	4.29 (1.39)	5.13 (1.33)	—	6.83	6.64	6.77
12. Self-sufficient	6.22 (.95)	4.41 (1.63)	6.00 (1.03)	4.12			
13. Dominant	5.11 (1.65)	2.48 (1.52)	3.75 (1.65)	1.53			
14. Masculine	6.37 (1.02)	1.51 (1.05)	5.12 (1.58)	—	6.96	1.83	6.16
15. Willing to take a stand	5.80 (.98)	4.17 (1.60)	5.46 (1.12)	—			
16. Aggressive	5.45 (1.32)	2.97 (1.61)	4.48 (1.69)	3.04	6.08	5.17	5.56
17. Acts as a leader	5.80 (1.18)	3.53 (1.63)	5.28 (1.27)	—			
18. Individualistic	5.58 (1.15)	4.15 (1.61)	5.20 (1.42)	4.67			
19. Competitive	5.67 (1.27)	3.46 (1.54)	5.19 (1.32)	—			
20. Ambitious	6.08 (1.00)	4.03 (1.65)	5.69 (1.10)	4.84	7.26	6.92	6.98
21. Yielding	3.12 (1.46)	4.45 (1.49)	3.48 (1.33)	—			
22. Cheerful	4.85 (1.28)	5.85 (1.01)	5.20 (1.30)	5.04			
23. Shy	1.80 (1.03)	3.22 (1.44)	1.98 (1.01)	2.91	3.76	4.11	4.49
24. Affectionate	4.63 (1.39)	6.04 (.93)	4.83 (1.29)	—	6.94	7.19	7.52
25. Flatterable	3.40 (1.67)	4.63 (1.58)	3.32 (1.58)	—			
26. Loyal	5.46 (1.27)	5.95 (1.11)	5.63 (1.27)	5.47			
27. Feminine	1.39 (.96)	6.17 (1.18)	4.28 (1.76)	—	1.22	6.78	5.38
28. Sympathetic	4.48 (1.32)	5.77 (1.06)	4.99 (1.27)	4.59	6.42	6.72	6.67

Table 1 (*continued*)

Trait	Man[a]	Woman[b]	Adult[c]	Anderson (1968)	Bryson and Corey (1977) Male	Female	General
29. Sensitive to the needs of others	4.78 (1.39)	6.03 (.94)	5.45 (1.29)	—	6.42	6.97	6.72
30. Understanding	4.94 (1.34)	6.08 (.88)	5.48 (1.18)	5.49			
31. Compassionate	4.54 (1.42)	5.99 (.99)	5.15 (1.29)	—			
32. Eager to soothe hurt feelings	3.96 (1.46)	5.58 (1.27)	4.43 (1.41)	—			
33. Soft-spoken	2.85 (1.44)	4.80 (1.62)	3.31 (1.43)	3.80			
34. Warm	4.63 (1.42)	5.98 (1.02)	5.04 (1.27)	5.22	7.04	7.66	7.75
35. Tender	4.13 (1.52)	6.03 (1.02)	4.57 (1.43)	4.56	6.18	6.65	6.74
36. Gullible	1.53 (1.04)	2.50 (1.57)	1.56 (.93)	2.19	2.69	2.56	3.29
37. Childlike	1.59 (1.00)	2.79 (1.63)	1.80 (1.14)	—			
38. Does not use harsh language	3.42 (1.49)	4.63 (1.89)	3.90 (1.60)	—			
39. Loves children	4.69 (1.30)	5.98 (1.22)	5.01 (1.40)	—			
40. Gentle	4.17 (1.55)	6.00 (1.08)	4.66 (1.47)	5.03	6.50	6.90	7.33
41. Helpful	4.97 (1.26)	5.75 (1.01)	5.30 (1.22)	4.92			
42. Moody	1.70 (.97)	1.83 (1.25)	1.61 (.95)	1.82			
43. Conscientious	5.79 (1.05)	5.34 (1.18)	5.88 (1.15)	4.81			
44. Theatrical	2.65 (1.34)	3.46 (1.51)	2.82 (1.25)	2.69			
45. Happy	5.15 (1.31)	5.81 (1.07)	5.60 (1.28)	5.14			
46. Unpredictable	2.18 (1.39)	2.87 (1.62)	2.07 (1.32)	2.90			
47. Reliable	5.95 (.99)	5.69 (1.05)	6.04 (.94)	5.27			
48. Jealous	2.73 (1.45)	2.92 (1.66)	2.03 (1.27)	1.04	3.01	2.24	2.97
49. Truthful	5.32 (1.37)	5.67 (1.17)	5.47 (1.38)	5.45			
50. Secretive	2.71 (1.50)	3.04 (1.62)	2.58 (1.46)	—			
51. Sincere	5.10 (1.33)	5.88 (1.05)	5.53 (1.28)	5.73			
52. Conceited	2.23 (1.44)	2.17 (1.33)	1.90 (1.27)	.74	1.81	1.78	1.69
53. Likable	5.19 (1.27)	5.74 (1.08)	5.46 (1.28)	4.97			
54. Solemn	3.08 (1.38)	2.74 (1.49)	2.80 (1.40)	2.89			
55. Friendly	5.16 (1.19)	5.88 (.97)	5.52 (1.13)	5.19			
56. Inefficient	1.30 (.84)	1.83 (1.29)	1.26 (.77)	1.78			

(*table continued*)

Table 1 (*continued*)

Trait	Man[a]	Woman[b]	Adult[c]	Anderson (1968)	Bryson and Corey (1977) Male	Female	General
57. Adaptable	5.21 (1.21)	5.16 (1.21)	5.58 (1.18)	—			
58. Unsystematic	1.85 (1.17)	2.30 (1.38)	1.79 (1.16)	—			
59. Tactful	4.87 (1.19)	5.50 (1.16)	5.23 (1.21)	4.94			
60. Conventional	3.94 (1.50)	4.34 (1.60)	4.13 (1.56)	2.60			

Note. N = 1,464. BSRI = Bem Sex Role Inventory. Standard deviations appear in parentheses. To facilitate reading and interpretation, traits have been grouped by scale rather than as they appear on the BSRI. Thus, Traits 1–20 are considered by Bem to be masculine; Traits 21–40, feminine; and Traits 41–60, social desirability. Traits on the BSRI and Anderson were rated on a 7-point scale, whereas traits on Bryson and Corey were rated on a 10-point scale. Since the BSRI scale ranges from 1 to 7, whereas the Anderson scale ranges from 0 to 6, for purposes of comparability a constant of 1 was added to each mean on the Anderson data.

[a] n = 493.
[b] n = 426.
[c] n = 545.

stantively meaningful nor statistically significant differences between the mean ratings of their likability for males and for females. Examples of such traits are Conceited, which was classified as masculine, and Gullible, which was classified as feminine. Using Bem's criterion for trait selection, both traits would have to be classified as neutral. Parker (1969) provides a list of feminine traits, which includes the following: Helpful, Moody, Conscientious, Sincere, Friendly, and Conventional. These are among the traits classified as neutral by Bem. Applying Bem's criterion for the designation of traits as sex related to the results of the present study, one would have to conclude that 16 of the traits classified as neutral on the BSRI are actually sex related because the differences in the mean ratings of their desirability for a man and for a woman are statistically significant. If one were to add a criterion of meaningfulness, some of the traits would still have to be classified as sex linked. For example, using both criteria, the traits Helpful, Theatrical, Unpredictable, Sincere, Happy, Friendly, Tactful would have to be classified as feminine, whereas on the BSRI they are treated as neutral.

The examples above highlight some of the difficulties that may be encountered when trait selection is based solely on the criterion of statistical tests of significance. The power of the statistical test depends, among other things, on the size of the groups used. Consequently, given sufficiently large groups, any difference between their means may reach a prespecified level of statistical significance. The contradictory findings reported above may also serve to underscore the dangers in drawing conclusions on the basis of results obtained from "samples" of convenience. While this state of affairs is not unique to research in sex roles, it is unfortunate that some researchers in this area present their findings as "norms," and that other researchers proceed and interpret their own findings in relation to such "norms."

Discriminant function analysis. It will be recalled that Bem has performed about 400 univariate tests of significance in the trait-selection phase of her study. In the present study, the ratings of the desirability of the 60 BSRI traits for an adult, for a man, and for a woman were subjected to a stepwise discriminant function analysis. In addition, a separate analysis was done of the ratings for a man and for a woman, in order to determine the type of results Bem might have obtained

Table 2
Discriminant Analysis of Trait Desirability for a Man, for an Adult, and for a Woman

Group	Traits	Function	R_c^2	Λ	χ^2	*df*	% of trace	% correct classification	Centroids 1	Centroids 2
Man									2.34	.63
		1	.817	.183	2434.47	61	89.35			
Adult	60							88.9	.22	−.94
		2	.348	.652	612.13	59	10.65			
Woman									−2.99	.48
Man									2.27	−.37
	Masculine	1	.803	.197	2372.66	3	95.40			
Adult								87.0	.16	.57
	Feminine	2	.164	.836	261.23	1	4.60			
Woman									−2.83	−.30
Man									2.66	
	60	1	.892	.108	1974.13	60	100.00	98.3		
Woman									−3.08	
Man	Masculine								−2.55	
		1	.883	.117	1965.35	2	100.00	97.7		
Woman	Feminine								2.95	

Note. R_c^2 = squared canonical correlation. All the chi-squares are significant beyond the .001 level. *n*s = 493 (man), 545 (adult), and 426 (woman).

had she performed a multivariate analysis.[1] Despite our expectation that the major contribution to the separation among the groups would be made by the two traits Masculine and Feminine, no hierarchy for the inclusion of traits in the analyses was preestablished. The results of the analyses are summarized in Table 2.

It will be noted that while both functions are statistically significant, the first one accounts for the bulk of the differences among the three groups, namely 89% of the trace. From the centroids associated with the first function, it is evident the three groups are separated about evenly on this dimension, with the adult group being in the middle. The second function, on the other hand, reflects primarily differences between trait desirability for a man and for a woman, as contrasted with the adult group.

Among indices used to study the relative contribution of each of a set of variables to a discriminant function are the standardized coefficients and the structure coefficients, the latter being the correlations of each variable with the discriminant function (see, for example, Cooley & Lohnes, 1971). Although we prefer the use of structure coefficients, both types of indices will be discussed. Because of space limitations the entire functions and the structure coefficients associated with them will not be reported. Instead, it will be noted that while the standardized coefficients associ-

[1] Each of the groups was composed of male and female respondents. Except for a very small number for whom sex identification is missing, the breakdown was as follows: 196 males and 343 females rated the desirability of the traits for an adult; 147 males and 317 females rated the traits for a man; 142 males and 266 females rated the traits for a woman. Discriminant analyses between males and females for each of the conditions were performed in order to determine whether there were meaningful differences in ratings due to the sex of the respondents. In the three analyses the differences between male and female raters were small, warranting pooling the data across sex. One interesting exception will be noted. There was a meaningful difference between male and female ratings of the desirability of the trait feminine for an adult, the mean rating by males being 3.69, and that by females 4.63. This result may be contrasted with the mean ratings of the two groups of the trait masculine for an adult: 5.20 and 5.09 by males and females, respectively.

ated with the 2 traits Feminine and Masculine in the first discriminant function are −.60 and .73, respectively, the absolute values of the coefficients associated with 44 other traits are less than .05, the remaining 14 traits having coefficients whose absolute values range from .05 to .12. The structure coefficients associated with the traits Feminine and Masculine are −.642 and .731, respectively, while those associated with the remaining traits are small except for 9 which are around .30.

The first trait to enter in the stepwise analysis was Masculine. This was followed by the trait Feminine. The remaining 58 traits add very little to the discrimination among the groups. Specifically, the overall λ (Lambda) associated with the 60 traits is .119, whereas that associated with the 2 traits Masculine and Feminine is .165. Also, the functions based on the 60 traits provide 89% of correct classification into the three groups, as compared with 87% correct classifications provided by the functions based on the traits Masculine and Feminine only. Looking back at the mean ratings of desirability (see Table 1), it is noted that the differences between the ratings of Masculine and Feminine for a man and for a woman far exceed the mean differences of the ratings of the remaining traits. Specifically, the mean ratings of Feminine for a woman and for a man are 6.17 and 1.39, respectively, and those of Masculine for a man and for a woman are 6.37 and 1.51, respectively.

The results of the present study are even clearer when one considers the discriminant analysis based on the ratings of desirability for a man and for a woman only—the two conditions used by Bem. The λ associated with the 60 traits is .108, while that associated with the traits Masculine and Feminine is .117. The use of the function based on the 60 traits provides 98.3% correct classification in the two groups, as compared with 97.7% accomplished by using the function based on the traits Masculine and Feminine only. Clearly, it is these two traits that carry the major burden of the discrimination between the two groups.

Factor analysis. As was noted earlier, the selection of traits for the BSRI and their classification into one of three categories—masculine, feminine, and neutral—was accomplished on the basis of tests of significance of the differences between the mean ratings of the desirability of each trait when applied to a man and when applied to a woman. Even if one were to accept this procedure, it is important to bear in mind that mean differences reveal nothing about the dimensions that underlie the relations among the ratings of the traits. The justification for treating a set of traits as a separate summated rating scale rests on the assumption that it is unidimensional. To the best of our knowledge, no attempt has been made by Bem, or by other researchers, to study the dimensions that underlie the ratings of desirability of the traits included in the BSRI.

In order to study the dimensionality of the BSRI, three separate factor analyses of the intercorrelations among the desirability ratings of the 60 traits when applied to a man, to a woman, or to an adult were performed.[2] Squared multiple correlations were used as initial estimates of the communalities. Both varimax orthogonal and oblique (delta = 0) rotations were used. The two solutions were very similar, and the correlations among the factors in the oblique solutions were close to zero. Consequently, only the orthogonal solutions are reported.

The rotated factor structures for the three referents are reported in Table 3. Three interpretable factors were retained in each analysis. Attempts to extract additional factors resulted in excessive fragmentation. It should be noted, however, that the three factor solutions account for about 80%, 81%, and 72% of the common factor variance for a man, a woman, and an adult, respectively.

Using a factor loading $\geq .40$ as criterion for meaningfulness, it will be noted that Factor 1 for the man and the adult referents and Factor 2, the woman referent, are very similar. The same 13 "feminine" traits and 8 "neutral" ones have meaningful loadings on

[2] Separate analyses were first performed for males and for females in each condition. Since the results were virtually the same, pooled analyses across sex within each condition are reported.

this factor. Among the traits with the highest loadings in the three solutions are the following: Compassionate, Sensitive to the needs of others, Understanding, Sympathetic, Warm, Tender, Gentle, Sincere, and Friendly. All express concern with, and positive affect toward, others. Consequently, the factor is named Interpersonal Sensitivity.

As will be noted from Table 3, several traits have meaningful loadings in only one of the solutions, the most notable being the trait Feminine, with a meaningful loading only when the referent is a woman. This is not surprising when one considers that all the items with high loadings on this factor have been rated relatively high in desirability for the three referents, except for the trait Feminine, which was rated high in desirability only when applied to a woman. Considering that the trait Feminine does not have a meaningful loading on the analogous factors associated with a man and an adult, and that it has the lowest of all the meaningful loadings on Factor 2 for a woman, it would be inappropriate to treat the traits that have high loadings on this factor as "feminine" traits, or to name the factor Femininity.

The second set of factors with similar patterns of loadings is composed of Factor 2 for a man and an adult, and Factor 1 for a woman. All the "masculine" traits have meaningful loadings for a man. Seventeen of these traits have meaningful loadings for an adult. Among the three traits that have no meaningful loadings on this factor is the trait Masculine, which is also the only trait that has no meaningful loading on Factor 1 for a woman. This latter exception is particularly notable considering that the meaningful loadings for a woman on Factor 1 are generally much larger than the loadings of the same traits on the respective factors for a man and for an adult. Moreover, while the loadings for the remaining 40 traits are low on Factor 2 for a man and for an adult, there is a trend toward bipolarity in Factor 1 when the referent is a woman. Note, for example, the negative loadings on this factor for the traits Yielding, Flatterable, Soft-spoken, Childlike, Conventional. The relative strength of this factor in the three solutions can be discerned from its eigenvalue, which is 5.77 for an adult, 8.12 for a man, and 13.43 for a woman.

The traits with the highest loadings on this factor for the three solutions are the following: Assertive, Forceful, Has leadership abilities, Acts as leader, Ambitious, Strong personality. The factor may be named Assertiveness or Instrumentality. Note, again, that the trait Masculine has a meaningful loading only when the referent is a man. Consequently, it would be a misnomer to name this factor Masculinity. (See Johnson, 1975, for a discussion that questions the validity of equating dependence and expressiveness with femininity, and equating independence, instrumentality, and assertiveness with masculinity.)

The third set of factors with similar patterns of factor loadings is composed of Factor 3 in the three solutions. Two of the "feminine" traits, Gullible and Childlike, have meaningful loadings in the three solutions. In addition, the trait Shy has a meaningful loading for a woman, and Flatterable for a man. Of the 10 "neutral" traits that are negative, 8 have meaningful loadings on this factor in the three solutions. Among traits with the highest loadings are Jealous, Secretive, Conceited, Gullible, and Childlike, all of which are probably best captured by the label Immaturity.

In sum, there is no evidence that the traits on the BSRI comprise three subsets of Masculinity, Femininity, and Social Desirability. While most of the "masculine" traits comprise a single factor, it appears that it would be most appropriate to refer to it as Instrumentality or Assertiveness. Moreover, it should be noted that this factor reflects a relatively narrow domain, as it is composed of some traits that are almost synonymous and others that are very close in meaning (e.g., Has leadership abilities and Acts as a leader; Self-reliant, Self-sufficient, and Independent).

The remaining two factors are each composed of a mixture of "feminine" and "neutral" traits. The distinction between the two factors is that one is composed of traits considered generally desirable for the three referents, whereas the other consists of traits considered relatively undesirable for the three

Table 3
Trait Desirability for an Adult, for a Man, and for a Woman, Four-Factor Solution, Orthogonal Rotation

Trait[a]	Man factors			Woman factors			Adult factors		
	1	2	3	1	2	3	1	2	3
1. Self-reliant	.084	.541	−.111	.741	−.003	−.151	.208	.427	−.226
2. Defends own beliefs	.348	.400	−.018	.759	.044	−.138	.347	.301	−.094
3. Independent	.057	.575	−.059	.833	−.068	−.163	.153	.504	−.166
4. Athletic	−.144	.530	.209	.591	.083	−.052	.084	.334	.199
5. Assertive	−.083	.667	−.107	.847	−.090	−.060	−.073	.628	−.042
6. Strong personality	.018	.688	−.014	.795	−.004	−.048	.136	.587	.082
7. Forceful	−.258	.658	.082	.783	−.098	.007	.098	.606	.085
8. Analytical	.148	.402	.060	.737	.071	.080	.251	.406	.028
9. Has leadership abilities	.069	.699	−.012	.871	−.059	−.100	.062	.630	−.092
10. Willing to take risks	.115	.503	.070	.710	−.006	−.044	.173	.406	.136
11. Makes decisions easily	−.012	.579	.070	.634	.105	−.129	.128	.454	−.037
12. Self-sufficient	.173	.619	−.069	.835	.030	−.146	.229	.539	−.245
13. Dominant	−.382	.561	.169	.666	−.204	.188	−.273	.541	.261
14. Masculine	−.109	.540	.049	.341	.245	.100	−.046	.326	.030
15. Willing to take a stand	.261	.666	.056	.837	−.026	−.082	.295	.542	−.058
16. Aggressive	−.273	.656	.190	.717	−.153	.154	−.266	.552	.196
17. Acts as a leader	−.060	.728	.031	.817	−.063	−.040	.065	.633	−.003
18. Individualistic	.231	.458	.046	.779	.011	−.139	.221	.434	.001
19. Competitive	−.228	.675	.119	.781	−.085	.045	−.184	.564	.134
20. Ambitious	.006	.648	−.021	.834	−.038	−.075	.036	.608	.058
21. Yielding	.471	−.176	.163	−.369	.260	.272	.231	−.099	.234
22. Cheerful	.602	−.012	−.001	−.103	.661	.072	.579	.085	.027
23. Shy	.260	−.193	.404	−.309	.246	.439	.274	−.158	.361
24. Affectionate	.733	−.083	.064	−.110	.632	.120	.642	−.000	.063
25. Flatterable	.187	.140	.405	−.367	.330	.378	.116	.023	.346
26. Loyal	.514	.232	−.163	−.124	.639	.013	.532	.097	−.096
27. Feminine	.234	−.244	.177	−.304	.479	.065	.203	.020	.054
28. Sympathetic	.764	−.083	.021	−.134	.658	.016	.724	.006	−.022
29. Sensitive to the needs of others	.846	−.118	−.001	.092	.700	.045	.746	.032	−.108
30. Understanding	.828	−.034	−.064	.012	.685	.012	.754	.026	−.047
31. Compassionate	.816	.098	.006	−.088	.722	−.072	.814	.010	−.063
32. Eager to soothe hurt feelings	.738	−.007	.227	−.254	.660	.146	.690	.038	.209
33. Soft-spoken	.454	−.159	.208	−.383	.488	.268	.408	−.067	.320
34. Warm	.773	−.113	.093	−.128	.701	.044	.791	.010	.101
35. Tender	.768	−.102	.136	−.161	.692	.153	.744	−.016	.141
36. Gullible	.012	−.205	.564	−.340	.044	.614	−.019	−.142	.590
37. Childlike	.160	−.177	.468	−.391	.120	.509	.049	−.105	.455
38. Does not use harsh language	.397	.016	.158	−.235	.230	.144	.261	.078	.083
39. Loves children	.643	.069	.149	−.303	.572	.187	.533	.038	.164
40. Gentle	.791	−.099	.083	−.121	.693	.133	.753	−.030	.161
41. Helpful	.654	−.020	.042	−.027	.604	.077	.632	.043	−.091
42. Moody	.021	.075	.541	.096	−.127	.480	−.002	−.026	.450
43. Conscientious	.354	.377	−.098	.415	.311	−.010	.444	.290	−.251
44. Theatrical	.212	.030	.474	.085	.205	.427	.133	.162	.448
45. Happy	.566	.106	−.185	.114	.599	−.039	.533	.090	−.017
46. Unpredictable	.023	.005	.564	.009	−.009	.554	.026	.081	.516
47. Reliable	.419	.349	−.203	.319	.541	−.130	.547	.297	−.221
48. Jealous	−.152	.243	.508	−.262	.101	.635	−.102	.059	.603
49. Truthful	.681	.037	−.142	.253	.485	−.138	.656	.019	−.123
50. Secretive	−.096	.197	.512	−.010	.063	.601	−.125	.158	.512
51. Sincere	.798	.020	−.092	.136	.722	.090	.773	.039	−.117
52. Conceited	−.336	.186	.497	.067	−.083	.579	−.248	.176	.558

Table 3 (*continued*)

Trait[a]	Man factors			Woman factors			Adult factors		
	1	2	3	1	2	3	1	2	3
53. Likable	**.525**	.230	−.035	−.075	**.644**	.066	**.436**	.138	−.013
54. Solemn	.060	.190	**.405**	.174	.036	.338	.180	.123	**.419**
55. Friendly	**.651**	.168	−.025	.113	**.655**	.054	**.676**	.055	−.027
56. Inefficient	−.015	−.093	**.532**	−.234	−.109	**.606**	−.113	−.153	**.551**
57. Adaptable	**.453**	.148	−.079	.201	.346	−.112	.319	.250	−.113
58. Unsystematic	.006	−.062	**.453**	−.201	−.033	**.521**	.042	−.047	**.430**
59. Tactful	**.618**	.187	.065	.104	**.535**	−.111	**.524**	.140	.002
60. Conventional	−.113	.149	.239	**−.413**	.207	.308	−.037	.089	.178
λ	11.699	8.116	3.757	13.431	9.592	4.257	10.377	5.772	3.960

Note. Factor loadings ≥.40 are in boldface. *ns* = 545 (adult), 493 (man), and 426 (woman).
[a] To facilitate reading and interpretation, traits have been grouped by scale rather than presented as they appear on the Bem Sex Role Inventory. Thus, Traits 1–20 are considered by Bem to be masculine; Traits 21–40, feminine; and Traits 41–60, neutral.

referents. The first was named Interpersonal Sensitivity, and the second, Immaturity.

Basically, the same factor structure is apparent in the three solutions, indicating that the crucial aspects in the rating of desirability of the traits is not the referent to which they are applied, but the respondents' conceptions of which traits are part of the same domain. A plausible explanation of the results could be made within the context of implicit personality theory (Schneider, 1973). The specific theory and the specific names of the factors notwithstanding, the results of this study indictate that it is inappropriate to treat the 20 "feminine" traits and the 20 "neutral" ones as two separate unidimensional sets.

Finally, it is obvious that one cannot obtain from a factor analysis more than one puts into it. Bearing in mind the approach in the selection of traits for the BSRI, it is not surprising that various dimensions and traits considered by other researchers as aspects of the domains of masculinity or femininity do not appear at all on the BSRI (see, for example, Bryson & Corey, 1977; Parker, 1969).

Study 2: Self-Ratings

This study was designed to investigate the pattern of responses to the BSRI traits, and the dimensions underlying the relations among them, when the instrument is used for self-rating. It should be noted at the outset that since the BSRI is a self-report inventory, it is subject to the limitations and deficiencies associated with such inventories, particularly those designed to measure self-concept (see, for example, Edwards, 1970; Messick, 1963; Wiggins, 1973; Wylie, 1968, 1974).

Only the 20 "masculine" and the 20 "feminine" traits were used in this study, because that is the form most often used in research related to sex roles, and because the dubious nature of the "neutral" traits was amply demonstrated in Study 1. A total of 571 graduate students of education attending one of two universities in New York City responded to the BSRI using Bem's directions.

Descriptive Statistics

Means and standard deviations for the 40 items for males and females are reported in Table 4. The most important thing to be noted is the general similarity of the mean ratings for males and for females. Except for the very large differences in the mean ratings of the traits Masculine (5.89 and 2.02 for males and females, respectively) and Feminine (5.85 and 2.09 for females and males, respectively), and the difference of 1.08 between the mean ratings of the trait Athletic, all other differences between the mean ratings of males and females are small, ranging (in absolute value) from .01 to .47, with a median of .16. The similarity of the ratings can also

Table 4
Mean Self-Ratings and Standard Deviations for the 20 Masculine Traits and 20 Feminine Traits of the BSRI for Males and for Females

Trait	Males		Females	
	M	*SD*	*M*	*SD*
1. Self-reliant	5.82	.94	5.96	.92
2. Defends own beliefs	5.63	1.11	5.79	1.04
3. Independent	5.79	.91	5.87	.95
4. Athletic	4.66	1.83	3.58	1.75
5. Assertive	4.93	1.14	5.04	1.18
6. Strong personality	5.39	1.19	5.40	1.19
7. Forceful	4.84	1.23	4.85	1.26
8. Analytical	5.61	1.02	5.29	1.29
9. Has leadership abilities	5.49	1.13	5.22	1.31
10. Willing to take risks	4.99	1.22	4.84	1.29
11. Make decisions easily	4.89	1.24	4.49	1.40
12. Self-sufficient	5.71	.89	5.77	.99
13. Dominant	4.76	1.15	4.57	1.36
14. Masculine	5.89	1.18	2.02	1.23
15. Willing to take a stand	5.60	1.02	5.46	1.11
16. Aggressive	4.58	1.16	4.30	1.39
17. Acts as a leader	5.06	1.17	4.74	1.33
18. Individualistic	5.56	1.05	5.47	1.14
19. Competitive	5.14	1.28	4.69	1.51
20. Ambitious	5.60	1.19	5.57	1.28
21. Yielding	4.44	1.10	4.29	1.16
22. Cheerful	5.47	.93	5.60	.98
23. Shy	3.75	1.41	3.69	1.34
24. Affectionate	5.42	1.09	5.74	1.07
25. Flatterable	4.46	1.41	4.32	1.41
26. Loyal	6.18	.86	6.37	.79
27. Feminine	2.09	1.25	5.85	1.06
28. Sympathetic	5.73	.93	5.94	.96
29. Sensitive to needs of others	5.92	.89	6.09	.86
30. Understanding	5.85	.76	5.98	.81
31. Compassionate	5.77	.96	5.91	.91
32. Eager to soothe hurt feelings	5.50	1.11	5.65	1.08
33. Soft-spoken	4.71	1.36	4.41	1.56
34. Warm	5.53	.95	5.78	.92
35. Tender	5.17	1.09	5.62	1.04
36. Gullible	3.15	1.35	3.62	1.55
37. Childlike	2.98	1.29	3.19	1.43
38. Does not use harsh language	4.43	1.66	4.44	1.81
39. Loves children	5.71	1.20	5.96	1.12
40. Gentle	5.51	.97	5.71	1.08

Note. *ns* = 171 (males) and 400 (females). BSRI = Bem Sex Role Inventory. Traits 1–20 are considered by Bem to be masculine and Traits 21–40, feminine.

be seen when one compares the overall mean ratings for males and females on Masculinity and Femininity, in which the traits Masculine and Feminine are included, and mean ratings in which those two traits are excluded. When the trait Masculine is included, the overall mean ratings of Masculinity for males and females, respectively, are 5.30 and 4.95, a difference of .35. When the trait Masculine is excluded, the respective means for males and females are 5.27 and 5.09, a difference of .18, or about half of the original difference. The overall mean ratings of Femininity, with the trait Feminine included, are 5.21 and 4.89 for females and males, respectively, the difference being .32. When the trait Feminine is excluded, the means for females and males are 5.17 and 5.03, respectively, a difference of .14, or less than half of the original difference.

From the overall mean ratings, it is clear that, on the average, both males and females indicate that "it is often true" (a rating of 5) that the traits in question describe them. This is even more pronounced when one excludes from consideration the traits Masculine and Feminine and the few traits that have been shown to be low in desirability (e.g., Gullible, Childlike). Since most of the traits have been rated toward the desirable end of the continuum when applied to an adult (see Table 1), it is conceivable that both males and females respond to the BSRI in a socially desirable manner or, as Smith (1968) has put it, with a "bias toward thinking as well of oneself as one can get away with" (p. 368). The available data cannot shed further light on this issue. It will be noted, however, that in a study by Parker and Veldman (1969), the traits Cheerful, Loyal, Sympathetic, Warm, and Gentle are part of a Social Desirability factor. Also, except for one or two traits, all the traits common to the BSRI and the Anderson (1968) list have been rated high on likability in several studies.

Lest one think that the relatively high mean ratings are unique to the present study, it is noted that other researchers have obtained similar results in the mean ratings of Masculinity and Femininity (cf. Bem, 1974, 1977). Since neither Bem nor other research-

ers report mean ratings on individual traits, more detailed comparisons are precluded.

Discriminant function analysis. Using overall mean ratings, Bem (1974) demonstrated that males score significantly higher than females on Masculinity, and that females score significantly higher than males on Femininity. If, as it appears, this was intended to lend support to the validity of the scales, several problems ensue, the most important being that Bem's contention that a person may be both masculine and feminine seems incompatible with the use of sex differences on Masculinity and Femininity as evidence of scale validity. This is particularly problematic when one uses groups that may be nonrepresentative of any population of males and females.

In any case, since Bem does not provide any information about mean ratings of individual traits, it is not possible to tell how much of the difference between males and females was due to the differences in the self-ratings on the traits Masculine and Feminine. It was shown above that the ratings on these two traits accounted for about half of the mean differences between males and females on Masculinity and on Femininity.

To further illustrate the relative potency of the traits Masculine and Feminine and that of the remaining 38 traits to discriminate between males and females, the responses of the two groups to the 40 traits were subjected to a stepwise discriminant function analysis. The findings are summarized in Table 5. Note that the λ associated with the 40 traits is .212 and the λ associated with the 2 traits Masculine and Feminine only is .241. Stated differently, the R^2 of the regression of group membership (male–female coded as a dummy variable) on the 40 traits is .788, while that on only the 2 traits Masculine and Feminine is .759. While 96.7% of the males and females are correctly classified on the basis of the full function, 93.5% of them are correctly classified by using the function based on the traits Masculine and Feminine. To further highlight the findings, it will be noted that the standardized coefficients associated with the traits Feminine and Masculine in the full function are −.70 and .55, respectively. Of the remaining traits, 32 have coefficients of <.10, and six have coefficients between .10 and .15. Probably even more revealing are the structure coefficients, which for the traits Feminine and Masculine are −.80 and .76, respectively. Except for the trait Athletic, whose structure coefficient is .14, and the trait Tender, whose structure coefficient is −.10, all the remaining traits have structure coefficients of <.10, with most being close to zero.

In sum, all the indices strongly support the notion that the discrimination between males and females is almost exclusively due to their self-ratings on the two traits Masculine and Feminine. Stated differently, knowledge of respondents' self-ratings on the remaining 38 traits adds little to the knowledge obtained from their ratings on Masculine and Feminine. Not surprisingly, males rate themselves high on Masculine and low on Feminine. The converse is true for females.

Factor analysis. It was stated earlier that we have not found any published reports of factor analyses of the desirability ratings of the BSRI traits. We did find one published study in which self-ratings on the BSRI were

Table 5
Discriminant Analysis of Trait Self-Ratings for Males and for Females

Traits	R^2	Λ	*df*	χ^2	% correct classification	Centroids
40	.788	.212	40	850.47	96.7	−1.26 2.94
Masculine	.759	.241	2	808.24	93.5	−2.71
Feminine						1.16

Note. *ns* = 171 (males) and 400 (females).

factor analyzed (Gaudreau, 1977). This study suffers from several methodological shortcomings. Not only were males and females used in a single analysis but also the subjects came from widely diverse groups (i.e., clerical workers, supervisors, middle managers, executives, police officers, and housewives). Warnings against inducing spurious correlations among variables by using heterogeneous groups have been sounded frequently throughout the brief history of social science research. Spearman (1904), for example, criticized researchers who "have purposely thrown together subjects of all sorts and ages, and thus have gone out of their way to invite fallacious elements into their work" (p. 223). Guilford (1952) elaborated on the dangers of using diverse groups in a single-factor analysis.

Another flaw in Gaudreau's (1977) analysis is that in addition to the individual traits, she included the scores on Masculinity, Femininity, and Androgyny (i.e., the discrepancy between Masculinity and Femininity), thus introducing three linear and experimental dependencies that may lead to artifactors rather than factors (see, for example, Gorsuch, 1974, pp. 267–268; Guilford, 1952). Because of these problems Gaudreau's findings will not be discussed.

In the present study, separate principal-axes factor analyses were performed for the male and female groups. Squared multiple correlations were used as initial estimates of the communalities. Both varimax orthogonal and oblique (delta = 0) rotations were used. Since the solutions were very similar, only the orthogonal ones are reported.

Four interpretable factors emerged from each analysis, accounting for about 83% and 73% of the common-factor variance for females and males, respectively. Extraction of a fifth factor for the males resulted in accounting for an additional 5% of the common-factor variance, but in a less interpretable factor structure. The two orthogonally rotated four-factor solutions are reported in Table 6.

Looking first at the factor structure for females and treating loadings of $\geq .40$ as meaningful, 17 traits load meaningfully on Factor 1. All but one are what Bem classifies as masculine traits, the exception being the "feminine" trait Shy, which loads negatively on this factor. The highest loadings are associated with the following traits: Assertive, Forceful, Acts as leader, Has leadership abilities, Strong personality, Dominant, and Aggressive. It is clearly an assertiveness factor. Factor 2 has 12 traits with meaningful loadings, all of which are classified as feminine by Bem. The traits with the highest loadings are Compassionate, Tender, Understanding, Sympathetic, Sensitive to the needs of others, and Gentle, all reflecting interpersonal sensitivity. Factor 3 is bipolar with three traits (Self-reliant, Independent, and Self-sufficient) loading positively, and two traits (Gullible and Childlike) loading negatively. This cluster of traits appears to reflect one's feelings of self-sufficiency; that is, individuals who perceive themselves as being self-sufficient, or relying on their own resources, quite naturally tend also to perceive themselves as not being gullible and childlike. Factor 4 is bipolar with only two traits having meaningful loadings on it. Feminine, with a positive loading, and Masculine, with a negative loading. Bem's contention that Masculinity and Femininity are orthogonal dimensions notwithstanding, females who rate themselves high on Feminine tend to rate themselves low on Masculine, regardless of their self-ratings on the remaining traits.

Compared with the females, the males tend to make a greater and stronger differentiation between assertiveness and self-sufficiency as reflected by Factors 2 and 3, respectively. Note that, as in the solution for females, the trait Shy has a negative loading on the Assertiveness factor. Also, Loyal, which is presumably a feminine trait, has a meaningful loading on the Self-sufficiency factor. Factor 1 for males is similar to Factor 2 for females, except that the loadings in the former tend to be higher than in the latter (eigenvalues of 5.46 and 4.90, respectively). Eleven of the "feminine" traits have meaningful loadings on this factor, which clearly reflects the notion of interpersonal sensitivity. As in the solution for the females, Factor 4 is bipolar. But it differs from the corresponding factor for females in that, in addition to the trait Masculine loading positively and the trait Feminine

Table 6
Masculine and Feminine BSRI Traits for Females and for Males, Four-Factor Solution, Orthogonal Rotation

Trait[a]	Female factors				Male factors			
	1	2	3	4	1	2	3	4
1. Self-reliant	.360	.135	**.492**	.007	.043	.152	**.524**	.099
2. Defends own beliefs	**.506**	.176	.106	−.133	.060	.192	**.422**	.085
3. Independent	**.442**	.098	**.480**	−.049	.033	.247	**.657**	.053
4. Athletic	.168	−.008	.018	−.192	.035	.178	.086	.350
5. Assertive	**.770**	−.015	.015	−.073	.027	.680	**.451**	−.042
6. Strong personality	**.689**	.044	.058	.005	.064	**.632**	**.466**	.013
7. Forceful	**.768**	−.016	−.011	−.047	.002	**.655**	**.377**	.145
8. Analytical	**.363**	.006	.178	−.037	.162	.151	.177	.097
9. Has leadership abilities	**.731**	.096	.127	.051	.287	**.568**	**.382**	.149
10. Willing to take risks	**.497**	.075	.036	−.071	.214	**.367**	**.405**	.109
11. Makes decisions easily	**.464**	.145	.236	−.009	.058	**.360**	.251	.171
12. Self-sufficient	**.468**	.109	**.460**	−.119	.040	.187	**.597**	.110
13. Dominant	**.687**	−.080	−.106	−.128	.004	**.648**	**.289**	.074
14. Masculine	**.133**	−.265	−.064	−.507	.130	.101	.059	**.416**
15. Willing to take a stand	**.615**	.144	.122	−.071	.215	**.315**	**.588**	.115
16. Aggressive	**.674**	−.151	−.190	−.177	.101	**.667**	**.227**	.048
17. Acts as a leader	**.738**	.108	.086	.075	.213	**.627**	**.366**	.090
18. Individualistic	**.519**	.216	.217	−.197	−.029	.181	**.431**	.073
19. Competitive	**.506**	−.091	−.081	.233	.219	.398	.078	.340
20. Ambitious	**.502**	.083	−.045	.168	.301	**.533**	−.062	.109
21. Yielding	−.262	**.340**	−.093	.083	.249	−.203	−.010	−.024
22. Cheerful	.192	**.423**	.174	.007	**.426**	.234	.188	.205
23. Shy	−.422	−.105	−.128	−.021	.055	−.400	−.102	−.007
24. Affectionate	.187	**.545**	−.195	.146	**.646**	.280	−.065	−.116
25. Flatterable	.134	.026	−.339	.183	.276	.075	−.188	−.215
26. Loyal	.088	**.396**	.081	−.008	.199	.126	**.422**	.092
27. Feminine	.120	**.447**	−.000	**.511**	−.065	−.056	.068	**−.748**
28. Sympathetic	.080	**.689**	−.039	.024	**.702**	.037	.117	.062
29. Sensitive to the needs of others	.115	**.679**	.066	−.097	**.687**	.131	.262	.021
30. Understanding	.110	**.713**	.039	.129	**.593**	.036	.295	.090
31. Compassionate	.127	**.741**	−.057	−.107	**.743**	.080	.242	.069
32. Eager to soothe hurt feelings	−.031	**.541**	−.206	−.018	**.534**	−.024	.007	.053
33. Soft-spoken	−.217	**.309**	.071	.149	**.336**	−.230	.042	.095
34. Warm	.146	**.626**	−.077	.219	**.734**	.245	.018	.017
35. Tender	.075	**.727**	−.057	.243	**.755**	.154	−.001	−.177
36. Gullible	−.140	.198	**−.480**	−.044	.085	−.011	−.243	**−.428**
37. Childlike	−.016	.025	**−.461**	−.120	.033	−.001	−.289	**−.489**
38. Does not use harsh language	−.113	.221	.078	.068	.175	−.064	.128	.104
39. Loves children	−.019	**.434**	.073	.045	**.547**	.068	−.048	.199
40. Gentle	.004	**.671**	−.040	.189	**.732**	.049	.039	−.029
λ	6.93	4.90	1.67	1.07	5.46	4.46	3.69	1.79

Note. Factor loadings ≥.40 are in boldface. *ns* = 400 (females) and 171 (males). BSRI = Bem Sex Role Inventory.
[a] To facilitate reading and interpretation, traits have been grouped by scale rather than as they appear on the BSRI. Thus, Traits 1–20 are considered by Bem to be masculine, and Traits 21–40, feminine.

negatively, the traits Gullible and Childlike also have negative loadings. Males who rate themselves high on Masculine tend to rate themselves low on Feminine as well as Gullible and Childlike, implying that they perceive these traits as related to their conception of Feminine. It will be recalled that Gullible and Childlike had negative loadings on the Self-sufficiency factor in the solution for females.

In sum, the factor analyses of self-ratings

for males and females do not reflect the dimensions of masculinity and femininity proposed by Bem. Not only did four factors emerge in each solution, but also, despite some similarities, they are sufficienty different from each other to lead one to believe that the patterns of self-ratings for males and for females differ. The fact that the traits Masculine and Feminine describe a separate bipolar factor also casts doubt on the validity of the classification of the remaining items as masculine and feminine. It will also be noted that some traits (e.g., Yielding, Soft-spoken) did not have meaningful loadings in either of the solutions.

Because some researchers seem to consider internal consistency reliability as evidence of the unidimensionality of summated rating scales, we present some reliability data on the Femininity scale to demonstrate the dubious nature of this approach. In the present study, the alphas for the Femininity scale (consisting of the 20 "feminine" items) are .79 and .77 for females and males, respectively. Contrast these with the alphas of .87 and .89 for females and males, respectively, on the 11 items with meaningful loadings on the Interpersonal Sensitivity factor (see Table 6, Factor 2 for females and Factor 1 for males; the trait Feminine, which has a meaningful loading on this factor for females only, was excluded from this analysis for comparative purposes).

It would be interesting to compare the factor structures obtained under the conditions of self-ratings with those obtained under the conditions of desirability ratings (Study 1). It should be noted, however, that the two studies used different subjects and that Study 1 included also the "neutral" traits. Nevertheless, on the basis of the differences in the nature of self-ratings and desirability ratings, and on the basis of the overall pattern of the two sets of factor structures, it seems plausible to speculate that the dimensions that underlie trait ratings differ when the task is to rate oneself and when it is to rate the desirability of the traits for an abstract referent. Assuming that similar findings are obtained in future research, one would have to question the validity of arriving at subscales for self-ratings on the basis of analysis of desirability ratings.

Discussion

Bem's effort to construct measures of masculinity and femininity was destined to fail, as it was based solely on an empirical approach in which trait selection was determined by a multitude of nonindependent univariate tests of significance. It is not surprising that some of the traits designated as sex related by Bem are considered neutral, and vice versa, by other researchers, who have resorted to different empirical approaches in their search for sex-related traits. The absence of theoretical definitions of the constructs precludes attempts to determine whether or not a given set of traits is representative of a given domain. How can one assess the validity of a measure when the construct it is supposed to be measuring is undefined? The present discussion is necessarily limited to a consideration of the properties of the BSRI regardless of one's definitions of sex roles.

It will be recalled that Bem's sole criterion for the designation of a trait as sex related was whether or not there was a significant difference in the ratings of its desirability for a man and for a woman. As was demonstrated in Study 1, this approach yielded 20 "masculine" traits which are all positive, or relatively high on desirability, and 20 "feminine" traits, some of which are negative even when the referent is a woman. This, of course, has implications for the total scores on Masculinity and Femininity, in view of people's general tendency to attribute to themselves positive traits, and their reluctance to attribute to themselves negative traits, when they respond to a self-report instrument. This problem is particularly serious because the difference between the scores on Masculinity and Femininity is used to arrive at an Androgyny score.

Before dealing with the issues of androgyny, however, it is important to note that no attempt was made by Bem, or other researchers using the BSRI, to study the dimensions that underlie the ratings of the desirability of the BSRI traits. It was demonstrated in Study 1 that the factor structure of the ratings of trait desirability is not consistent with the

BSRI trait classification. While most of the "masculine" traits have meaningful loadings on a single factor, reflecting a narrow domain of assertiveness, the "feminine" and the "neutral" traits are indistinguishable from each other, and break down into two factors: one consisting of positive traits reflecting interpersonal sensitivity; the other being composed of negative traits reflecting emotional immaturity. Consequently, it is inappropriate to treat the 20 "feminine" traits as a single summated rating scale.

The situation becomes even more complicated when one recalls that the factor structure of self-ratings (Study 2) was more complex in that the "masculine" items also split into two factors. Moreover, the factor structure for females differed from the one for males.

The above noted findings are sufficient to reject Bem's operational definition of androgyny solely on empirical grounds. Since, however, the popularity of the BSRI is probably due primarily to the Androgyny score that it presumably provides, it is of interest to sketch the metamorphosis that the operational definition of androgyny has undergone during its brief history. Except for indicating that being androgynous means being "*both* masculine and feminine, *both* assertive and yielding, *both* instrumental and expressive" (Bem, 1974, p. 555), Bem provides no discussion of androgyny. Instead, it is operationally defined on the basis of a test of significance between the means of Masculinity and Femininity, treating the self-ratings of one individual on 40 traits as being independent of each other. It seems unnecessary to discuss the problems in such an approach, as this has already been done by Strahan (1975). Nor is it necessary to discuss the methodological problems attendant on the use of difference scores, as detailed discussions are available (e.g., Harris, 1963). It will only be pointed out that the use of a difference score as a definition of a construct is highly questionable (see, for example, Cronbach, 1958; Cronbach & Furby, 1970).

Responding to criticisms (e.g., Spence, Helmreich, & Stapp, 1975; Strahan, 1975) that her operational definition of androgyny does not distinguish between different kinds of the same numerical discrepancy between Masculinity and Femininity (e.g., high–high, average–average, low–low), Bem proposed that median splits be used instead. According to the new definition, people who are above the median on both scales are classified as androgynous; those who are above the median on Masculinity and below the median on Femininity are classified as masculine; people who are above the median on Femininity and below the median on Masculinity are classified as feminine; and those below the median on both scales are considered undifferentiated. This is probably the crudest and the least useful method for arriving at a typology (for a review of available methods see Bailey, 1974). When using median splits one runs the risk of classifying some of the people whose scores on both scales are relatively similar as being different types, and of classifying some whose scores are relatively dissimilar as being of the same type. If, in addition, it is noted that all the groups used in studies with the BSRI were groups of convenience and not samples of any defined population, it becomes evident that a person may be classified as being one type or another depending upon the aggregate of people to which he, or she, is considered to belong because the BSRI was administered to them (e.g., introductory psychology classes). Finally, the use of median splits is unwarranted in view of the factorial complexity of the scales.

In a recent discussion of the relative merits of difference scores and median splits, Bem (1977) relies again on statistical tests of significance as the sole criterion. Using an assortment of scales (e.g., Attitudes Toward Women, Internal–External Locus of Control, Machiavellianism, Self-Esteem), Bem tested whether high–high individuals differed significantly from low–low individuals on each. Having found significant differences on some scales, but not on others, she concluded that a distinction between high–high and low–low was warranted, though she acknowledged that she was unable to specify the conditions under which the high–high will differ from the low–low. Is one to deduce that sometimes

high–high and low–low are the same kind of androgynous people, and sometimes they are different (androgynous?) types, though it is not possible to say when?

Bem (1977) concludes her discussion by stating, "Finally, we urge investigators to further analyze their data without categorizing individual subjects in any way, i.e., through the use of multiple regression techniques" (p. 204). While endorsing what appears to be a suggestion to conduct studies within the framework of trait–treatment interactions, one cannot help wondering: Where has androgyny gone?

Reference Notes

1. Spence, J. T. *Traits, roles and the concept of androgyny.* Paper presented at the Conference on Perspectives of Women, Michigan State University, East Lansing, May 13–14, 1977.
2. Bem, S. L. *Beyond androgyny: Some presumptuous prescriptions for a liberated sexual identity.* Keynote address for the American Psychological Association – National Institute of Mental Health Conference on the Research Needs of Women, Madison, Wisconsin, May 1975.

References

Anderson, N. H. Likeableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology,* 1968, *3,* 272–279.

Bailey, K. D. Cluster analysis. In D. R. Heise (Ed.), *Sociological methodology: 1975.* San Francisco: Jossey-Bass, 1974.

Bem, S. L. The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology,* 1974, *42,* 155–162.

Bem, S. L. Sex role adaptability: One consequence of psychological androgyny. *Journal of Personality and Social Psychology,* 1975, *31,* 634–643.

Bem, S. L. On the utility of alternative procedures for assessing psychological androgyny. *Journal of Consulting and Clinical Psychology,* 1977, *45,* 196–205.

Bem, S. L. Theory and measurement of androgyny: A reply to the Pedhazur-Tetenbaum and Locksley–Colten critiques. *Journal of Personality and Social Psychology,* 1979, *37,* 1047–1054.

Bernard, J. *Women, wives, mothers: Values and options.* Chicago: Aldine, 1975.

Bernard, J. Change and stability in sex-role norms and behavior. *Journal of Social Issues,* 1976, *32,* 207–223.

Block, J. H. Conceptions of sex role: Some cross-cultural and longitudinal perspectives. *American Psychologist,* 1973, *28,* 512–526.

Block, J. H., Van der Lippe, A., & Block, J. H. Sex role and socialization patterns: Some personality concomitants and environmental antecedents. *Journal of Consulting and Clinical Psychology,* 1973, *41,* 321–341.

Brannon, R. The male sex role: Our culture's blueprint of manhood, and what it's done for us lately. In D. David & R. Brannon (Eds.), *The forty-nine percent majority: The male sex role.* Reading, Mass.: Addison-Wesley, 1976.

Bryson, J. B., & Corey, D. M. Sex-diagnosticity in 100 personality trait adjectives. *Personality and Social Psychology Bulletin,* 1977, *3,* 301–304.

Bullough, V. L. *The subordinate sex.* Urbana: University of Illinois Press, 1973.

Clifton, A. K., McGrath, D., & Wick, B. Stereotypes of woman: A single category? *Sex Roles,* 1976, *2,* 135–148.

Constantinople, A. Masculinity–femininity: An exception to a famous dictum? *Psychological Bulletin,* 1973, *80,* 389–407.

Cooley, W. W., & Lohnes, P. R. *Multivariate data analysis.* New York: Wiley, 1971.

Cronbach, L. J. Proposals leading to the analytic treatment of social perception scores. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior.* Stanford, Calif.: Stanford University Press, 1958.

Cronbach, L. J., & Furby, L. How should we measure "change"—Or should we? *Psychological Bulletin,* 1970, *74,* 68–80.

Cronbach, L. J., & Meehl, P. E. Construct validity in psychological tests. *Psychological Bulletin,* 1955, *52,* 281–302.

Edwards, A. L. *The measurement of personality traits by scales and inventories.* New York: Holt, Rinehart & Winston, 1970.

Ehrlich, H. J., & Rinehart, J. W. A brief report on the methodology of stereotype research. *Social Forces,* 1965, *43,* 564–575.

Fishman, J. A. An examination of the process and function of social stereotyping. *The Journal of Social Psychology,* 1956, *43,* 27–64.

Gaudreau, P. Factor analysis of the Bem Sex Role Inventory. *Journal of Consulting and Clinical Psychology,* 1977, *45,* 299–302.

Gorsuch, R. L. *Factor analysis.* Philadelphia, Pa.: Saunders, 1974.

Gough, H. G. *California Psychological Inventory: Manual.* Palo Alto, Calif.: Consulting Psychologists Press, 1957.

Green, B. F. Attitude measurement. In G. Lindzey (Ed.), *Handbook of social psychology* (Vol. 1). Reading, Mass.: Addison-Wesley, 1954.

Guilford, J. P. When not to factor analyze. *Psychological Bulletin,* 1952, *49,* 26–37.

Guilford, J. P., & Zimmerman, W. S. *The Guilford-Zimmerman Temperament Survey: Manual of instructions and interpretations.* Beverly Hills, Calif.: Sheridan Supply, 1949.

Harris, C. W. (Ed.) *Problems of measuring change.* Madison: University of Wisconsin Press, 1963.

Hathaway, S. R., & McKinley, J. C. *The Minnesota*

Multiphasic Personality Inventory. New York: Psychological Corp., 1943.

Hochschild, A. R. A review of sex role research. In J. Huber (Ed.), *Changing women in a changing society*. Chicago: University of Chicago Press, 1973.

Hunter, J. E. Images of woman. *Journal of Social Issues*, 1976, *32*, 7–17.

Johnson, M. M. Fathers, mothers and sex typing. *Sociological Inquiry*, 1975, *45*, 15–26.

Lipman-Blumen, J. Role de-differentiation as a system response to crisis: Occupational and political roles of women. *Sociological Inquiry*, 1973, *13*, 105–129.

Maccoby, E. E., & Jacklin, C. N. *The psychology of sex differences*. Stanford, Calif.: Stanford University Press, 1974.

Messick, S. Personality measurement and college performance. In *Educational Testing Service Invitational Conference, 1963*. Princeton, N.J.: Educational Testing Service, 1963.

Oskamp, S. *Attitudes and opinions*. Englewood Cliffs, N.J.: Prentice-Hall, 1977.

Parker, G. V. C. Sex differences in self-description on the adjective check list. *Educational and Psychological Measurement*, 1969, *29*, 99–113.

Parker, G. V. C., & Veldman, D. J. Item factor structure of the adjective check list. *Educational and Psychological Measurement*, 1969, *29*, 605–613.

Parsons, J. E., Ruble, D. N., Hodges, K. L., & Small, A. W. Cognitive-developmental factors in emerging sex differences in achievement-related expectancies. *Journal of Social Issues*, 1976, *32*, 47–61.

Pleck, J. H. Masculinity–femininity: Current and alternate paradigms. *Sex Roles*, 1975, *1*, 161–178.

Pleck, J. H. The male sex role: Definitions, problems, and sources of change. *Journal of Social Issues*, 1976, *32*, 155–164.

Rebecca, M., Hefner, R., & Oleshansky, B. A model of sex-role transcendence. *Journal of Social Issues*, 1976, *32*, 197–206.

Rosenkrantz, P. S., Vogel, S. R., Bee, H., Broverman, I. K., & Broverman, D. M. Sex-role stereotypes and self-concepts in college students. *Journal of Consulting and Clinical Psychology*, 1968, *32*, 287–295.

Rosnow, R. L., Wainer, H., & Arms, R. L. Anderson's personality-trait words rated by men and women as a function of stimulus sex. *Psychological Reports*, 1969, *24*, 787–790.

Ruble, D. N., Frieze, I. H., & Parsons, J. E. (Eds.). *Journal of Social Issues*, 1976, *32*, entire issue.

Sarbin, T. R., & Allen, V. L. Role theory. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (2nd ed., Vol. 1). Reading, Mass.: Addison-Wesley, 1968.

Schneider, D. J. Implicit personality theory: A review. *Psychological Bulletin*, 1973, *79*, 294–309.

Sherman, J. A. *On the psychology of women: A survey of empirical studies*. Springfield, Ill.: Charles C Thomas, 1971.

Sherman, J. A. Social values, femininity, and the development of female competence *Journal of Social Issues*, 1976, *32*, 181–195.

Smith, M. B. The self and cognitive consistency. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook*. Chicago: Rand McNally, 1968.

Spearman, C. "General Intelligence," objectively determined and measured. *American Journal of Psychology*, 1904, *15*, 201–293.

Spence, J. T., Helmreich, R., & Stapp, J. Ratings of self and peers on sex role attributes and their relation to self-esteem and conceptions of masculinity and femininity. *Journal of Personality and Social Psychology*, 1975, *32*, 29–39.

Steinmann, A., & Fox, D. J. Male–female perceptions of the female role in the U.S. *Journal of Psychology*, 1966, *64*, 265–276.

Strahan, F. Remarks on Bem's measurement of psychological androgyny: Alternatives, methods, and a supplementary analysis. *Journal of Consulting and Clinical Psychology*, 1975, *43*, 568–571.

Taylor, G. R. *Sex in history* (2nd ed.). New York: Harper & Row, 1973.

Travers, R. M. W. Rational hypotheses in the construction of tests. *Educational and Psychological Measurement*, 1951, *11*, 128–137.

Vinacke, W. E. Stereotypes as social concepts. *The Journal of Social Psychology*, 1957, *46*, 229–243.

Wiggins, J. S. *Personality and prediction: Principles of personality assessment*. Reading, Mass.: Addison-Wesley, 1973.

Wrightsman, L. S. *Social psychology* (2nd ed.). Monterey, Calif.: Brooks/Cole, 1977.

Wylie, R. C. The present status of self theory. In E. F. Borgatta & W. W. Lambert (Eds.), *Handbook of personality theory and research*. Chicago: Rand McNally, 1968.

Wylie, R. C. *The self concept* (Rev. ed., Vol. 1). Lincoln: University of Nebraska Press, 1974.

Received March 18, 1978 ■

Postscript

We confine this comment on Bem's (1979) reply to the present article to the question of the theory behind the construction of the BSRI. Nowhere in her earlier statements regarding the construction of the BSRI (e.g., Bem, 1974) were we able to find an allusion to the theory of sex roles that is said by Bem (1979) to be the basis for the BSRI. Bem does make reference to statements by other people about possible motivational processes of sex-typed individuals. But nowhere does she present "a theory about both the cognitive processing and the motivational dynamics of sex-typed and androgynous individuals" (Bem, 1979, p. 1048). Quite the contrary, Bem

has stated, "My hypotheses have derived from *no* formal theory" (Bem, Note 2, p. 1, italics added). In our opinion, the rationale presented in Bem's (1979) reply to this article does not constitute even a rudimentary theory. Asserting, as Bem does, that some individuals are motivated to conform to sex-typed cultural norms and that others are not motivated to do so, or that some individuals are "consistent" and others are "inconsistent", is tautological, unless one articulates a theoretical explanation for such phenomena.

Thus, in our opinion, it is Bem's approach to scale construction, *not* the culture, that "has *clustered* a quite heterogeneous collection of attributes into two mutually exclusive categories" (Bem, 1979, p. 1048, italics added). It is Bem, not the culture, that "groups a hodgepodge of attributes into a category *it calls* 'femininity' or 'masculinity' " (Bem, 1979, p.1049, italics added). And it is Bem who treats a hodgepodge of items as if it comprised a unidimentional scale. Bem maintains that her "theory deliberately does not specify the particular contents of these definitions, however, because these will vary from culture to culture" (Bem, 1979, p. 1049) and because hers is a theory of "process" not of "content." How, then, does one reconcile this orientation with the approach used by Bem in the revision of the BSRI (Bem, 1979, p. 1051) in which items are eliminated because their correlations with the total score are low, or because they load on separate factors? Seeking to determine which items are "clustered," Bem appropriately resorts to some of the same methods whose absence in the construction of the original BSRI we have criticized. In sum, we do not see how the retained items in the short BSRI reflect a theory of "process" that is not enunciated.

Journal of Personality and Social Psychology
1979, Vol. 37, No. 6, 1017-1031

Psychological Androgyny: A Case of Mistaken Identity?

Anne Locksley
New York University

Mary Ellen Colten
University of Michigan

This article critically discusses theoretical and methodological issues raised by psychological androgyny research and theory. A number of problematic assumptions shared by this approach and the previous masculinity/femininity approach are detailed. The first part of the article considers whether inventories devised to tap general sex stereotypes should be used as individual personality measures. Alternative forms of cognitive structures, linking sex, other person features, and behavioral rules, are described and hypothesized to have sex-differentiating effects on behavior. The second part of the article discusses problems created by the persistent use of indicators of adaptation and mental health as criterion variables in research on sex identity and sex roles. It is suggested that a psychological theory of sex identity and sex roles should recognize the fact that sex is a structural feature of situations and of ongoing organizations of life experience. In a society in which sex plays a role in the very structuration of experience, the notion of psychological androgyny, with its implication of freedom from sex-related social and biological effects on personality and behavior, is arbitrary.

Reviews of 40 years of psychological research and theory about masculinity/femininity (Constantinople, 1973; Harrison, 1975; Maccoby & Jacklin, 1974; Pleck, 1975, 1976) reveal that Terman and Miles's (1936) original formulation of the masculinity/femininity construct initiated a line of research that consistently failed to provide a coherent, reliable, empirical basis for psychological theories of sex identity. Indeed, the presuppositions of this research were never substantiated. Recently, the concept of psychological androgyny (Bem, 1974, 1975, 1976; Block, 1973; Heilbrun, 1973; Spence & Helmreich, 1978; Spence, Helmreich, & Stapp, 1974, 1975; Bem, Note 1) has emerged in the wake of masculinity/femininity research. It has already been hailed as an exemplar for a new paradigm for psychological research on sex identity (cf. Harrison, 1975; Kaplan & Bean, 1976; Pleck, 1975, 1976; Rebecca, Hefner, & Oleshansky, 1976).

At first glance, the concept of psychological androgyny is very appealing. The concept, and the theory in which it is nested, seems to be saying not only that there is no intrinsic link between one's anatomical sex and one's behavior and interests, but also that people whose personalities fail to conform to sex-role standards are better off for it. However, research and theory about psychological androgyny deserve more than a first glance for their proper evaluation. The politicization of the argument has tended to obscure the fact that it makes specifically psychological assertions about the nature of sex identity, definitions of masculinity and femininity, and appropriate means of measuring these phenomena. The validity of these psychological assumptions and techniques for measurement should be assessed on scientific as well as political grounds. After all, the general area of

The first author is grateful for the facilities permitting the preparation of the final draft of this manuscript that were generously provided by the Laboratory for Research in Social Relations at the University of Minnesota.

The authors would like to express their appreciation for the helpful comments of Judith Bardwick, Eugene Borgida, Bruce Fireman, Susan Harding, Robert Helmreich, David Mann, Mark Snyder, Janet Spence, and two anonymous reviewers on earlier drafts of this article.

Requests for reprints should be sent to Anne Locksley, Department of Psychology, New York University, 6 Washington Place, New York, New York 10003.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3706-1017$00.75

sex roles and sex differences in psychology can only benefit from testing the theoretical and methodological mettle of what has already been acclaimed as an important new contribution to the field.

Accordingly, the present article is intended to contribute to a reflective and critical analysis of psychological androgyny research and theory. It should be noted at the outset that the primary focus of this paper is on the logic of psychological androgyny research, as expressed in theoretical formulations and methodological practices. We will not address psychometric issues raised by existing measures of psychological androgyny, except when they are indicative of the conceptualization of the phenomena in question. Thus recent studies of the instability of Bem's (1974) item-selection procedure (cf. Edwards & Ashworth, 1977; Walkup & Abbott, 1978) or of the Bem Sex Role Inventory (BSRI) factor structure (Gaudreau, 1977; Moreland et al., 1978) will not be discussed. Nor will other purely psychometric considerations, such as the absence of homogeneity across classes of trait terms (cf. Hogan, DeSoto, & Solano, 1977) on the BSRI, (cf. also Strahan, 1975), be explored. The fundamental issues addressed here are (a) the feasibility of using inventories developed to tap general perceptions of aggregate differences as measures of individual differences and (b) the appropriateness of a traditional individual-differences approach to the phenomena of sex roles, sex differences in personality or behavior, and sex identity. It will become apparent that psychological androgyny research and theory, far from constituting a radical departure, share a number of problematic assumptions with masculinity/femininity research and theory. We discuss some of these problems with the hope of encouraging a better approach to the phenomena of sex roles, sex differences in personality, and sex identity.

Definition of Psychological Androgyny

The concept of psychological androgyny has been operationally defined in differing ways, though there is a general consensus about the properties that define an androgynous person. These properties are best understood in relation to the earlier formulation of masculinity/femininity. The various masculinity/femininity measures (Gough, 1952; Guilford & Guilford, 1936; Hathaway & McKinley, 1951; Strong, 1936; Terman & Miles, 1936) had all been constructed with the assumption that masculinity and femininity were opposite poles of a single dimension; thus the less masculine a given respondent's score, the more feminine the respondent was deemed to be, ipso facto. In contrast, the concept of psychological androgyny presupposes that masculinity and femininity are orthogonal personality constructs, making it possible to categorize respondents on the basis of their scores on both masculinity and femininity scales.

Psychological androgyny has been defined simply as a function of masculinity and femininity scores. Currently two scales are widely used in psychological androgyny research: the Bem Sex Role Inventory (BSRI), developed by Bem (1974), and the Personal Attributes Questionnaire (PAQ), developed by Spence, Helmreich, and Stapp (1974) with items culled from the Sex Role Stereotype Questionnaire (Rosenkrantz, Vogel, Bee, Broverman, & Broverman, 1968). Bem originally used the BSRI to categorize respondents into three groups: masculine, feminine, and androgynous. However, the Spence group argued that truly androgynous people are those who endorse both masculine and feminine traits, while those who endorse neither are undifferentiated with respect to masculinity and femininity. Accordingly, Bem (1977) explored the utility of this distinction. Although her reanalyses of data from earlier experimental studies, and her investigation of relationships between sex type and other personality constructs, produced equivocal results, she concluded that the distinction between undifferentiated and androgynous respondents is warranted. So psychological androgyny is the presence of both masculinity and femininity. In turn, masculinity and femininity are defined as additive combinations of trait terms judged to be significantly more desirable for, or more characteristic of, each sex relative to the other.

Bem and Spence and Helmreich are in accord with respect to the essential content of

the masculinity and femininity constructs. Both consider femininity to involve nurturance, expressivity, and empathy, and masculinity to involve agency, instrumentality, and dominance. However, they interpret scores on their own instruments somewhat differently. Bem deliberately invokes cognitive and dispositional processes for interpreting responses to the BSRI. At one point she writes that cross-situational consistency in behavior can be attributed to a "motivational disturbance" that "causes a person to respond to dissimilar situations *as if* they were equivalent" because she or he "is defensively motivated to maintain some sort of image" (Note 1, p. 8). Highly "sex-typed" persons are seen to be defensively motivated in this sense. Thus Bem presents the BSRI as a measure of the "degree of sex-role stereotyping in the person's self-concept" (1976, p. 51).

In contrast, Spence and Helmreich avoid appeals to cognitive or motivational processes and argue that responses to the PAQ can be understood as relatively veridical assessments of individuals' repertoires of traits or personality characteristics (cf. Spence & Helmreich, 1978, pp. 14–16). They are especially committed to distinguishing masculinity and femininity as personality constructs from sex role-related behavior, as well as from dispositions to actualize concepts of masculinity and femininity.

From our perspective, the theoretical and methodological similarities of the two bodies of work outweigh their differences and justify their joint treatment. Most of the criticisms to be presented here are intended to speak both to Bem's and to Spence and Helmreich's work on psychological androgyny. However, as a result of their differing interpretations of their scales, some criticisms will refer only to Bems work,[1] and other criticisms only to Spence and Helmreich's work. In such cases the referent will be explicit.

Problems With the Conceptualization and Measurement of Psychological Androgyny

The development and labeling of any personality typology involve claims of reference and of identity. The investigator in effect asserts that all persons falling within each category of the typology can be considered as equivalent, at least with respect to the quality identified by the category title. The strategy by which the typology or measure is constructed essentially represents the logic of these claims. The labeling of the typology, particularly when appropriating terms current in ordinary discourse, conveys the nature of the quality or construct at issue and thereby evokes commonly associated expectations and assumptions. This section considers the logic governing the development and application of the BSRI and the PAQ, traces the origin of that logic in masculinity/femininity research, and seriously questions its validity.

Whether defining masculinity and femininity as opposite ends of a continuum or as orthogonal dimensional constructs, psychologists have invariably derived their content from aggregate differentiation of the two sexes,

---

[1] Because Bem invokes cognitive and dispositional processes when interpreting responses to the BSRI, the burden of proof, as it were, is heavier for her than for Spence and Helmreich. Bem wants to interpret item endorsement as indicative of the extent to which sex stereotypes have been internalized and consequently control one's behavior, proscribing and prescribing various acts. However, the nature of the BSRI items and the manner of the administration of the BSRI do not warrant such an inference. For example, consider two items on the BSRI Masculinity scale: "Acts as a leader" and "has leadership abilities." In order to know whether or not one has leadership abilities, one has to have been in situations in which such a position was accessible. It hardly needs to be pointed out that tremendous discrimination against women's assumption of leadership positions is a pervasive feature of our society. Endorsement of such an item, or lack thereof, may be a function of experiences and of the cumulative effects of those experiences resulting from a number of causes other than identification with a sex stereotype. In general, responses may reflect behavior induced by systematic situational factors as well as behavior indicative of personal dispositions. The lack of homogeneity (cf. Hogan et al., 1977) among the trait terms also undermines the assumption that exclusive endorsement of masculinity items or exclusive endorsement of femininity items necessarily reflects a global underlying disposition to emulate masculine or feminine stereotypes. Factor analyses (Gaudreau, 1977; Moreland et al., 1978) have produced four factors more internally consistent than the original three BSRI scales, further weakening the assumption of unidimensional constructs conforming to the original labels.

biologically defined, and have created constructs subsequently applied to members of either sex. Masculinity/femininity research was generally based on the assumption that sex type is a function of behaviors or attributes significantly more characteristic of one sex than of the other. Sex-differing attributes were established empirically; sex differences in response constituted the criterion for item selection in developing the masculinity/femininity scales.

Psychological androgyny researchers differ from masculinity/femininity researchers not only in treating masculinity and femininity as orthogonal constructs but also in relying on general perceptions of aggregate sex differences in personality attributes. Items for both the BSRI and the PAQ masculinity and femininity scales were selected because they were judged, on the average, to be differentially desirable or characteristic of the typical man and the typical woman. This concept of masculinity and femininity, as composites of attributes distinguishing males from females in general, has been a persistent feature of sex-stereotype research (e.g., Broverman, Vogel, Broverman, Clarkson, & Rosenkrantz, 1972; McKee & Sherriffs, 1957, 1959; Rosenkrantz et al., 1968; Sherriffs & Jarrett, 1953; Sherriffs & McKee, 1957). Indeed, the BSRI and the PAQ are constructed in exactly the same manner as the sex-stereotype scales. The psychological androgyny researchers have assumed that sex type is a function of covariation between individual traits and traits popularly thought to be differentially desirable or typical of men and of women.

The distinction between actual and stereotypically perceived sex differences is an important one. Obviously the effect on behavior of generally held notions about personality covariates of sex may be quite different from the effect of anatomical sex on behavior. There may be sex differences in behavior caused by social or biological factors that are not perceived to be a function of sex and that are consequently not even partially attributable to the press of differential socialization practices or normative expectations. The two sorts of effects on behavior were confounded in masculinity/femininity research by the use of sex differences in item response as the criterion for item selection. In addition, the restriction of scale content to items perceived to be linked to sex forces consideration of the cognitive mediation of a sense of self by concepts connecting sex and other attributes, and of the effects of these concepts on behavior. The reliance on self-report for the measurement of individual differences further underscores the need for recognizing the role of cognitive aspects of sex typing or sex identity on sex differentiation in behavior.

But like the masculinity/femininity researchers, the psychological androgyny researchers use personality inventories developed around aggregate distinctions between the categories of biological sex to measure individual differences in personality regardless of biological sex. This procedure raises two basic theoretical and methodological questions. First, if the intent is to define and measure individual differences in masculinity and femininity, should the definition and measurement of these constructs be based on stereotypically perceived aggregate sex differences? Second, can an inventory developed to tap beliefs about aggregate sex differences be used as a measure of individual differences? Specifically, can the sense provided by the context in which the measure is developed be sustained in the context of its application?

Limitations of the Stereotypic Conception of Masculinity and Femininity

The rationale for deriving the BSRI and the PAQ from sex stereotypes is based on the assumption that different frequencies of attributing adjectives to the typical woman or the typical man indicate beliefs about covariates of sex. The findings of sex-stereotype research do not necessarily support this assumption. Because sex segregation in family and work roles is a widespread feature of American society, those adjectives that meet the criterion may be linked to prototypically female and male family and work roles rather than to sex per se. The evidence suggests that this might be the case. Clifton, McGrath, and Wick (1976) asked a sample of respondents to check those adjectives, from a list of 153, that describe the typical housewife, bunny,

club woman, career woman, and woman athlete. They reasoned that if a general stereotype of women existed, then similar clusters of adjectives should be used across all roles. Instead, they found a surprisingly low commonality of adjectives used across all roles. Indeed, only one adjective was used to describe each of the five types of woman: active. The adjectives used to describe the typical housewife were very similar to those found on the BSRI and the PAQ femininity scales. However, none of the adjectives used to describe bunny appear on these scales. The other three roles, club woman, career woman, and woman athlete, were all described with a preponderance of adjectives similar to those on the BSRI and the PAQ masculinity scales.

Inspection of the content of the PAQ and the BSRI further corroborates the premise that respondents ideas of characteristics generally distinguishing males and females and their ideas of characteristics specific to highly sex-segregated roles (e.g., housewife/mother and career man) may be confounded. It is not difficult to derive traits of nurturance, empathy, and expressiveness from the American ideal of motherhood, or to derive traits of agency, instrumentality, and dominance from the American ideal of the father providing for his family.

Like all sex-stereotype scales developed to date, the BSRI and the PAQ development procedures left respondents with no choice but to try to convey the significance of nontrait person features with trait terms. The consequences may be misleading. For example, Johnson (1975) has pointed out that the instrumentality/expressiveness dichotomy is not a good indication of the masculinity/femininity dichotomy because traditionally female activities, such as mothering, require instrumental behavior in order to be carried out effectively. Nonetheless, the instrumentality/expressiveness dichotomy recurs consistently throughout the sex-stereotype literature and is the basis of the BSRI and the PAQ. We suspect that this distinction derives from the prototypical middle-class housewife who, as a result of present-day affluence, does not procure resources for the family. When the family unit is less affluent, mothering responsibilities often include actions that obtain resources needed for family survival. When trait terms are the only means by which respondents can distinguish females from males, the terms may be used to signify something other than their ordinarily understood meanings. Although the typical housewife may not work for money or use social networks to procure major survival resources, her actual behavior may require "agency" or "instrumentality," and her instrumentality is distinguishable from that of her husband not by adjectival characterizations of behavior but by the context in which it occurs and the purposes it serves. These observations suggest that the content of general sex stereotypes may be nothing other than reified personality characteristics associated with ideal representatives of adult, sex-segregated social roles.

The probable nonindependence of conceptions of prototypically sex-segregated social roles and conceptions of general covariates with sex has several problematic implications for psychological androgyny research and theory. To begin with, most psychological androgyny research has been conducted on college students and even on high school students. The variance in scores permitting the categorization of androgynous persons may indicate only the lack of press from the definitions and constraints of adult sex-segregated social roles rather than the absence of relevant sex-differentiated traits. In fact, the preoccupation of adolescents and young adults with sexuality and courtship suggests an alternative list of adjectives to that of the BSRI or the PAQ. Furthermore, Clifton et al.'s (1976) failure to obtain general stereotypes when information in addition to that of the stimulus person's sex was provided is corroborated by reflection on ordinary language rules. For example, we make discriminations based on such person features as age, sex, and kinship position when we characterize people with sex-related qualities. We may say, "The Marlboro man is the epitome of masculinity," but we don't say, "He's a masculine father," even though it is syntactically acceptable to do so. Similarly, we may characterize a man as effeminate, but not a woman. These observations suggest that general sex stereotypes may

be too global for interpreting and guiding behavior at the level of individual self-perception or self-direction.

The assumption that traits differentially attributed to the typical woman and the typical man retain their sex-associated character when used for self-description purposes is undermined by a recent series of experiments by Nisbett and Zukier (Note 2). They demonstrate that the impact of stereotypes on social inference and prediction is diluted by the presence of "nondiagnostic" information, that is, information unrelated to the stereotype. Obviously, personal experience provides a wealth of nondiagnostic information with respect to global sex stereotypes. The results of Nisbett and Zukier's studies suggest that global sex stereotypes are rarely activated for self-perception or self-description purposes. In fact it is probable that general sex stereotypes are most likely to have effects under either of the following conditions: (a) in situations comparable to those in which sex stereotypes are assessed, namely, in situations in which only the sex of the target person is known or in which sex is the most informative characteristic known, and (b) in situations in which sex is the salient factor distinguishing two persons and in which its general covariates are representative of the situational requisites. Support for the first condition is provided by studies of preferential hiring decisions. Blau and Jusenius (1976) show that sex discrimination in employment is most extensive at the port of entry for internal job ladders precisely because few other person features (e.g., dependability, companionability) relevant for long-term association are known. Support for the second condition is suggested by the popular recognition of the impact of childbirth and of parenting during early infancy on parents' acceptance of, and behavioral adherence to, sex stereotypes.

Psychological androgyny researchers define masculinity and femininity as stereotypically perceived composites of traits believed to differentiate the sexes on the average. However, if the intent is to use concepts of sex covariates at the individual level of analysis, in an effort to discern either the influence of these concepts on behavior or the presence and ramifications of those covariates, it may be more reasonable and effective to tap those concepts most pertinent for self-description and self-perception. Otherwise, psychological androgyny may reflect only the failure of the particular masculinity and femininity measures used to tap the most salient dimensions of sex differentiation in personality and behavior. Certainly the form and content of ideas of sex covariates have yet to be determined. They may be like disconnected sets of propositions (e.g., propositions about lesbians may be quite different from propositions about glamorous actresses); they may distinguish people on the basis of kinship categories or on the basis of traits; they may be embedded in concepts of settings or environments. Wittgenstein's (1958) notion of a family of concepts with varying criteria of family resemblances may be more appropriate here than the notion of a solitary and general sex stereotype. In any case, the determination of the nature of ordinary concepts linking sex, other attributes, and behavioral rules is important because of its ramifications for the types of predictions one would make about the effects of such concepts or cognitive structures on behavior.

Recent work in cognitive psychology offers hypotheses about the nature and structure of ideas of masculinity and femininity alternative to those presupposed by Bem and by Spence and Helmreich. The BSRI and PAQ development and scoring procedures attribute masculinity to an individual when all traits believed to be significantly more desirable for, or characteristic of, males rather than females are properties of that individual. The attribution of femininity follows a parallel logic. However, the structure of such concepts need not be so Aristotelian. For example, Rosch (1975) demonstrated that the structure of cognitive representations of categories can be of a prototypical sort. Such a category is represented by a particular object, called a good instance of the category, surrounded by other objects of decreasing similarity to the prototype and correspondingly decreasing degrees of membership. Analogously, popular ideas of what it is to be a woman and what it is to be a man may be represented by proto-

typical persona, or conceptual structures that cluster salient information definitive of a type or category of individuals. Unlike trait-dimensional summaries of people, person prototypes may cluster diverse types of information (e.g., physical appearance, occupation, social prestige, as well as behavioral dispositions and characteristics), including information about vivid, immediately perceptible person features. Such features may activate the prototype, causing the individual to generate numerous inferences about the stimulus person's probable behavioral dispositions in the absence of any information about their actual behavior (e.g., Snyder, Tanke, & Berscheid, 1977; Taylor & Crocker, Note 3). This type of reasoning can affect self-perception as well. For example, when the poet Maya Angelou was 15 years old, she decided that she must be a lesbian because, unlike her girlfriends, her breasts remained small, her hips narrow, and her voice deep (Angelou, 1971). For her, a boyish physique was as evocative of the category of lesbianism as homoerotic feelings.

Certainly the investigator retains the right to label operationally defined personological phenomena as she or he chooses. But should the scientific meaning of terms be arbitrary with respect to their ordinary sense? The term *psychological androgyny* clearly implies "of indeterminate sex, or not sex related." However, its scientific operationalization defines persons as androgynous only in relation to global sex stereotypes. Such persons may still be as influenced as sex-typed persons by concepts of sex-linked attributes, or may evidence sex-differentiated personality or behavioral qualities.

Methodological Problems of Using Sex Stereotypes for Individual-Difference Measures

The validity of transferring the sense provided by the context in which the BSRI and the PAQ were developed to the context in which they are administered can be questioned on certain methodological grounds. The conceptual and referential context in which judgments are made about the self is entirely different from the context in which judgments are made about such abstractions as the typical man and the typical woman.

To begin with, self-rating on an attribute entails comparison processes. When female respondents rate themselves, are they comparing themselves to their female peers, their peers of both sexes, to a general notion of the average woman, or to some specific exemplar of femininity? The same question can be asked regarding male respondents. For example, a woman may endorse "athletic" as a true description of herself because she is athletic relative to other women, although she may not be particularly athletic relative to most men. Similarly a man may endorse "compassionate" because he considers himself to be more compassionate than most men, though less compassionate than most women. Furthermore, the behaviors referred to by these terms may not be the same for men as for women. The compassionate man may be thinking of his indignation over inequities and injustices in the world, whereas the compassionate woman may be thinking of her empathy for the problems of her friends. Endorsement of the same items may thus conceal sex differences, again raising the question of the validity of the behavioral reflections of a label such as *androgyny*.

In addition, respondents asked to describe themselves use a different inferential context, which may alter the sense of a given term from that afforded by its original context. For example, when thinking of the ideal housewife/mother, loyalty (an item on the BSRI femininity scale) would seem to be a salient feature of the mother's feelings about her family. On the other hand, a football jock may think of loyalty in the sense of loyalty to his buddies, or to his team and to his school. This concept of loyalty is hardly indicative of femininity or of emulation of a female sex stereotype.

It could be argued that all personality inventories feature indeterminant inferential contexts, and so singling out the BSRI and the PAQ is unfair. But the problem is not so much one of general ambiguity, but rather one of biases in self-report, or automatic adjustments to one's sex, systematically related to the notion of what is measured. The sorts

of referential vagaries noted above contribute to a surface impression of equivalence not necessarily supported by the underlying evidence. Indeed, sex differences in the results of Bem's (1976) experimental studies lend credence to this premise, that responses differ in their significance by sex.

Fundamentally, the appropriateness of a traditional individual-differences approach to this general domain is questionable. There are special characteristics of masculinity, femininity, sexual identity, and sex differentiation, such as it is, in personality or behavior, that do not hold for any other individual-difference construct (with the possible exception of intelligence). For one thing, sex, the anchor for concepts of masculinity and femininity, is a universal attribute of the organism. The lay person and, until recently, the psychologist tend to assume the bipolar nature of masculinity and femininity because biological sex is thought to be causally implicated in trait differentiation. Masculinity and femininity are conceptually related to biological sex. They are also, at least in their ordinary sense, conceptually related to sexuality and sexual relations, universal phenomena of extraordinary salience and consequence. Finally, their relation to biological sex suggests another consideration. Sex is a major criterion for the structural differentiation of experience over the life cycle, given an extensively sex-segregated society. The implications of these considerations for the viability of psychological androgyny research and theory are explored in the following section.

Masculinity, Femininity, and Adaptation

Masculinity/femininity research and theory had always been characterized by an emphasis on the contribution of the construct to overall adaptation or adjustment. As a personality construct, masculinity/femininity was postulated to account for individual variation in indicators of adaptation or adjustment. The construct was assumed to be valid if it discriminated criterion groups that could be distinguished on the basis of a "relevant" adjustment indicator. Homosexuality in men, adjudged to be pathological and maladaptive, was one diagnostic criterion advocated by Terman and Miles (1936), and a notable amount of research was conducted investigating the predictive validity of various masculinity/femininity scales regarding male sexual object choice (e.g., Burton, 1947; Krippner, 1964; Manosevitz, 1970). In a similar vein, research was conducted to determine the contribution of masculinity/femininity to global adjustment (e.g., Mussen, 1961), schizophrenia (e.g., Reed, 1956), anxiety (e.g., Consentino & Heilbrun, 1964; Gotts & Phillips, 1968; Gray, 1957), obesity (e.g., Lefley, 1971), and alcoholism (e.g., Jensen & Hoffman, 1973), among other adjustment indicators. The persistent assumption was that masculinity in males and femininity in females is functional for adaptation. The assumption was never substantiated, though the limitations of the masculinity/femininity scales make it difficult to know whether these findings are attributable to construct validity problems or if they constitute valid disconfirmations of the hypothesis.

Psychological androgyny research has inherited the focus on adaptation from the masculinity/femininity research. However, the direction of the argument has been reversed: Androgyny (rather than exclusive masculinity or exclusive femininity) is postulated to facilitate adaptation or adjustment. Bem (1974, 1976, Note 1) has always nested androgyny in a broader notion of the basis of mental health, and her experimental studies are designed to demonstrate that androgynous subjects are more able to cope with the requisites of situations than are masculine or feminine subjects. Indeed, she describes her sex-typed subjects as having "behavioral deficits," arguing that they don't have the abilities necessary for effective action in certain types of situations. Similarly, Spence et al. (1975) demonstrated that psychologically androgynous subjects have higher self-esteem than do masculine, feminine, or undifferentiated subjects. They explain this finding as indicative of androgynous subjects' greater personal and social effectiveness.

Two major problems, however, stem from the focus of this research on the prediction of indicators of adaptation or adjustment. The first involves a persistent blurring between no-

tions of predictive validity and of construct validity. The second and more serious problem involves contradictions between psychological and sociological levels of analysis.

Predictive validity is equivalent to the "truth value", or "power" of a theory, which increases as the hypotheses deduced from the theory are confirmed (Wiggins, 1973); it is not synonymous with construct validity. Construct validity is a feature of a particular measure of a personality construct and is established by a complex set of procedures determining the plausibility that the measure measures what it purports to measure (Campbell & Fiske, 1959; Cronbach & Meehl, 1955; Nunnally, 1967; Wiggins, 1973). The construct validity of a measure cannot be established by demonstrating its predictive validity. Yet the psychological androgyny researchers have attempted to do precisely that. They have reasoned that if androgynous subjects are more flexible in handling diverse situations or evidence more positive self-regard compared to the other categories of subjects, this must be due to differences in the sex type between the four categories of subjects. The notion of the function of sex type for adaptability or adjustment is so entrenched that the construct validity of the BSRI and the PAQ is taken for granted. Oddly, there are numerous behaviors that are popularly recognized to covary with masculinity and femininity, but neither the masculinity/femininity researchers nor the psychological androgyny researchers have tried to determine their relation to the sex-type measures. Consider, for example, masculine males. They are generally assumed to "sow their wild oats," to put a high priority on physical attractiveness in the choice of a mate, to value male friendships, and to socialize frequently in male groups. In contrast, feminine females are generally assumed to invest considerable amounts of time in their physical appearance, to prefer the company of males over females, and to subordinate commitments to female friends to commitments to male friends. Feminine males are assumed to be given to stereotypically feminine gestures and to enjoy sitting about and gossiping. These examples are obviously suggested by popularly disseminated caricatures. Readers may think of still other behaviors that are widely believed to be indicative of masculinity and femininity. What is important for establishing the construct validity of the BSRI and the PAQ is their discrimination of behaviors that reflect dispositional sex-differentiated preferences on the part of the actor, rather than behaviors that reflect such relatively non-sex-related attributes as self-esteem, self-confidence, intelligence, or conservatism.

The confusion between construct validity and predictive validity characteristic of psychological androgyny research and theory can easily be rectified by further research. A more serious problem of contradictions between psychological and sociological levels of analysis is created by the focus on adaptation or adjustment as criterion variables for sex-identity research. Given the overall argument of psychological research and theory, the use of criterion indicators of adaptability or adjustment ignores the fact that many situations continue to be structured with respect to sex, such that sex-appropriate behaviors generally procure psychological and material rewards, whereas sex-inappropriate behaviors may be costly. If situational appropriateness is to be a criterion of adjustment, it would be "maladjusted" to behave in a sex-inappropriate manner. And, given that behavior is at least partially a function of situations (e.g., Mischel, 1968), men and women will act differently because they have different ongoing tasks, responsibilities, and access to resources. The existence of sex segregation and discrimination in the paid labor force has been amply documented, revealing differential sex-dependent contingencies surrounding entry into occupations, considerable sex segregation in occupational distribution, and striking sex differences in earnings (Blau & Jusenius, 1976; Kanter, 1977; Tangri, 1972). Also, in spite of the increasing participation of married women in the work force, time-budget studies show that sex-based division of labor in the family has persisted (Robinson, 1978; Walker & Woods, 1976). Within the coeducational college environment, which is the setting for practically all of the psychological androgyny research, and which is, perhaps, the least sex-segregated

setting of any institutionalized environment, one still discerns ample evidence for differential constraints: in athletic programs, in student political organizations, and in extracurricular activities (Douvan, 1975).

At a level that is more immediately experienced by the average college student, norms governing appropriate behavior in interpersonal relationships in informal as well as formal settings still differ depending on the actor's sex. People are relatively aware of these norms. Their impact on behavior is suggested by the rather distressing findings of a study (Zanna & Pack, 1975) that found that when a potential male partner was described with desirable characteristics, female subjects' self-descriptions were influenced by information about his ideal woman: They tended to describe themselves in conventional terms when the partner's ideal was conventional and in unconventional terms when the partner's ideal was unconventional. When the partner was undesirable, information about his ideal had little impact on the subjects' self-descriptions. The women students clearly geared their behavior to reward contingencies present in the situation.

Snyder and Skrypnek provide further evidence (described in Snyder, Note 4) of the behavioral impact of sex-contingent expectations. They located two individuals in separate rooms and required the pair to divide a series of tasks varying in their sex specificity. Since the dyads could communicate only by means of a signaling system, neither member could detect the sex of the other. In one condition, the "perceiver" was told that the "target" was female; in a second condition, the perceiver was told that the target was male. Snyder and Skrypnek found not only that the perceived sex of the target influenced the outcome of the division of labor process, but also that the targets themselves began to initiate choices consistent with their assigned sex.

Once we consider that situations involve different contingencies and behavioral norms for males and for females, the use of adaptability or adjustment in androgyny research becomes contradictory. Bem's experimental studies demonstrate that androgynous subjects are capable of sex-inappropriate behavior when the demand characteristics of the experiment clearly tell the subjects that sex-inappropriate behavior is situationally appropriate. The studies do not demonstrate that androgynous subjects engage in sex-inappropriate behavior when sex-appropriate behavior is situationally appropriate or rewarding. And if androgynous subjects are to be distinguished by their ability to adapt to situational requisites, will they not act in sex-role-congruent ways when such acts enable them to obtain psychological, interpersonal, and material rewards? The fact that androgynous subjects endorse both masculine and feminine traits in no way militates against this point. Consider the female college student who is analytical, assertive, competitive, independent, individualistic, and self-reliant in her studies, as all good students are expected to be, and affectionate, childlike, flatterable, loyal, soft-spoken, sympathetic, understanding, warm, and yielding with her boyfriend, as all good girlfriends are expected to be. She would be classified as androgynous by Bem, yet her behavior is not independent of sex role norms.

Even though many of the traits may be statistically orthogonal and although individuals, for example, may be both dominant and submissive between situations, within situations their behavior may still be constrained by interaction rules that affect the sexes differently. Bem argues that androgynous individuals are better off than masculine or feminine individuals because they are capable of being, to take one example, yielding in situations in which it is appropriate to yield and, to take another example, assertive in situations in which it is appropriate to be assertive. This argument neglects the fact that norms and rules governing behavior in settings are often sex specific and are contingent on the sex composition of the actors in the setting (cf. Ruble & Higgins, 1976). For example, in a given social interaction it may be appropriate for a woman to yield to a man, but highly inappropriate for a man to yield to a woman. In other settings, the woman may be assertive, but such behavior may not procure or maintain authority as the man's behavior does in interaction with the woman.

Because many situations are structured with

respect to sex in such a way that sex-inappropriate behavior is costly, we should expect to find sex differences in the frequencies of certain types of behavior, which, in turn, have implications for how one might realistically describe oneself as an actor in general. Bem's (1974) data does provide some evidence for this line of reasoning, though she has never considered the implications of the following findings. She found significant sex differences in androgyny scores, with females scoring on the feminine side of the scale and males scoring on the masculine side. In addition, only 6% of the males at Stanford and 9% at Foothill were feminine, and only 8% of the females at Stanford and 8% at Foothill were masculine. These findings suggest that there are definite constraints either influencing respondents' willingness to endorse cross-sex-associated items exclusively or influencing respondents' frequency of acting in ways that could be described with cross-sex-associated items. Either explanation would be consistent with the notions that situations are structured with respect to sex at both institutionalized and informal social levels, and that differential contingencies influence the behavior of males and females, their dispositions to behave in varying ways, and consequently their perceptions of self as an actor in general.

Spence and Helmreich (1979) argue that the influence of sex-contingent norms of behavior "cannot be used as a criticism of personality instruments such as the PAQ; it merely serves as a warning that one should not make predictions about behavior as a function of PAQ scores without taking situational variables into account" (pp. 1037–1038). Elsewhere they stipulate that "expression of the personality traits tapped by the PAQ is likely to be exaggerated when these response dispositions are congruent with situational demands . . . and likely to be suppressed or modified when they are incongruent" (p. 1037). For support of this point they cite a study by Megargee (1969), who fully crossed sex and high- and low-dominant subjects in dyads and asked each pair to choose a leader for working on one of two laboratory tasks. In by far the majority of mixed-sex pairs, the male wound up as the leader, regardless of either partner's dominance rating. Dominant women were more likely to participate verbally in the choice of a leader but were not more likely to be a leader than nondominant women were, whether paired with a high- or a low-dominant male. Spence and Helmreich neglect to point out that this effect occurred regardless of whether the task was of a "masculine" sort, involving industrial mechanical work, or of a "feminine" sort, involving clerical work and verbal talent.

Clearly situational demands and press can inhibit the expression of a personality trait. However, the investigator ought to be able to specify, in a theoretically plausible fashion, those situations in which the personality trait should predict behavior and those situations in which the trait should not predict behavior. Congruence with situational demands is not a valid criterion, at least unless the contribution of situational demands can be distinguished from the contribution of the personality disposition. After all, there is no a priori reason to assume that situational demands always overshadow the influence of personality dispositions. The Megargee study is a case in point. It is not clear to us that one would have predicted no effect of the sex-typing of the task on the expression of the personality trait *dominance* on the part of the women subjects. Furthermore, if situational contingencies and social expectations repeatedly overwhelm the influence of assumed dispositions on behavior, one can hardly expect the targets of these powerful forces to sustain dispositions or self-perceptions largely independent of their behavior.[2]

---

[2] Interestingly, most of Bem's experimental studies involve behaviors directly tapping the behavioral content of the items on the BSRI masculinity and femininity scales. Spence (Note 5) cautions against assuming that either the PAQ or the BSRI masculinity and femininity scales relate to other sex-related aspects of behavior or personality. No one has found any association, for example, between psychological androgyny and attitudes towards women's rights (Jones et al., 1978; Spence, Note 5). Spence and Helmreich observe that in general "the psychological dimensions of masculinity and femininity . . . are only weakly related within each sex to the broad spectrum of sex-role behaviors" (1978, p. 3). They argue that this state of affairs is legitimate and even

Underlying the notion of psychological androgyny is the assumption that sex typing per se is dysfunctional or maladaptive, and that men are as disadvantaged as women, since they are expected to conform or aspire to a stereotypic ideal of masculinity, just as women are expected to conform to a stereotypic ideal of femininity.

Neither data produced by Bem and by Spence and Helmreich nor observations of the relative well-being of men and women in our society support the conclusion that individuals evidencing exclusively masculine traits are in fact as disadvantaged as those evidencing exclusively feminine characteristics. Bem's findings show that feminine females are the least capable of the lot and that little difference obtains between masculine and androgynous females. Spence and Helmreich's subjects scoring high-feminine/low-masculine display considerably lower self-esteem than those scoring high-masculine/low-feminine.

Most of the masculinity traits included in the BSRI and the PAQ reflect skills that enable individuals to access resources and control situations and social interactions. Although masculine individuals may forfeit some emotional gratifications, they have survival skills. On the other hand, most of the femininity traits do not reflect skills: Compare child-like, shy, guillible, soft-spoken, to analytical, assertive, independent, self-reliant, items on the BSRI femininity and masculinity scales, respectively.

Jones, Chernovetz, and Hansson (1978) conducted a series of studies designed to investigate Bem's (Note 1) hypothesis that psychologically androgynous individuals are more adjusted than sex-typed persons, in consequence of their greater behavioral flexibility or more extended behavioral repertoire. They used a variety of dependent measures, including self-esteem, locus of control, problems with alcohol, helplessness, sexual maturity, and neurosis, as adjustment indices, as well as a series of attitudinal variables. They found that masculinity rather than androgyny generally predicted greater adjustment and flexibility, for both sexes. They also found that both male and female feminine and androgynous subjects expressed a desire to become more masculine. These results do not vindicate neglect of the fact that behavior occurs in situations all too often governed by asymmetrically sex-differentiated norms and contingencies. More important, these observations suggest that the term *psychological androgyny* may be utterly arbitrary, inasmuch as persons classified as androgynous are no less aware or concerned about the social implications of their sex and its ramifications for their behavior than are sex-typed persons.

Indeed, how could they be? It is impossible to escape the social consequences of one's sex. Sex is an immediately perceptible feature of every person. Regardless of one's own inclinations, one's sex evokes and elicits sex stereotypes in others. Thus one is repeatedly assaulted by the influence of sex-contingent expectations and norms. Furthermore, as the studies by Megargee (1969), Ruble and Higgins (1976), Snyder and Skrypnek (Snyder, Note 4), and Zanna and Pack (1975) suggest, individuals are easily swayed by sex-contingent expectations, quickly adopting these norms as their own. Both the influence of one's sex on others' perceptions and expectations on oneself and one's actions are unavoidable. The negotiation of social experience involves participating in systems of behavioral rules that mark the sex of each and

---

to be expected, given that the strength of some sex role norms can create homogeneity in behavior regardless of personal inclination, (1978, cf. p. 28). This caveat creates certain difficulties. On the face of it, at least, it is unsatisfactory to show that the BSRI and the PAQ scales simply tap the traits they contain. If they are only trait lists, why are they labeled Masculinity and Femininity scales? It is insufficient to derive their sense purely from aggregate distinctions. That would be equivalent to developing and labeling a scale "Whiteness" because it comprises intelligence items that discriminate the races on the average, and using the scale to categorize individuals as white, nonwhite, and beyond coloration, regardless of their race. The labels and the research that has been conducted with these scales clearly suggest the assumptions that there is something about individual personalities that is being measured, and that that something is sex related. Furthermore, there are situations so weakly governed by sex-role norms that personality dispositions ought to predict whether the person's behavior is sex role conforming or deviant. These situations ought to be specifiable in advance.

every actor. Considering the role of sex in the very architecture of experience and behavior, the notion of psychological androgyny, with its implication of freedom from sex-related social effects on personality and behavior, is arbitrary at best.

The present article has raised a number of questions about the explicit and implicit claims of psychological androgyny research and theory. We have addressed issues concerning the interpretation of respondents' scores on the BSRI and the PAQ, assumptions about the definitions of masculinity and femininity, and the general focus on adjustment and mental health as criterion variables. Throughout our questioning of psychological androgyny research and theory, we have repeatedly emphasized that sex is a structural feature of situations and of ongoing organizations of life experience, in that settings and ongoing organizations are characterized by sex-differentiated behavioral norms, cost/reward contingencies, and access to resources. In general a psychological theory of sex identity or sex differences should recognize this fact, because individuals apprehend the significance of their sex at least partly as a function of their experience with sex-differentiating social and institutional environments.

The diversity and complexity of the social and sociological functions of sex do not automatically exile the factor from the province of psychology, though. Psychological androgyny research introduced the role of cognitive and motivational processes in sex differentiation of behavior, if only by deriving masculinity and femininity scales from general sex stereotypes. A potentially more fruitful approach to the measurement of cognitive structures linking sex, other person features, and behavior rules would begin with concepts of more immediate relevance for self-perception and self-direction than global sex stereotypes and would take advantage of recent developments in cognitive psychology regarding formal and structural features of cognitive organizations (e.g., Abelson, 1976; Cantor & Mischel, 1977; Minsky, 1975; Mischel, 1977; Neisser, 1976; Rosch, 1975). One possible conceptual form suggested earlier is the person prototype. Psychological androgyny researchers define masculinity and femininity in personality-trait terms, as clusters of attributes fixed across time and across situations. Possibly a more adequate representation of person factors that tend to sex-differentiate behavior would focus on the cognitive mediation of a sense of self by concepts connecting sex, other attributes, and their general valence and would examine the effects of those concepts on behavior. As noted earlier, Wittgenstein's (1958) notion of a family of concepts, with varying overlap or criteria of resemblance, is probably preferable to the notion of a general sex stereotype, for purposes of investigating individually initiated, sex-differentiated behavior. In any case, it is important to recognize that existing measures of sex stereotypes carry implicit assumptions about the nature of cognitive representations of world knowledge involving gender. Those assumptions require explicit examination.

Finally, the profound and extensive distinction by sex still characteristic of our society is precisely what has been brought to our attention by the feminist movement. Individuals are willy-nilly caught in the process of contending with social as well as biological implications of their sex that have serious ramifications for their self-regard and for major and minor life decisions. From this perspective, there is no such thing as psychological androgyny. Individuals living within a culture in which sex is deeply structurally embedded are inevitably subject to a complex nexus of social relations that locate them in social space partly as a function of their sex. Rather than constructing personality typologies in which some types of people are accorded privileges over others, even so bland a privilege as the mental health of conformity to situational requisites, it would be far more beneficial for psychologists to investigate the means by which people in general participate in social processes that persistently rely on sex for differential allocation of resources, bases of power, social authority, and prestige.

Reference Notes

1. Bem, S. L. *Psychology looks at sex roles: Where have all the androgynous people gone?* Paper presented at the UCLA Symposium on Women, Los Angeles, May 1972.

2. Nisbett, R., & Zukier, H. *The dilution effect: Producing "regressive" predictions by exposure to nondiagnostic information.* Unpublished manuscript, University of Michigan, 1978.
3. Taylor, S. E., & Crocker, J. *Schematic bases of social information processing.* Paper presented at the First Ontario Symposium on Personality and Social Psychology, London, Canada, August 1978.
4. Snyder, M. *On the self-perpetuating nature of stereotypes.* Unpublished manuscript, University of Minnesota, 1978.
5. Spence, J. *Traits, roles, and the concept of androgyny.* Paper presented at the Conference on Perspectives on the Psychology of Women, Michigan State University, East Lansing, May 13–14, 1977.

References

Abelson, R. P. Script processing in attitude formation and decision-making. In J. S. Carroll & J. W. Payne (Eds.), *Cognition and social behavior.* Hillsdale, N.J.: Erlbaum, 1976.

Angelou, M. *I know why the caged bird sings?* New York: Bantam Books, 1971.

Bem, S. L. The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology,* 1974, *42,* 155–162.

Bem, S. L. Sex role adaptability: One consequence of psychological androgyny. *Journal of Personality and Social Psychology,* 1975, *31,* 634–643.

Bem, S. L. Probing the promise of androgyny. In A. G. Kaplan & J. P. Bean (Eds.), *Beyond sex-role stereotypes.* Boston: Little, Brown, 1976.

Bem. S. L. On the utility of alternative procedures for assessing psychological androgyny. *Journal of Consulting and Clinical Psychology,* 1977, *45,* 196–205.

Blau, F., & Jusenius, C. Economists' approaches to sex segregation in the labor market: An appraisal. *Signs,* 1976, *1,* 181–199.

Block, J. H. Conceptions of sex-role: Some cross-cultural and longitudinal perspectives. *American Psychologist,* 1973, *28,* 512–526.

Broverman, I., Vogel, S., Broverman, D., Clarkson, F., & Rosenkrantz, P. Sex-role stereotypes: A current appraisal. *Journal of Social Issues,* 1972, *28,* 59–78.

Burton, A. The use of the Masculinity–Femininity scale of the MMPI as an aid in the diagnosis of sexual inversion. *Journal of Psychology,* 1947, *24,* 161–164.

Campbell, D. T., & Fiske, D. W. Convergent and discriminant validation by the multitrait–multimethod matrix. *Psychological Bulletin,* 1959, *56,* 81–105.

Cantor, N., & Mischel, W. Traits as prototypes: Effects on recognition memory. *Journal of Personality and Social Psychology,* 1977, *35,* 38–48.

Clifton, A. K., McGrath, D., & Wick, B. Stereotypes of women: A single category? *Sex Roles,* 1976, *2,* 135–148.

Consentino, F., & Heilbrun, A., Jr. Anxiety correlates of sex-role identity in college students. *Psychology Reports,* 1964, *14,* 729–730.

Constantinople, A. Masculinity–femininity: An exception to a famous dictum? *Psychological Bulletin,* 1973, *80,* 389–407.

Cronbach, L. J., & Meehl, P. E. Construct validity in psychological tests. *Psychological Bulletin,* 1955, *52,* 281–302.

Douvan, E. Feminine socialization and higher education. *New Directions for Higher Education,* No. 9, Spring, 1975, 3–47.

Edwards, A., & Ashworth, C. A replication study of item selection for the Bem Sex Role Inventory. *Applied Psychology Measurement,* 1977, *1,* 501–507.

Gaudreau, P. Factor analysis of the Bem Sex Role Inventory. *Journal of Consulting and Clinical Psychology,* 1977, *45,* 299–302.

Gotts, E., & Phillips, B. The relation between psychometric measures of manifest anxiety and masculinity–femininity. *Journal of School Psychology,* 1968, *6,* 123–130.

Gough, H. Identifying psychological femininity. *Educational and Psychological Measurements,* 1952, *12,* 427–439.

Gray, S. Masculinity–femininity in relation to anxiety and social acceptance. *Child Development,* 1957, *28,* 203–214.

Guilford, J., & Guilford, R. Personality factors S, E, and M and their measurement. *Journal of Psychology,* 1936, *2,* 109–127.

Harrison, J. *A critical evaluation of research on "masculinity/femininity" and a proposal for an alternate paradigm for research on psychological differences and similarities between sexes.* Unpublished doctoral dissertation, New York University, 1975.

Hathaway, S., & McKinley, J. *The Minnesota Multiphasic Personality Inventory.* New York: Psychological Corp., 1951.

Heilbrun, C. G. *Toward a recognition of androgyny.* New York: Knopf, 1973.

Hogan, R., DeSoto, C. B., & Solano, C. Traits, tests, and personality research. *American Psychologist,* 1977, *32,* 255–264.

Jensen, D., & Hoffman, H. Demographic and MMPI characteristics of male and female state hospital alcoholic patients. *Psychological Reports,* 1973, *33,* 561–562.

Johnson, M. Fathers, mothers, and sex typing. *Sociological Inquiry,* 1975, *45,* 15–26.

Jones, W., Chernovetz, M. E., and Hansson, R. O. The enigma of androgyny: Differential implications for males and females? *Journal of Consulting and Clinical Psychology,* 1978, *46,* 298–313.

Kanter, R. M. *Work and family in the United States: A critical review.* New York: Russell Sage Foundation, 1977.

Kaplan, A. G., & Bean, J. P. *Beyond sex-role stereotypes.* Boston: Little, Brown, 1976.

Krippner, S. The identification of male homosexuality with the MMPI. *Journal of Clinical Psychology,* 1964, *20,* 159–161.

Lefley, H. Masculinity–femininity in obese women.

Journal of Consulting and Clinical Psychology, 1971, *37,* 180–186.

Maccoby, E. E., & Jacklin, C. *The psychology of sex differences.* Stanford, Calif.: Stanford University Press, 1974.

Manosevitz, M. Item analysis of the MMPI *Mf* scale using homosexual and heterosexual males. *Journal of Consulting and Clinical Psychology,* 1970, *35,* 395–399.

McKee, J. P., & Sherriffs, A. C. The differential evaluation of males and females. *Journal of Personality,* 1957, *25,* 356–371.

McKee, J. P., & Sherriffs, A. C. Men's and women's beliefs, ideals, and self-concepts. *American Journal of Sociology,* 1959, *64,* 356–363.

Megargee, E. I. Influence of sex roles on the manifestation of leadership. *Journal of Applied Psychology,* 1969, *53,* 377–382.

Minsky, M. A framework for representing knowledge. In P. Winston (Ed.), *The psychology of computer vision.* New York: McGraw-Hill, 1975.

Mischel, W. *Personality and assessment.* New York: Wiley, 1968.

Mischel, W. On the future of personality measurement. *American Psychologist,* 1977, *32,* 246–254.

Moreland, J. R., Gulanick, N., Montague, E. K., & Harren, V. A. Some psychometric properties of the Bem Sex Role Inventory. *Applied Psychological Measurement,* 1978, *2,* 249–256.

Mussen, P. Some antecedents and consequences of masculine sex-typing in adolescent boys. *Psychological Monographs,* 1961, *75*(Whole No. 516).

Neisser, U. *Cognition and reality.* San Francisco: Freeman, 1976.

Nunnally, J. C. *Psychometric theory.* New York: McGraw-Hill, 1967.

Pleck, J. H. Masculinity–femininity: Current and alternative paradigms. *Sex Roles,* 1975, *1,* 161–177.

Pleck, J. H. The psychology of sex roles: Traditional and new views. In L. A. Cater & A. F. Scott (Eds.), *Women and men: Changing roles, relationships, and perceptions.* New York: Aspen Institute for Humanistic Studies, 1976.

Rebecca, M., Hefner, R., & Oleshansky, B. A model of sex-role transcendence. *Journal of Social Issues,* 1976, *32,* 197–206.

Reed, M. The M–F dimension in normal and psychotic subjects. *Journal of Abnormal and Social Psychology,* 1956, *55,* 289–294.

Robinson, J. *How Americans use time: A social-psychological analysis.* New York: Praeger, 1978.

Rosch, E. Cognitive representations of semantic categories. *Journal of Experimental Psychology: General,* 1975, *104,* 192–233.

Rosenkrantz, P., Vogel, S., Bee, H., Broverman, I., & Broverman, D. Sex-role stereotypes and self-concepts in college students. *Journal of Consulting and Clinical Psychology,* 1968, *32,* 287–295.

Ruble, D. N., & Higgins, E. T. Effects of group sex composition on self-presentation and sex-typing. *Journal of Social Issues,* 1976, *32,* 125–132.

Sherriffs, A. C., & Jarrett, R. F. Sex differences in attitudes about sex differences. *The Journal of Psychology,* 1953, *35,* 161–168.

Sherriffs, A. C., & McKee, J. P. Qualitative aspects of beliefs about men and women. *Journal of Personality,* 1957, *25,* 451–464.

Snyder, M. N., Tanke, E. D., & Berscheid, E. Social perception and interpersonal behavior: On the self-fulfilling nature of social stereotypes. *Journal of Personality and Social Psychology,* 1977, *35,* 656–666.

Spence, J., & Helmreich, R. *Masculinity & femininity.* Austin: University of Texas Press, 1978.

Spence, J., & Helmreich, R. The many faces of androgyny: A reply to Locksley and Colten. *Journal of Personality and Social Psychology,* 1979, *37,* 1032–1046.

Spence, J., Helmreich, R., & Stapp, J. The Personal Attributes Questionnaire: A measure of sex role stereotypes and masculinity–femininity. *JSAS Catalog of Selected Documents in Psychology,* 1974, *4,* 43. (Ms. No. 617)

Spence, J., Helmreich, R., & Stapp, J. Ratings of self and peers on sex role attributes and their relation to self-esteem and conceptions of masculinity and femininity. *Journal of Personality and Social Psychology,* 1975, *32,* 29–39.

Strahan, R. F. Remarks on Bems measurement of psychological androgyny: Alternative methods and a supplementary analysis. *Journal of Consulting and Clinical Psychology,* 1975, *43,* 568–571.

Strong, E. Interests of men and women. *Journal of Social Psychology,* 1936, *7,* 49–67.

Tangri, S. S. Occupational role innovation among women. *Journal of Social Issues,* 1972, *28,* 177–199.

Terman, C. M., & Miles, C. C. *Sex and personality.* New York: McGraw-Hill, 1936.

Walker, K., & Woods, M. *Time use: A measure of household production of goods and services.* Washington, D.C.: American Home Economics Association, 1976.

Walkup, H., & Abbott, R. Cross-validation of item selection on the Bem Sex Role Inventory. *Applied Psychological Measurement,* 1978, *2,* 63–71.

Wiggins, J. S. *Personality and prediction.* Reading, Mass.: Addison-Wesley, 1973.

Wittgenstein, L. *The blue and brown books.* New York: Harper & Row, 1958.

Zanna, M. P., & Pack, S. J. On the self-fulfilling nature of apparent sex differences in behavior. *Journal of Experimental Social Psychology,* 1975, *11,* 583–591.

Received April 14, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 6, 1032–1046

The Many Faces of Androgyny: A Reply to Locksley and Colten

Janet T. Spence and Robert L. Helmreich
University of Texas at Austin

In their methodological and theoretical critique of the Bem Sex Role Inventory (BSRI) and the Personal Attributes Questionnaire (PAQ), Locksley and Colten assume that there is a singular androgyny theory to which the rationale and the psychometric properties of these instruments are inexorably tied and that each is intended to be a broad-gauged measure of masculinity and femininity or of global self-images of these concepts. The present authors, however, conceive of the PAQ as a specialized measure of socially desirable instrumental and expressive characteristics, objectively defined trait dimensions that distinguish between the sexes to some degree and thus may be labeled "masculine" and "feminine." Further, they hypothesize that these dimensions have complex and frequently weak relationships with other components of masculinity and femininity, although simultaneously having implications for significant areas of functioning. When the PAQ is viewed from this theoretical perspective, the authors suggest that Locksley and Colten's criticisms directed at its rationale, its empirical properties, and its definition of androgyny become largely inappropriate or irrelevant.

In their article "Psychological Androgyny: A Case of Mistaken Identity?", Locksley and Colten (1979) vigorously criticize the theory of psychological androgyny as well as both the rationale and the methodology of the major instruments used to measure androgyny—the Bem Sex Role Inventory (BSRI; Bem, 1974) and the Personal Attributes Questionnaire (PAQ; Spence & Helmreich, 1978; Spence, Helmreich, & Stapp, 1974, 1975). We find ourselves in agreement with a number of their specific points. However, our overall assessment of the research done with the PAQ in particular leads us to different conclusions, in large part because we view our instrument and its relevance to androgyny theory, and to such broad concepts as masculinity and femininity and sex role identity, from a perspective that differs from that of Locksley and Colten and that of many other androgyny researchers. Within the context of our theoretical perspective, we will discuss some of Locksley and Colten's major arguments, in an effort to clarify further conceptual and methodological issues.

What has come to be known as androgyny theory rests essentially on the proposition that the perpetuation of traditional sex role distinctions is dysfunctional; a society in which differences between the sexes were reduced to the minimum dictated by anatomy would be not only more just but also healthier for its members. A companion proposition, articulated most clearly by Bem (1974), is that masculinity and femininity, so long regarded as terms representing two ends of a single bipolar dimension, the independent dimensions. Both masculine and feminine characteristics may coexist in the same individual, with those individuals exhibiting a relatively high degree of both being labeled "androgy-

Preparation of the manuscript was partially supported by National Science Foundation Grant BNS 76–17376 (Janet T. Spence and Robert L. Helmreich, principal investigators).

Requests for reprints should be sent to Janet T. Spence, Department of Psychology, University of Texas at Austin, Austin, Texas 78712.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3706-1032$00.75

nous." [1] In this theoretical context, masculinity and femininity are global, unidimensional concepts, referring to constellations of covarying attributes and behaviors that normatively distinguish between men and women, and/or to individuals' conceptions of their own masculinity and femininity.

Locksley and Colten (1979) express reservations about accepting the assumption that individuals high in both masculinity and femininity (in these global senses) are better adjusted than sex-typed individuals are, in both a psychological and a sociological sense. Other major criticisms concern the usefulness of these global notions of masculinity and femininity, and the conceptual and methodological deficiencies of the BSRI and PAQ in measuring these constructs. It is to these latter issues that we shall largely address ourselves, confining our discussion to the PAQ, since both its theoretical rationale and its psychometric properties are not identical to those of other androgyny measures.

Although acknowledging at several places that our views and Bem's (e.g., 1975, 1977) are not identical, a pervasive theme underlying Locksley and Colten's critique is that there is a single androgyny theory to which the construction and content of the BSRI and the PAQ, as well as the research done with these instruments, is firmly tied. They also imply that the PAQ, as well as the BSRI, is—or is intended to be—a measure of such general notions as sex identity, sex stereotypes, or sex differences in personality. However, our conception of what the PAQ measures is quite narrow, and both our definition of androgyny and the theoretical suppositions in which our research with the PAQ has been embedded depart in several critical respects from androgyny theory as just described. It is not so much that we disagree with the tenets of that theory as the fact that we are not talking about quite the same things.

To respond to the criticisms offered by Locksley and Colten (1979), it is crucial to present first the theoretical and empirical rationale underlying the development of the PAQ. The 24-item version of the PAQ, used in all of our research following our initial investigation (Spence et al., 1974, 1975), consists of certain clusters of socially desirable socioemotional trait descriptions reflecting what are typically labeled *personality* characteristics, to distinguish them from such other psychological attributes as cognitive abilities or styles, values, attitudes, and so forth. The instructions are general, making no reference to gender, and nothing in the content of the items makes the intent of the instrument obvious. Indeed, the items are ones that appear in identical or similar form on other conventional personality inventories, whose developers and users have, at most, only incidental interest in sex differences.

The development of the PAQ was guided simultaneously by theoretical and empirical considerations. The initial (55-item) version of the scale was drawn from a larger pool of items for which ratings of the typical man and the typical woman (personality stereotypes) and the ideal man and the ideal woman had been obtained. The 55 items chosen for the PAQ were selected from those judged by both sexes to distinguish between the typical man and the typical woman (both adults and individuals of college age).

In using stereotype data to select items for the PAQ, we were in part expressing our conviction that traits about which there was consistent agreement would be a good source of items that would yield actual differences between the sexes and thus permit us to devise an instrument with which we could test a number of hypotheses, such as a dualistic versus a bipolar conception of masculine and feminine *personality traits*. The usefulness of this method of selection is demonstrated by the fact that sex differences in self-reports appeared in the expected direction on every item.

Items were assigned to subscales on the

---

[1] Bem's (1974) androgyny measure, the BSRI, was initially scored by finding the difference between scores on masculinity and femininity scales, and androgyny was defined as a relative balance between the two scores, regardless of their absolute level. We, along with others, have argued that, to correspond more closely to accepted usage, the term should be reserved for those relatively high in both dimensions, a usage that Bem (1977) now accepts and that was also adopted by Locksley and Colten.

basis of these stereotype ratings as well as ratings of the ideal man and the ideal woman. Traits that were stereotypically more characteristic of males but socially desirable to some degree in both sexes were assigned to a Masculinity (M) scale, while traits that were more stereotypically characteristic of females but socially desirable in both sexes were assigned to a Femininity (F) scale. This method of assignment resulted in two groups of items whose content by and large confirmed the speculations of a number of theorists, as well as popular thought, about the fundamental differences between the personalities of men and women. These differences have been characterized by Bakan (1966) as a distinction between masculine *agency* (a sense of self-assertion and self-protectiveness) and feminine *communion* (a sense of selflessness and need to identify with others) and by Parsons and Bales (1955) as a distinction between instrumental and expressive traits. Analyses of the data from several samples of college students indicated significant sex differences in M and F scale scores and also a slightly positive rather than strongly negative correlation between scales in each sex. The latter outcome refuted the implications of the bipolar M–F model of these personality attributes, suggesting instead the usefulness of a dualistic approach.

Use of the ideal ratings resulted in a third scale with more bipolar properties. This scale, labeled Masculine–Feminine (M–F), contains items for which ratings fell toward opposite poles for the ideal man and the ideal woman. The scale is mixed in content, containing items dealing with instrumental qualities, such as "aggressive" and "dominant," which are seen as somewhat desirable for men but not for women, and other items, such as "excitable in a major crisis" and "cries easily," which reflect emotional vulnerability and are seen as somewhat desirable in women but not in men.

Following our initial study, we shortened each scale to eight items, retaining those that showed the best psychometric properties and that also illustrated instrumental and expressive personality traits. (Thus, such items as "athletic" and "good at math and science" were dropped.) The current PAQ may therefore be described as a quite conventional personality test in the self-report mode, consisting of clusters of socially desirable instrumental (masculine) and expressive (feminine) traits. The ultimate justification for the appellations Masculinity and Femininity lies in the demonstration that the clusters of items discriminate between the sexes in their self-reports.

The decision to limit the content of the scales to personality traits describing instrumental and expressive qualities was a theoretical one. Space limitations preclude an explication of the considerations underlying this decision. They led us to conclude, however, that the essential core both of the socially sanctioned and expected differences in the personalities of men and women and of the actual differences between them, as indicated in self-reports, lay within the realm of instrumentality and expressiveness. Implied in this statement is our expectation that the instrumental–expressive conceptualization has considerable transcultural and transgenerational validity in distinguishing between the sexes.[2]

Purportedly biological sex differences in expressive and instrumental personality characteristics have frequently been used to explain traditional instrumental versus expressive role divisions along sexual lines, and acceptance of the functional value of these divisions has led to justifications for socialization practices that attempt to assure the development of a high degree of expressiveness in girls and of instrumentality in boys. Development of a personality scale narrowly confined to "mas-

---

[2] The PAQ does not contain all possible exemplars of instrumental and expressive traits. We do not anticipate that in any given group, actual or expected sex differences will necessarily be found for the entire universe of instrumental and expressive qualities or that exactly the same instrumental and expressive qualities will distinguish between men and women in all cultures and nationalities (see Spence & Helmreich, 1978, for a further discussion of these issues). However, in several completed and ongoing cross-national investigations with the PAQ, limited as it is in content, we have found more similarities than differences with comparable United States samples. As we will discuss shortly, our PAQ results with United States samples are quite consistent across a broad range of chronological ages and socioeconomic groups.

culine" instrumental characteristics and "feminine" expressive characteristics allows exploration of the validity of many suppositions about the relation between these personality dimensions and other types of gender-related phenomena. Our suspicions (which formal evidence is now verifying) were, on the one hand, that the relation between instrumental–expressive personality traits and other gender-related phenomena would turn out to be considerably weaker and/or more complex than had been presumed and, on the other hand, that these trait continua would have important implications for significant areas of functioning without regard to gender.

Independent investigation of the effects of PAQ-defined masculine instrumentality and feminine expressiveness on criterion variables could proceed because the M and F scales were found to be essentially orthogonal. Once the separate relations of M and of F to criterion measures are discovered, their simultaneous contribution to criterion variance becomes important to determine. We have found that the median split method of jointly classifying subjects on M and F scores is a useful conceptual scheme for representing the combined influence of these scales on many dependent variables.[3] For mnemonic purposes, the four resulting cells have been given verbal labels: *Androgynous* (above the median on both M and F), *Masculine* (above the median on M, below on F), *Feminine* (below on M, above on F), and *Undifferentiated* (below on both).

As should be obvious from this description, we introduced the term *androgyny* simply as a convenient label to identify individuals who score relatively high on both the M and the F scales of our particular instrument. Within this context, it is merely a label that corresponds to the definition of the term in standard dictionaries (e.g., "1. having the characteristics or nature of both male and female," *Webster's New Collegiate Dictionary,* 1974). Locksley and Colten's statement that with respect to its "ordinary sense . . . the term *psychological androgyny* clearly implies 'of indeterminate sex, or not sex related' " (1979, p. 1023) is based on an extraordinary sense with which we are not familiar. In any event, androgyny, as we have used the term, has no particular theoretical import, being intended to indicate nothing more than a relatively high degree of both instrumental and expressive personality traits, as defined by the PAQ. The relation between androgyny in this personality sense and other types of androgyny (e.g., interest in both "masculine" and "feminine" hobbies) must be determined empirically.

Locksley and Colten express concern that "most psychological androgyny research has been conducted on college students and even on high school students," and they suggest that "the variance in scores permitting the categorization of androgynous persons may indicate only the lack of press from the defini-

---

[3] Widespread confusion surrounds the use of this method. On the one hand, it has been accepted by some with uncritical enthusiasm as a kind of genuine typology, suitable for use on all occasions. However, the conceptual model that best represents the conjoint influence of M and F must be sought for each new type of data, not assumed a priori. The median-split method has proved to be broadly useful for this purpose, but there is no theoretical reason to anticipate that this will be universally true. On the other hand, the median-split method has come under attack primarily on the grounds of crudeness, loss of information, and inaccuracy of classification into types (e.g., Pedhazur & Tetenbaum, 1979). The frequently suggested solution to the problem is to apply more sophisticated techniques, such as multiple regression, to the data, a suggestion not further amplified. The median-split method is indeed crude and is not the method of choice if one is seeking to maximize predictive accuracy for individual subjects. Our intent, however, has been to provide a conceptual heuristic scheme that will reveal, in a simple and easily communicable form, the nature of the conjoint influence of M and F on the criterion variable. Assuming that the method is properly used; more complex analyses, including regression techniques, should provide no new conceptual information. One should seek other mathematical schemes, such as regression models, when the median-split method proves inadequate or when one is seeking to maximize variance accounted for by using the measures appropriately as continuous variables. In the case of such variables as our measure of self-esteem, the maximum R^2 is produced by an equation containing the partials of M and F and the interaction term ($M \times F$). However, without a priori predictions concerning the joint influence of M and F on criteria, the mechanical use of stepwise regression can lead to conceptually misleading outcomes. See Spence and Helmreich (in press) for an extended discussion of these issues.

tions and constraints of adult sex-segregated social roles" (1979, p. 1021). We have now collected data from more than 1,300 middle-aged adults (Spence & Helmreich, 1978, Note 1). The most striking finding revealed by these data is their parallelism to the data obtained from college students in absolute scores on means and in variance of the scales, in the strength of sex differences, and in the percentage classified into each of the four M–F categories we have defined—a parallelism that exists even when the middle-aged sample is classified using medians drawn from a college population. Significant sex differences have also been found in highly selective populations, for example, between male and female PhD-holding scientists (Beane, 1976), and between male and female athletes (Helmreich, Spence, & Vacalis, Note 2). Further, our data (Spence & Helmreich, 1978) from "even high school students," who were much more heterogeneous in social origins than the samples described above, suggest that the pattern of sex differences, scale correlations, and distributions among PAQ categories is quite impervious to social class. We have recently developed a child's version of the PAQ (Simms, Davis, Foushee, Holahan, Spence, & Helmreich, Note 3) and have found similar sex differences on each scale among 6- and 7-year-old children, as well as variability within each sex.

In criticizing the BSRI, Locksley and Colten (1979) cite factor-analytic studies (e.g., Gaudreau, 1977; Moreland, Gulanick, Montague, & Harren, 1978) that produced four factors, and they suggest that this further weakens the assumption of unidimensional constructs conforming to the original labels assigned to the inventory's scales. Similarly, factor-analytic and discriminant-function data obtained from the BSRI by Pedhazur and Tetenbaum (1979) led these investigators to question the BSRI and the concept of androgyny on both theoretical and methodological grounds. Although there is a fair amount of overlap between the PAQ and the BSRI, and a moderate correlation between parallel scales, there are enough differences between the two instruments to suggest that the type of results found with the BSRI would not necessarily be found with the PAQ. The items contained on the BSRI are not confined to expressive and instrumental traits, and the methods of item-to-scale assignment were somewhat different (resulting in three gender-related scales on the PAQ vs. two on the BSRI). Finally, Gilbert and Strahan (1978) and Pedhazur and Tetenbaum (1979) have provided data suggesting that several of the BSRI F items are judged to be socially *undesirable* for both sexes.[4]

We have undertaken additional psychometric investigations of the PAQ that suggest the unidimensionality of the M and F scales. These investigations included examination of the structure of the item-correlation matrices obtained from high school, college, and adult respondents of each sex along with factor analyses. Using the techniques recommended by Dziuban and Shirkey (1974), the correlation matrices comparing the M and F scales of the PAQ were classified as highly appropriate for factor analysis in each sex in each sample.[5]

Both maximum likelihood and principal factor analyses of the M and F items—computed separately for each sex and each sample—yielded similar results in each instance. Two large, oblique factors emerged, one reproducing the M scale, the other the F scale.[6]

---

[4] The significance of inclusion of such items is suggested by results obtained with our recent extension of the PAQ (EPAQ; Spence, Helmreich, & Holahan, in press) to include Masculinity and Femininity scales containing socially undesirable traits. Very low correlations were found between parallel Masculinity scales and Femininity scales.

[5] The Kaiser-Meyer-Olkin measure of sampling adequacy and Bartletts test of sphericity were computed, and the off-diagonal elements of the anti-image covariance matrix were inspected for the 16 items of the M and F scales and the full 24-item PAQ in six samples, 1,010 female and 757 male high school students, 670 female and 477 male college students, and 243 female and 208 male adults (average age, 36 years). The results for each measure are reported in Helmreich, Spence, and Wilhelm (Note 4).

[6] Comrey (1978) discusses methodological problems in factor-analytic studies. Many of these are evident in analyses of masculinity and femininity measures (see, for example, the discussion by Pedhazur and Tetenbaum, 1979) of the BSRI. Our analyses

Consistent with the correlations obtained with the unit-weighted M and F scales, the two factors had a low positive correlation (average $r = .08$ across the six samples).

Of greatest interest are the results obtained from adult respondents. Our sample consisted of 243 females and 208 males, all of whom were married and the parents of elementary school children. These respondents were clearly familiar with the definitions and constraints of adult, sex-segregated social roles. The fact that the same clusters of traits emerged in this sample provides reassuring evidence for the theoretically postulated trait dimensions. These data further suggest that the derivation of constituent masculine and feminine trait constellations from stereotypically perceived aggregate sex differences is, contrary to Locksley and Colten (1979), a reasonable and valid method of operationally defining these constructs. The soundness of the method is shown in the stability of sex differences across disparate groups as well as in the homogeneity of the trait clusters. (Indeed, the use of perceptions about aggregates to develop measures of individual differences represents a traditional approach to personality test construction.)

In sum, the PAQ is an instrument of restricted content, containing items describing personality traits of an expressive or instrumental nature. Our theoretical conception of such traits is that they are internally located response predispositions or capacities that have considerable transituational significance for behavior but are neither conceptually equivalent to behavior nor its sole determinant. Trait dispositions interact both with situational factors and with other properties of the individual to determine the form and the intensity of his or her responses in any given instance. As Fishbein and Ajzen (1974) have demonstrated, general trait measures are capable of a high degree of predictive accuracy if behavior is aggregated over multiple acts and types of situations. Adequate prediction of singular acts or of behavior in limited settings typically requires additional information.

Overt expression of the personality traits tapped by the PAQ is likely to be exaggerated when these response dispositions are congruent with situational demands (including role-related demands) and likely to be suppressed or modified when they are incongruent. An elegant example may be found in a study by Megargee (1969) in which high- and low-dominant individuals (as predetermined by a personality test) were paired and asked to choose a leader before beginning to work jointly on a laboratory task. In same-sex pairs and in mixed-sex pairs in which the male was high dominant (but *particularly* in the latter), the high-dominant member of the pair was chosen as the leader in the vast majority of instances. In mixed-sex pairs in which the female was high dominant, the low-dominant males ended up as the leaders about three quarters of the time. However, in most pairs, it was the dominant female who selected the leader, almost without exception designating the male. Thus, while the majority of these women conformed to implied sex role pressures, they actively utilized their disposition to dominate by manipulating the situation in a way that preserved the proprieties of male–female interaction. In this type of situation, dominant, assertive women, who neither valued conventional rules for the interactions between the sexes nor anticipated that role violation would result in negative consequences, would presumably have instead let nature (their own) take its course.

Other studies, some of them cited by Locksley and Colten, also demonstrate the effects of role demands on behavior. The fact that behavior is jointly determined both by situational variables and by relatively stable properties of the individual cannot be used as a

---

(Helmreich, Spence, & Wilhelm, Note 4) involved use of several techniques, construction of varying numbers of factors, rotations varying the obliqueness of the solution, and replications in independent samples. The convergence on a two-factor solution for the 16 items of the M and F scales thus seems methodologically justified.

Parallel factor analyses of the full 24-item instrument (including items assigned to the M–F scale) were also computed. These also yielded two main factors, with the F factor remaining unchanged and all of the M–F items loading on the M factor. The resultant M factor combines instrumentality with a lack of emotional vulnerability (i.e., feelings not easily hurt, never cries, etc.).

criticism of personality instruments such as the PAQ; it merely serves as a warning that one should not make predictions about behavior as a function of PAQ scores without taking situational variables into account.

Locksley and Colten take us to task for having neglected to point out (in an earlier version of our reply) that in the Megargee (1969) study, the "effect occurred regardless of whether the task was of a 'masculine' sort, involving industrial mechanical work, or of a 'feminine' sort, involving clerical work and verbal talent" (p. 1027). However, Megargee selected the clerical-verbal task as an example of a *sex-neutral* task, the label of "feminine" being supplied by Locksley and Colten. Our vote goes with Megargee. It remains an open question as to how feminine a task must be before men, as a group, are willing to relinquish leadership and women willing to assume it. Our reading of the attribution literature makes us suspect that when the task becomes one at which women are expected to be more competent than men, the leadership role will be shifted to women, with a concomitant downgrading of the importance of the task and the significance of being leader.

Locksley and Colten go on to comment that an "investigator ought to be able to specify, in a theoretically plausible fashion, those situations in which the personality trait should predict behavior and those situations in which the trait should not predict behavior. Congruence with situational demands is not a valid criterion, at least unless the contribution of situational demands can be distinguished from the contribution of the personality disposition. After all, there is no a priori reason to assume that situational demands always overshadow the influence of personality dispositions" (p. 1027). Apparently Locksley and Colten intend these remarks to be critical of our position, but we are hard put to determine what we are supposed to be disagreeing about. The importance of the Megargee study was in its demonstration that while certain forms of behavior were suppressed by high-dominant women, at the same time dominance behavior appeared in modified form. The study thus illustrated both the importance of situational factors, such as sex role demands, and that of trait dispositions such as measured by the PAQ. Criticizing the PAQ or our conceptions (or those of any investigator) because we cannot yet make detailed predictions about specific situational variables fails to recognize the very primitive state of knowledge in this field. Many of the theoretical presuppositions underlying decades of masculinity–femininity research have been found to be wanting. Those of us in this general area of research are attempting to develop a conceptual framework that will serve as a guide to more fruitful empirical research that will, in turn, allow the development of a more valid and sophisticated theoretical model.

One of the assumptions that has been rejected is that masculine and feminine behaviors or attributes are, of necessity, psychological opposites, two endpoints of a continuum. (We should note, however, that formal demonstration that masculine and feminine qualities are in fact essentially independent has been conducted only for *personality* instruments of limited content, such as the PAQ or the BSRI). A presupposition continuing to dominate this area of research is that gender-related phenomena, whatever their nature or level of operation, tend to covary: preferences or adoptions of the many categories of role-related behavior, sex differences in non-role-related attributes, sexual predilections, and the like. This assumption has led to the use of such broad concepts as sex role identity or masculinity and femininity, without any systematic attempt to decompose them. It has also resulted in attempts to measure these constructs by instruments of omnibus content, surveying a diverse array of gender-related phenomena, or by specialized instruments such as the PAQ or the BSRI with limited content domains. We have been critical both of these broad concepts and of the methods used to measure them. Omnibus measures, as we have noted elsewhere, would be perfectly satisfactory if, as is frequently assumed in both lay and scientific circles, various classes of masculine and feminine attributes and behaviors were highly correlated. Similarly, if such correlations could be demonstrated, it would be legitimate to generalize from a measure of a narrow subclass of mas-

culine and feminine attributes to other subclasses or to the class as a whole. However, a number of sources of evidence suggest that gender-related phenomena are only loosely associated, and that the relationships among various kinds of masculine and feminine behaviors tend to be weak or complexly determined. Thus, global constructs of masculinity and femininity or of sex role identification, based on an additive composite of the specific ways in which an individual resembles the typical member of his or her own sex, are likely to have little utility.

A profitable way to proceed, we have suggested, is to identify the principle components or classes of psychological phenomena related to gender and to devise methods to measure them so that their interrelationships may be determined. Part of our current research program is devoted to the development of additional measures which, like the PAQ or our Attitudes Toward Women Scale (AWS; Spence & Helmreich, 1972), tap a specific subset of sex-related phenomena, for example, overt behaviors in interactions with members of the opposite versus same sex, actual and preferred divisions of responsibilities in various areas within the homes of married couples, and so forth. Although the evidence supports our view that the intercorrelations among these measures and their interactions with situational variables are complex, it is premature to speculate on the exact nature of these complexities. The outcome of these investigations should suggest better articulated theoretical assumptions that should, in turn, suggest the directions in which future empirical research might fruitfully go. (It may turn out, of course, that gender-related phenomena are more tightly bound together than we suspect. This, however, must be demonstrated empirically, not taken for granted.)

Returning to the PAQ and the concept of androgyny, as defined by that instrument, our preceding discussion should make it obvious that we differ in several important respects from the position of Bem and her followers. First, we do not *define* the individual who is androgynous on the PAQ or BSRI (i.e., who has high scores on both the M and F scales) as a person who is "flexible" (i.e., who exhibits both masculine and feminine behaviors) across the general category of sex role behaviors. Nor do we *predict* that these androgynous individuals are likely to differ markedly from others in sex role behaviors that do not directly involve expressive or instrumental skills. Bem, in contrast, implies that the BSRI can be used as a measure of general sex role flexibility and hence assumes that androgynes, as defined by their high scores on the M and F scales, will not be sex-typed in role behaviors.

In her only experimental test of the sex role flexibility hypothesis, Bem (Bem & Lenney, 1976) preselected groups of individuals who, on the BSRI, were balanced ("androgynous," according to her initial definition), extremely sex-typed, or extremely cross-typed. In the experimental setting, subjects indicated their preference for, and comfort in, performing sex-appropriate versus sex-inappropriate tasks such as ironing a napkin or pounding a nail. Sex-typed individuals showed somewhat more avoidance of sex-inappropriate tasks than did the other groups, thus providing some support for the flexibility hypothesis. In a recently completed conceptual replication of the Bem and Lenney study (Helmreich, Spence, & Holahan, in press) using unselected subjects, we have found only minimal correlations between M and F scores on the PAQ and preferences for stereotypically masculine and feminine tasks, and no particular elevation of this type of role flexibility among Androgynous (high M and F) individuals of either sex. (Stronger relations, however, were found with a sex role attitudes measure.)

These findings lend credence to our suspicion that the personality traits measured by the M and F scales of the PAQ and the BSRI are only minimally related to many sex role behaviors that do not quite directly require instrumental or expressive skills. Thus we are extremely hesitant to generalize the findings obtained with these scales to other gender-related phenomena. It should not be assumed without empirical demonstration, for example, that masculine and feminine attributes in domains other than socially desirable instrumental and expressive personality traits form independent, orthogonal dimensions. Nor can

our data on the positive relation between self-esteem and the PAQ M and F scales (Spence & Helmreich, 1978; Spence et al., 1975) be used to conclude that individuals who are androgynous in a role sense are necessarily "healthier" or exhibit higher self-esteem than nonandrogynous individuals.

We indicated earlier that on theoretical grounds, we were impressed that the empirical process by which the PAQ scales were generated supported the conceptions of theorists such as Parsons and Bales (1955) concerning the distinction between masculine instrumental traits and feminine expressive traits. Locksley and Colten, however, express discomfort over this outcome. The instrumental–expressive distinction, they state, is not a good indication of the masculine–feminine dichotomy because "traditionally female activities, such as mothering, require instrumental behavior in order to be carried out effectively" (1979, p. 1021). Once more we see the invidious effects, endemic in the sex role literature, of the failure to distinguish between personality traits and sex role behaviors and responsibilities. Ironically, Talcott Parsons, who originated the proposal that the division in role responsibilities between the sexes could be characterized by the terms *instrumental* and *expressive,* and who also postulated that effective role performance demanded instrumental or expressive qualities of personality, *did not confuse the conceptual distinction between traits and roles.* However, we concur with Locksley and Colten, and, indeed, have said elsewhere (e.g., Spence & Helmreich, 1978), that it is fallacious to assume that the execution of expressive, feminine *roles* requires exclusively expressive personality traits; similarly, implementation of instrumental, masculine roles can benefit from expressive personality characteristics (e.g., management roles involving interpersonal relations). But as long as the difference in the two uses of *instrumental* and *expressive* is understood, each may legitimately be used to describe the differences between masculine versus feminine *roles,* and between masculine versus feminine *personality traits.*

We also reject the notion that, on the level of personality, a high degree of instrumentality is necessarily required for the effective performance of all stereotypically masculine tasks or a high degree of expressiveness for the performance of all stereotypically feminine tasks. Nonetheless, task analyses suggest that there is some degree of conjunction and that Parsonian theory, while perhaps overdrawn, is not without foundation. For example, some female-dominated professions, such as elementary education and social work, place a special premium on expressive or communal capacities and motives. There is also evidence (Wertheim, Widom, & Wortzel, 1978) to suggest that individuals of both sexes who are attracted to such professions score higher on the PAQ F scale than do their same-sex peers entering male-dominated professions. The latter, in turn, score higher on the M (instrumental) scale. Similarly, we have found that academic scientists, particularly female scientists, are higher in M (but no different in F) than unselected samples of men and women.

Data such as these suggest that instrumental and expressive characteristics have important implications for functioning in a number of significant areas. They further suggest that many activities that are normatively masculine (i.e., male dominated) demand instrumental attributes for their execution (whether or not they also benefit from expressivity) and that many activities that are normatively feminine (female dominated) demand expressive talents (whether or not they also demand instrumental qualities). Evidence also shows that differential socialization practices employed with boys and with girls actively encourage the development of instrumental attributes in the boys and expressive attributes in the girls (e.g., Block, 1973), in the at least partially correct belief that these personal qualities will be required for the discharge of traditional sex roles. The wisdom and justice of perpetuating sex role distinctions aside, there is failure to recognize that functioning may be enhanced by both types of qualities. Even worse, children—boys in particular—may be actively discouraged from developing and manifesting cross-sex characteristics, in the mistaken belief that instrumentality and expressivity preclude one an-

other (Foushee, Helmreich, & Spence, 1979).

Locksley and Colten (1979), however, denigrate the value of identifying the instrumental–expressive personality distinction as one component of masculinity–femininity by attacking the use of stereotype data to derive the PAQ and BSRI scales. A variety of arguments is advanced to buttress this position, several of them following from the assumption that the intent of these instruments was to measure global self-concepts of masculinity and femininity and/or concepts of masculinity and femininity in prototypic males and females. We will not respond to this set of arguments because, as we have indicated, our theoretical conception of the PAQ is that it is a more specialized instrument.

Other criticisms, however, are germane. They dispute the notion of a general stereotype, even on the specialized level of personality-trait attributions, citing as evidence a study by Clifton, McGrath, and Wick (1976). These investigators provided respondents with a list of adjectives, asking them to check those items that characterized the typical housewife, bunny, club woman, career woman, and woman athlete. Little commonality was found across roles in adjective choices. These results, while interesting for the light they shed on specific stereotypes, are quite to be expected. To use another example, one would not be surprised to find that, if asked to characterize men who are professional football players, people would agree that they are taller, stronger, and have fewer front teeth than nonathletically inclined males of the same age (all stereotypes that objective evidence would doubtless show are valid). This demonstration would not destroy the importance of the known difference between men and women in height and muscular strength, nor would it tell us much about people's beliefs (or the objective facts) about the number of missing front teeth in the typical member of each sex. We would also expect that female football players would be judged to be bigger and stronger than women in general—but not as big or strong as men. The existence of within-groups stereotypes does not demonstrate the nonexistence of between-groups stereotypes.

Locksley and Colten (1979), however, appear to argue from the Clifton et al. (1976) data that general stereotypes about the personality traits of the typical or the ideal man and of the typical or the ideal woman do not exist, or that, if they do exist, they are to be found only among high school and college students and others who have not experienced adult role constraints and conceptions, and/or that they represent stereotypes about the prototypic "housewife/mother and career man." In developing the PAQ, we obtained two types of stereotypic ratings (from independent samples), one for the typical male and the female adult and the other for the typical male and the female college student (who has not yet married or launched upon a career). Only those items for which significant stereotyping occurred under both sets of rating instructions were chosen for our questionnaire. Further, inspection of stereotype data obtained from older samples (e.g., Broverman, Vogel, Broverman, Clarkson, & Rosenkrantz, 1972; Rosenkrantz, Vogel, Bee, Broverman, & Broverman, 1968) does not confirm the contention that these stereotypes can be found only among younger inexperienced groups. Although cultural and subcultural differences in stereotypes about specific traits may well exist (Spence & Helmreich, 1978), what is most striking about the extant findings is the amount of agreement, over a wide age span of raters and rating targets, about the types of characteristics that distinguish the sexes. Finally, as we have noted, sex differences in self-reports on the PAQ M and F scales have also been found not merely in college students but also in diverse samples varying widely in age, ethnicity, and socioeconomic background.

Locksley and Colten claim further that "general sex stereotypes may be too global for interpreting and guiding behavior at the level of individual self-perception or self-direction" (1979, pp. 1021–1022). Two issues are involved here: the influence of sex stereotypes in guiding reactions toward others and in guiding reactions toward the self. With respect to the former, the danger of stereotypes

is not only that they tend to exaggerate both the number of temperamental characteristics that differentiate the sexes and the magnitude of these differences, but also that they too often result in crucial policy decisions discriminating against women and girls, based on the actual or supposed characteristics of females as a class versus males as a class. Women should be excluded from positions of authority because they are too emotional; girls should not be allowed to play Little League baseball because they are too fragile; women should not be admitted to graduate school because they are not highly motivated and are likely to drop out to get married and have babies; and so forth ad nauseum. Determination of the conditions under which information supplied about an individual man or woman will reduce the role of sex stereotypes in judgments about that individual is an important area of investigation. However, instances in which sex stereotypes have been the primary determinant in reaching decisions are too numerous and too serious to dismiss as trivial the influence of general stereotypes on people's lives.

The influence of sex stereotypes about masculine versus feminine personality characteristics in guiding self-behavior is a complex question. We should point out first that we as investigators, rather than the respondents, have labeled the PAQ scales "masculine" and "feminine," labels justified by the fact that sex differences in self-report are consistently found. The same appellation can be applied to other types of personal qualities and behavior that objectively differentiate the sexes, whether or not there is general awareness about the existence of particular types of sex differences and whether or not they are positively sanctioned. But, as we have written elsewhere:

> Masculinity and femininity can also be regarded as global aspects of the self-concepts that men and women directly identify in these or equivalent terms (such as being a "real man" or a "real woman").

We then went on to state:

> Individuals . . . have organized belief systems, however poorly articulated, about the psychosocial meaning of being "a man" or "a woman" and can be expected to have incorporated these beliefs into their sense of self. These belief systems, our observations suggest, are compounded of many elements: assumptions about appropriate sex-roles, characteristics of the self such as personality attributes and cognitive abilities, physique and physical appearance, styles of speech and body movement, sexual behavior, and so forth. One can expect that individuals belonging to the same culture or subculture will be reasonably similar in identifying the factors contributing to masculinity and femininity in themselves and others in the sense that they will be drawn from a common pool and will not be completely idiosyncratic. Individuals' definitions of masculinity and femininity, however, are likely to be based on complexly weighted sets of indicators that not only vary from one person to another but also change with age and may even differ when individuals are assessing themselves as opposed to others. (Spence & Helmreich, 1978, pp. 115–116)

How salient instrumental and expressive characteristics tend to be in the modal individual's self-concept of masculinity or femininity remains problematical. Hence the availability of these characteristics as models for more or less self-conscious guidance of behavior in situations in which gender per se is salient is unknown.

We have also suggested the importance of determining the nature of individuals' self-concepts of their own masculinity or femininity and their global conceptions of masculinity or femininity in others, as well as the importance of outlining some of the elements that might enter into these conceptions. Locksley and Colten have made the valuable proposal that techniques from cognitive psychology could fruitfully be employed to discover the structure of these global concepts. We regard such efforts as complementing—rather than supplanting—the study of masculinity and femininity, objectively defined from the perspective of the investigator. Whatever the place of instrumentality and expressivity in these global concepts, the fact remains that these personality dimensions have been shown to be significant determinants of behavior and that they do differentiate the sexes to some degree.

In the final section of their critique, Locksley and Colten observe that research and theory on masculinity and femininity have always emphasized the contribution of sex roles to adaptation or adjustment. In fact, they state, claims for the validity of the constructs

have often been based on their ability to discriminate between the adjustment levels of criterion groups. In contemporary theory, superior adjustment is claimed for androgynous individuals, whereas previously, superior adjustment was attributed to sex-typed men and women. The authors contend that this emphasis on adjustment by current androgyny theorists has led to two major problems. The first, they suggest, is a confusion between predictive validity and construct validity. (As they define the terms, predictive validity refers to a confirmation of the hypotheses derived from a theory, whereas construct validity refers to demonstrations that a measure taps what it purports to measure.) They also imply that efforts to establish the construct validity of the BSRI and the PAQ have been lacking, predictive validity having been mistaken for construct validity. The second problem, they state, reflects a confusion between psychological and sociological levels of analysis.

We agree that there has been confusion between construct and predictive validity, but our analysis differs from Locksley and Colten's. With the exception of her study with Lenney (Bem & Lenney, 1967), Bem's experimental studies have set up situations that were designed to elicit quite directly either instrumental or expressive behaviors. The results of these studies (Bem, 1977) generally showed that in situations in which instrumental behaviors were elicited, individuals with high M scores on the BSRI (Androgynous and Masculine individuals, according to the four-way categorical scheme) showed more of that type of behavior than did those low in M; similarly, in situations in which expressive behavior was appropriate, individuals high in F (Androgynous and Feminine) showed more of that type of behavior than did those low in F. Alone among the categorical groups, Androgynous subjects were thus high in both types of behavior. These data simultaneously provide evidence for the construct validity of the BSRI M and F scales as well as a behavioral demonstration of what was found in the correlational analysis of M and F self-report scores: namely, that it is possible to possess a high degree of both instrumental and expressive qualities.

In the Bem and Lenney (1976) study, a relation was determined between BSRI scores and preference for simple gender-related tasks whose performance had no discernible connection with instrumentality and expressiveness. Whether this study is relevant to the construct or predictive validity of the BSRI depends on one's theory. Bem apparently regards her instrument as a general measure of sex role preferences and behaviors. According to this theoretical conception of the BSRI, demonstration that Androgynous individuals show less sex typing in reaction to gender-related tasks than do other types of individuals (or that individuals who are sex typed are more sex typed in their task reactions) would thus fall into the category of construct validity. Our conception of the BSRI, and even more particularly of the PAQ, is that these measures primarily tap instrumental and expressive trait dispositions. Our theory further leads us to anticipate, at best, only minimal relation between M and F scores and the type of tasks devised by Bem and Lenney (1976). The evidence seems to favor the weak relation postulated by our theory rather than the strong ones postulated by Bem. While resolution of these rival views awaits additional data, the methodological point is clear: according to our theory, this type of experiment is addressed to predictive, not construct, validity. Negative results (i.e., failure to find a relation to behavior as a function of PAQ or BSRI scores) would strengthen the predictive validity of our theory but would not undermine the construct validity of these instruments, as we conceptualize them.

In commenting on our contention that the masculine and feminine characteristics tapped by the PAQ cannot necessarily be assumed to be highly related to other components of masculinity and femininity, Locksley and Colten note that it is "unsatisfactory to show that the . . . PAQ scales simply tap the traits they contain" (1979, p. 1028, Footnote 2). This is not the place to review a growing literature that shows the relation between the PAQ scales and a number of other types of self-report and behavioral measures (e.g., Helmreich & Spence, 1978; Klein, 1978; Spence & Helmreich, 1978). We note only that these

other measures sometimes show significant main effects for sex of subject and sometimes do not; when sex differences do occur, they often—but not universally—disappear when PAQ scores are taken into account. The instrumental and expressive characteristics measured by the PAQ show promise of contributing to our understanding of significant real-life behaviors, regardless of the gender of the individual in which they occur, and also of helping us to unravel some of the mysteries of sex differences.

One of our early findings, replicated numerous times with many types of samples, is that both M and F scores (but particularly the former) were positively correlated with a self-report measure of social competence and self-esteem. Androgynous individuals of both sexes —those relatively high in both M and F— scored highest, followed by Masculine, Feminine, and Undifferentiated. Similarly, we have found that clients of a student counseling center scored markedly lower on the M and F scales than unselected students did. We can find nothing that we have said about these data or about the rationale of the PAQ that would permit Locksley and Colten's curious conclusion that we regard these findings as support for the construct validity of our questionnaire.

We indicated above that empirical data gathered with the PAQ suggest that both the M and F scales are positively related to self-esteem and negatively related to neuroticism, anxiety, and other problem behaviors. Since M is more strongly correlated with the indices of adjustment than F, men and women classified as Masculine (high M, low F) receive scores that are only slightly different from those classified as Androgynous (high M, high F). Locksley and Colten, however, suggest that feminine characteristics are not merely more weakly related to adjustment than are masculine characteristics but that the relationship may be negative in direction, thus giving the edge to Masculine rather than to Androgynous individuals. In support of this claim, they cite a recent study by Jones, Chernovetz, and Hansson (1978), employing the BSRI, in which "masculine" subjects were found to have greater self-esteem, fewer neurotic and alcohol problems, and the like, than "androgynous" subjects had. The composition of these groups, however, was based on the now discarded subtractive method of scoring the BSRI, in which androgyny is defined as a balance between M and F, regardless of level, rather than as a relatively high degree of both M and F. However, Jones et al. (1978) went on to isolate groups of subjects who corresponded to our definitions of androgynous and undifferentiated subjects (above the median on both scales vs. below the median on both scales). Although not always significant, the means of indices of adjustment were uniformly higher in both sexes for the androgynous individuals than for the Undifferentiated individuals. What a comparison of Androgynous and Masculine groups would have shown is unknown. Even more troublesome, from our perspective, is the use of the BSRI rather than the PAQ. The BSRI F scale contains a number of items (some of which are listed by Locksley and Colten) that are not clearly expressive, communal traits, as we conceive of them, and/or, as we noted earlier, are socially undesirable for women as well as men. In our recent work with an expanded version of the PAQ that includes scales tapping masculine and feminine traits that are socially undesirable, we have found correlations with measures of neuroticism and acting-out behaviors that are opposite in sign for socially desirable and undesirable characteristics (Spence et al., in press).

It is also critical to note that we have not used the data indicating a positive relation between the socially desirable M and F scales and several measures of adjustment and self-esteem to predict that Androgyny (as defined by the PAQ) will always lead to the most "desirable" or "adaptive" behavior. Our reluctance is twofold. First, our empirical findings suggest that, whether considered separately or in combination, the M and F scales do not always relate to criterion variables in exactly the same way with respect either to magnitude or to direction. Therefore, the order of the categorical groups in terms of criterion responses does not always put Androgynous subjects at one end of a continuum and Undifferentiated subjects at the other. Further, as Locksley and Colten

(1979) note, terms such as "adaptive" or "effective" are often value laden and politicized. One example from our research will serve as an illustration. We have recently developed a multifactor measure of achievement motivation (Helmreich & Spence, 1978) that has demonstrated considerable construct and predictive validity. Significant differences in achievement-scale scores as a function of PAQ M and F category have also been found. For one of the scales, Competitiveness, the order in each sex is Masculine, Androgynous, Undifferentiated, and Feminine. Dependent on political philosophy, this result could be viewed as evidence for the superiority of the Androgynous individual over the Masculine—or the reverse. We have for these reasons rejected the denotations *androgyny researchers* or *androgyny theorists* as descriptions of our own efforts, writing:

> We are wary . . . of the surplus meaning, both ideological and empirical, that has accrued to "androgyny" as a label for the general model. . . . Our predictive model is . . . an open, evolving, dualistic one and in this sense is not "androgynous." (Spence & Helmreich, 1978, p. 109)

In suggesting a confusion between psychological and sociological levels of analysis, Locksley and Colten indicate that individuals who ignore sex-contingent expectations may do so at great cost and in this sense they may show behavior that is dysfunctional and maladaptive. Self-acknowledged androgyny theorists would be the first to admit, indeed insist, that this is true in contemporary society. In the long view, they argue that a society in which sex role distinctions were minimized would be healthier for all of its members. In the world of the here and now, they deny that those who fly in the face of traditional expectations on what Locksley and Colten describe as the sociological level are, on the psychological level, neurotic misfits, uncertain in their gender identification, sexually abnormal, and so forth. To the contrary, such individuals may be "healthier" than those who conform to conventional norms.

These speculations may well be valid. However, individuals who are androgynous in a role sense are not necessarily the same individuals who are androgynous on the PAQ or the BSRI. Data bearing on the hypotheses of androgyny theorists, to say nothing of Locksley and Colten's counterpropositions, have yet to be gathered.

Conclusions

From our theoretical perspective, the search for global measures of masculinity and femininity or sex role identity—measures that allow individuals to be placed along a quantitative dimension—is a snare and a delusion. The classes of psychological attributes and behavioral patterns that distinguish between men and women at a given time and a given culture are not only multitudinous but also may have different roots and may vary relatively independently across individuals. Still different findings may emerge when individuals' conceptions of their own masculinity or femininity are probed. There are many masculinities and femininities, and androgyny has many identities. The masculinity and femininity measured by the PAQ refers to a limited set of socially desirable instrumental and expressive personality traits, which are but one possible component of the global self-conceptions that Locksley and Colten discussed. Thus, from our theoretical vantage point, many of Locksley and Colten's methodological criticisms are rendered irrelevant. Androgyny on the level of sex role behaviors (the area of inquiry to which their critique is primarily directed) is a worthy topic of investigation that awaits the development of more appropriate measuring instruments than the BSRI or the PAQ.

Reference Notes

1. Spence, J. T., & Helmreich, R. L. *Familial antecedents of masculinity, femininity, and achievement motivation.* National Science Foundation project, in progress, 1979.
2. Helmreich, R. L., Spence, J. T., & Vacalis, D. *Masculinity, femininity, and achievement motivation in varsity athletes.* Manuscript in preparation, 1979.
3. Simms, R. E., Davis, M. H., Foushee, H. C., Holahan, C. K., Spence, J. T., & Helmreich, R. L. *Psychological masculinity and femininity in children and their relationships to trait stereotypes and toy preference.* Paper presented at the meeting of the Southwestern Psychological Association, New Orleans, April 1978.
4. Helmreich, R. L., Spence, J. T., & Wilhelm, J. A.

A psychometric analysis of the Personal Attributes Questionnaire. *Sex Roles,* in press.

References

Bakan, D. *The duality of human existence.* Chicago: Rand McNally, 1966.

Beane, W. *Life history and personality variables in scientific attainment.* Unpublished doctoral dissertation, University of Texas at Austin, 1976.

Bem, S. L. The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology,* 1974, *42,* 155–162.

Bem, S. L. Sex role adaptability: One consequence of psychological androgyny. *Journal of Personality and Social Psychology,* 1975, *31,* 634–643.

Bem, S. L. On the utility of alternate procedures for assessing psychological androgyny. *Journal of Consulting and Clinical Psychology,* 1977, *45,* 196–205.

Bem, S. L., & Lenney, E. Sex-typing and the avoidance of cross-sex behavior. *Journal of Personality and Social Psychology,* 1976, *33,* 48–54.

Bem, S. L., Martyna, W., & Watson, C. Sex-typing and androgyny: Further explorations of the expressive domain. *Journal of Personality and Social Psychology,* 1976, *34,* 1016–1023.

Block, J. H. Conceptions of sex roles: Some cross-cultural and longitudinal perspectives. *American Psychologist,* 1973, *28,* 512–526.

Broverman, I., Vogel, S., Broverman, D., Clarkson, F., & Rosenkrantz, P. Sex-role stereotypes: A current appraisal. *Journal of Social Issues,* 1972, *28,* 59–78.

Clifton, A. K., McGrath, D., & Wick, B. Stereotypes of women: A single category? *Sex Roles,* 1976, *2,* 135–148.

Comrey, A. Common methodological problems in factor analytic studies. *Journal of Consulting and Clinical Psychology,* 1978, *46,* 648–659.

Dziuban, C. D., & Shirkey, E. C. When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychological Bulletin,* 1974, *81,* 358–361.

Fishbein, M., & Ajzen, I. Attitudes towards objects as predictors of single and multiple behavioral criteria. *Psychological Review,* 1974, *81,* 26–43.

Foushee, H. C., Helmreich, R. L., & Spence, J. T. Implicit theories of masculinity and femininity: Dualistic or bipolar? *Psychology of Women Quarterly,* 1979, *3,* 259–269.

Gaudreau, P. Factor analysis of the Bem Sex Role Inventory. *Journal of Consulting and Clinical Psychology,* 1977, *45,* 299–302.

Gilbert, L., Strahan, R. F., & Deutsch, C. L. Feminine and masculine dimensions of the typical, desirable and ideal woman and man. *Sex Roles,* 1978, *4,* 767–778.

Helmreich, R. L., & Spence, J. T. The Work and Family Orientation Questionnaire: An objective instrument to assess components of achievement motivation and attitudes toward family and career. JSAS *Catalog of Selected Documents in Psychology,* 1978, *8,* 35. (Ms. No. 1677)

Helmreich, R. L., Spence, J. T., & Holahan, C. K. Psychological androgyny and sex role-flexibility: A test of two hypotheses. *Journal of Personality and Social Psychology,* in press,

Jones, W., Chernovetz, M. E., & Hanssen, R. O. The enigma of androgyny: Differential implications for males and females? *Journal of Consulting and Clinical Psychology,* 1978, *46,* 298–313.

Klein, H. *Psychological masculinity and femininity, self-consciousness, and typical and maximal dominance expression in women.* Unpublished doctoral dissertation, University of Texas at Austin, 1978.

Locksley, A., & Colten, M. E. Psychological androgyny: A case of mistaken identity? *Journal of Personality and Social Psychology,* 1979, *37,* 1017–1031.

Megargee, E. I. Influence of sex roles on the manifestation of leadership. *Journal of Applied Psychology,* 1969, *53,* 377–382.

Moreland, J. R., Gulanick, N., Montague, E. K., & Harren, V. A. Some psychometric properties of the Bem Sex Role Inventory. *Applied Psychological Measurement,* 1978, *2,* 249–256.

Parsons, T., & Bales, R. F. *Family socialization and interaction process.* Glencoe: Free Press, 1955.

Pedhazur, E. J., & Tetenbaum, T. J. Bem Sex Role Inventory: A theoretical and methodological critique. *Journal of Personality and Social Psychology,* 1979, *37,* 996–1016.

Rosenkrantz, P. S., Vogel, S. R., Bee, H., Broverman, I. K., & Broverman, D. M. Sex-role stereotypes and self-concepts in college students. *Journal of Consulting and Clinical Psychology,* 1968, *32,* 287–295.

Spence, J. T., & Helmreich, R. L. The Attitudes Toward Women Scale: An objective instrument to measure attitudes toward the rights and roles of women in contemporary society. JSAS *Catalog of Selected Documents in Psychology,* 1972, *2,* 66. (Ms. No. 153)

Spence, J. T., & Helmreich, R. L. *Masculinity and femininity: Their psychological dimensions, correlates and antecedents.* Austin: University of Texas Press, 1978.

Spence, J. T., & Helmreich, R. L. On assessing "androgyny." *Sex Roles,* in press.

Spence, J. T., Helmreich, R. L., & Holahan, C. K. Negative and positive components of psychological masculinity and femininity and their relationship to neurotic and acting-out behaviors. *Journal of Personality and Social Psychology,* in press.

Spence, J. T., Helmreich, R., & Stapp, J. The Personal Attributes Questionnaire: A measure of sex-role stereotypes and masculinity–femininity. JSAS *Catalog of Selected Documents in Psychology,* 1974, *4,* 43. (Ms. No. 617)

Spence, J. T., Helmreich, R., & Stapp, J. Ratings of self and peers on sex role attributes and their relation to self-esteem and conceptions of masculinity and femininity. *Journal of Personality and Social Psychology,* 1975, *32,* 29–39.

Wertheim, E. G., Widom, C. S., & Wortzel, L. H. Multivariate analysis of professional career choice correlates. *Journal of Applied Psychology,* 1978, *63,* 234–242.

Received June 2, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 6, 1047–1054

Theory and Measurement of Androgyny: A Reply to the Pedhazur–Tetenbaum and Locksley–Colten Critiques

Sandra Lipsitz Bem
Cornell University

The theory behind both the Bem Sex Role Inventory (BSRI) and Bem's androgyny research, as well as particular issues raised in the critiques of Pedhazur and Tetenbaum and of Locksley and Colten, are discussed. It is noted that (a) the BSRI is based on a theory about both the cognitive processing and the motivational dynamics of sex-typed and androgynous individuals; (b) the strategy of item selection for the BSRI followed directly from the theory and utilized established techniques for test construction; (c) a short BSRI has been developed in accordance with the results of various factor analyses; (d) current research is testing the hypothesis that sex-typed and androgynous individuals differ in the extent to which gender serves as a cognitive schema; and (e) the concept of androgyny contains an inner contradiction and hence a built-in obsolescence.

The methodological critique of the Bem Sex Role Inventory (BSRI) by Pedhazur and Tetenbaum (1979) rests on a misunderstanding of both the purpose of the instrument and the theory underlying it. The more conceptual critique by Locksley and Colten (1979) reduces in most of its particulars to empirical questions about the generality of the concept of androgyny and about the domain of applicability of its associated measuring instruments. Accordingly, this response begins with an explication of the theory on which the BSRI is based, then proceeds to a discussion of some of the specifics of the Pedhazur–Tetenbaum critique, and ends with an overview of research currently in progress, research that ought to answer some of the questions raised by Locksley and Colten.

Preparation of this article was supported in part by National Science Foundation Grant BNS-78-22637 to Sandra Lipsitz Bem. The author would like to express her thanks to Daryl J. Bem for critical comments on the manuscript and to Richard Darlington for sharing his statistical expertise.

Requests for reprints should be sent to Sandra Lipsitz Bem, Department of Psychology, Uris Hall, Cornell University, Ithaca, New York 14853.

The Theoretical Rationale Behind the BSRI

The recent debate in personality–social psychology over the degree to which individuals show transituational consistencies in their behavior has led to the suggestion that we ought to reverse our usual assumption of consistency as given and inconsistency as problematic and, instead, adopt the view that it is the phenomenon of consistency that requires explanation (D. J. Bem, 1972). Such an approach shifts the burden of proof and leads us to ask why a person's behavior might display consistency rather than why it does not. Such an approach also suggests that behavioral consistency might itself be an important individual difference variable (e.g., Bem & Allen, 1974; Campus, 1974).

In the special case of sex roles, this shift of emphasis brings two idealized groups of individuals into focus: those "sex-typed" individuals who restrict their behavior in accordance with cultural definitions of sex-appropriate behavior, and those "androgynous" individuals who do not. This leads us for the first time to view the situational adaptability ("inconsistency") of the androgynous group

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3706-1047$00.75

as the "unmarked" norm—the given, the baseline—and to regard the sex-stereotyped consistency of the sex-typed group as "marked" or problematic, as the phenomenon to be explained.

Because the BSRI was developed to "capture" these particular groups of individuals, its construction was based on two specific theoretical assumptions: (a) Largely as a result of historical accident, the culture has clustered a quite heterogeneous collection of attributes into two mutually exclusive categories, each category considered both more characteristic of and more desirable for one or the other of the two sexes. These cultural expectations and prescriptions are well known by virtually all members of the culture. (b) Individuals differ from one another in the extent to which they utilize these cultural definitions as idealized standards of femininity and masculinity against which their own personality and behavior are to be evaluated. In particular, the sex-typed individual is highly attuned to these definitions and is motivated to keep her or his behavior consistent with them, a goal she or he presumably accomplishes both by selecting behaviors and attributes that enhance the image and by avoiding behaviors and attributes that violate the image. In contrast, the androgynous individual is less attuned to these cultural definitions of femininity and masculinity and is less likely to regulate her or his behavior in accordance with them. The BSRI is thus based on a theory about both the cognitive processing and the motivational dynamics of sex-typed and androgynous individuals. Moreover, empirical research on the behavioral correlates of sex typing and androgyny has so far confirmed that it is serving its intended conceptual purposes (cf. S. L. Bem, 1975; Bem & Lenney, 1976; Bem, Martyna, & Watson, 1976; Ickes & Barnes, 1978; Russell, 1978).

The Construction of the BSRI

The strategy utilized in the construction of the BSRI follows directly from these theoretical premises. Because the BSRI is founded on a conception of the sex-typed individual as someone who is highly attuned to cultural definitions of sex-appropriate behavior and who uses such definitions as the ideal standard against which her or his own behavior is to be judged, items for the BSRI were selected not on the basis of sex differences in self-report, as most previous masculinity–femininity inventories have been compiled, but on the basis of judges' ratings of the culturally defined desirability of various attributes for each of the two sexes. In contrast to self-reports on such items, these cultural definitions were expected to be widely known and to be quite stable across time and from sample to sample. The BSRI is thus designed to assess the extent to which the culture's definitions of desirable female and male attributes are reflected in an individual's self-description. More specifically, the BSRI presents an individual with a heterogeneous collection of attributes and assesses the extent to which the individual clusters this collection into the two categories designated by the culture as more desirable for one or the other of the two sexes.

In order to select items for the Femininity and Masculinity scales of the BSRI, undergraduate judges were asked to rate the desirability of approximately 200 personality attributes either "for a woman" or "for a man." No judge was asked to rate both. A personality characteristic was then defined as feminine or masculine (and hence eligible for the Femininity and Masculinity scales of the BSRI) if, and only if, it was judged to be significantly more desirable in American society for one sex than for the other by four independent samples of judges. The assumption that the culture's definitions of desirable female and male personalities are widely known implies that virtually any sample of American adults would be qualified to serve as knowledgeable informants with respect to these cultural definitions. Moreover, our instructions to the judges emphasized that we were interested not in the judges' personal opinions of how desirable these various attributes were but in their judgment of how American society would evaluate the various attributes. The consistency of the ratings across four independent samples of judges is strong evidence that the BSRI is tapping

widely known cultural definitions, as stipulated by the theory.

As Pedhazur and Tetenbaum (1979) point out, there has recently been a discussion in the literature as to how best to score the BSRI. When the BSRI was first developed, I set forth a definition of androgyny based on the *t* ratio between an individual's endorsement of feminine and masculine attributes, with small *t* ratios indicating androgyny and large *t* ratios (significant differences) indicating sex typing (S. L. Bem, 1974). This method of scoring the BSRI follows directly from the theory in that it distinguishes between those individuals who cluster the attributes on the BSRI into the two categories designated by the culture as more desirable for one or the other of the two sexes, and those individuals who do not. The later emergence of *behavioral* differences between those who score high on both the Femininity and the Masculinity scales (the androgynous high-highs) and those who score low on both (the "undifferentiated" low-lows)—even though both groups achieve small *t* ratios—led to the suggestion that the BSRI be scored on the basis of a median split on both dimensions, a scoring procedure yielding four rather than three distinct groups of individuals (S. L. Bem, 1977). For empirical purposes, this was an entirely sensible proposal, but it did tend to obscure the original theoretical rationale behind the BSRI.

The Pedhazur and Tetenbaum Critique

The BSRI as Atheoretical

The most pervasive criticism advanced by Pedhazur and Tetenbaum (1979) is that the BSRI is atheoretical: "Instead of defining the domains of masculinity and femininity and attempting to construct measures consistent with the definitions, Bem has chosen a strictly empirical approach" (p. 998), an approach that "was destined to fail" (p. 1012). The above discussion, however, contradicts Pedhazur and Tetenbaum's conclusion. The theory underlying the BSRI asserts that sex-typed individuals will conform to whatever definitions of femininity and masculinity the culture happens to provide. The theory deliberately does not specify the particular contents of these definitions, however, because these will vary from culture to culture. The theory is a theory of process, not content, and the use of judges as "native informants" about the particular contents of the culture's prescriptions flows directly from the theory itself. Contrary to Pedhazur and Tetenbaum, this is not dust-bowl empiricism.

Closely related to the criticism that the BSRI is atheoretical is the objection that the Femininity and Masculinity scales are not unidimensional, and that one should employ a methodology for constructing the scales that would guarantee their unidimensionality. But Pedhazur and Tetenbaum are putting the methodological cart before the theoretical horse. The culture has arbitrarily clustered together heterogeneous collections of attributes into the two categories prescribed as more desirable for one sex or the other. The very concept of androgyny is a positive assertion that these arbitrary clusters of apples and oranges need not—and for some individuals do not—"hang together." If the culture groups a hodgepodge of attributes into a category it calls "femininity" or "masculinity," then that hodgepodge is what sex-typed individuals will take as the standard for their behavior. The purpose of the BSRI is to discriminate between those individuals for whom this hodgepodge does form a unitary cluster and those individuals for whom it does not.

Multiple t *Tests as Inappropriate for Item Selection*

Pedhazur and Tetenbaum (1979) criticize the use of item-by-item *t* tests as the basis for item selection, presumably because of a concern that this strategy might capitalize on chance findings. They did not note, however, that the initial list of 200 personality characteristics was rated by four independent groups of judges and that an item was defined as feminine or masculine if, and only if, it was consistently and reliably rated as significantly more desirable for one or the other of the two sexes by all four groups of judges. The probability of this occurring by chance for any individual item is 1/160,000; the number of items out of 200 expected by

chance to reach significance in all four groups is .00125. Moreover, a recent replication at the University of Washington cross-validated this pattern for 37 of the 40 items on the BSRI (Walkup & Abbott, 1978). The results for the three exceptions were in the predicted direction, but reached significance only for female judges. Accordingly, the BSRI would appear to tap relatively enduring definitions of sex-appropriate behavior, culturally defined standards of sex-appropriate attributes that have not given way even in the face of a strong feminist critique in the culture at large. (The BSRI also contains 20 items that were judged by the original four samples of judges to be no more desirable in American society for one sex than for the other. The Walkup and Abbott replication cross-validated this "neutral" pattern for only about half of these items, a finding of little practical or theoretical significance because these items serve as filler and do not enter into the assessment of an individual's sex role.) [1]

Moreover, the use of item-by-item tests (whether *t* tests, *D*, or item-total correlations) is established practice in test construction (Anastasi, 1968), and it is the use of stepwise discriminant analysis that seems quite curious in this context. Because discriminant analysis weights each item only with respect to the amount of *incremental* discriminating power it has over and above the discriminating power of the items already entered into the equation, discriminant analysis is exceedingly sensitive to the composition of the particular items in the item pool. In the present case, for example, discriminant analysis would have yielded dramatically different weightings if the two items "feminine" and "masculine" had been excluded from the initial item pool, but the content of the culture's definitions would not have changed. Indeed, I suspect that even these two items would have been eliminated by a discriminant analysis if we had simply included the items "female" and "male" in the initial item pool. The goal of the social desirability ratings was to identify a comprehensive complement of items that the culture consistently and reliably designates as more appropriate for one or the other of the two sexes. Given this goal, it is simply not meaningful to conclude—as a discriminant analysis suggests we must—that only the two items "feminine" and "masculine" need to be considered.

In fact, there is now a body of accumulated evidence that suggests that—far from being the best two items on the scale—"feminine" and "masculine" are actually the worst two items on the scale. It is true, as Pedhazur and Tetenbaum (1979) point out, that these two items are responsible in large part for sex differences in self-report on the BSRI, but the goal of the BSRI is to measure and to facilitate the investigation of within-sex differences, not between-sex differences. The difference between females and males on the scales of the BSRI has never been used either as a criterion for item selection or as an indicator of validity (unlike the sex difference on Spence and Helmreich's Personal Attributes Questionnaire, 1978), but only as an empirical datum about a particular subject sample at a particular point in time. The theory behind the BSRI in no way commits itself even to the existence of a sex difference in any particular sample of women and men. Once again, the focus is on individual differences, not sex differences. Moreover, as factor analyses by Pedhazur and Tetenbaum and others (e.g., Berzins, Welling, & Wetter, 1978; Feather, 1978; Gaudreau, 1977; Waters, Waters, & Pincus, 1977) have indicated, these two items do not load highly with other items on the BSRI; rather, they form a bipolar factor of their own, a factor correlated

---

[1] It should be noted that an earlier study, also conducted at the University of Washington, had previously replicated this pattern for only 2 of the 40 attributes—"feminine" and "masculine"—a finding that was interpreted by the authors as indicating that the original social desirability ratings were not stable (Edwards & Ashworth, 1977). However, that study did not utilize our original instructions, but asked judges instead to rate "how desirable or undesirable you judge [each attribute] to be in an American male/female." In contrast, our instructions emphasized that we were only interested in each judge's assessment of how American society would evaluate the various attributes. The two kinds of ratings are simply not equivalent—in 1972 or 1977. The Edwards and Ashworth study thus constitutes a failure to replicate our methodology, not a failure to replicate our results.

in several analyses with gender. Finally, these two items are not highly correlated with their own total scale scores, but serve primarily as gender markers.

It is important in this context not to confuse the individual items "feminine" and "masculine" with the Femininity and Masculinity scales. The labels for the scales are a shorthand means of summarizing that all of the items were selected in the first place because they were judged more desirable either for women or for men.

Factor Analysis and the Short BSRI

As Pedhazur and Tetenbaum (1979) suggest, factor analyses of the feminine and masculine items on the BSRI generally yield four factors that account for 75% or more of the common variance: (a) a single feminine factor defined by such items as "warm," "gentle," and "eager to soothe hurt feelings"; (b) two masculine factors, one defined by such items as "dominant," "aggressive," and "assertive," and the other defined by such items as "independent," "self-reliant," and "self-sufficient"; and (c) a factor correlated in several analyses with gender and defined by the items "masculine" and "feminine" (and occasionally "athletic"). According to Pedhazur and Tetenbaum, these results are devastating to the BSRI. I disagree. Because the theory underlying the BSRI does not require that the domains of femininity and masculinity be unidimensional, it is only the existence of that small fourth factor that is unanticipated by the theory.

The results of these factor analyses do suggest ways in which the BSRI might be refined, however; and in fact, a short BSRI has recently been developed that contains exactly half of the original items. Two groups of feminine and masculine items in particular were eliminated during the development of the short BSRI: (a) The few items, including "feminine" and "masculine," that defined the factor correlated with gender, as noted above, and (b) a group of feminine items with relatively low social desirability, none of which correlated highly with the Femininity score or loaded on the feminine factor for either sex, and a few of which even had high loadings on the masculine factor (e.g., "yielding," "shy," and "soft-spoken"). These feminine items were included on the original BSRI in order to balance the overall social desirability of the feminine and masculine attributes, but as the concept of androgyny has evolved, it has seemed increasingly inappropriate to define androgyny in terms of these relatively undesirable attributes. Accordingly, the Femininity and Masculinity scales of the short BSRI consist of items that represent the most desirable personality characteristics for a given sex, and the variances of their social desirability ratings are quite comparable as well. Finally, the short BSRI also includes 10 filler items that were judged to be no more desirable for one sex than for the other in all four of our samples of judges as well as in the Walkup and Abbott (1978) replication.[2]

The Locksley and Colten Critique

In contrast to the critique by Pedhazur and Tetenbaum (1979), the Locksley and Colten (1979) critique raises strategic issues of a broader and more conceptual nature, issues that turn on the eventual heuristic payoff of differing approaches. For example, Locksley and Colten question the feasibility of basing the measurement of individual differences in femininity and masculinity on broadly based cultural stereotypes about women and men, and they believe that a more cognitive approach to sex typing would be more fruitful. Again, only future research can decide such issues. They also question the concept of androgyny itself: "Sex is an immediately perceptible feature of every person. . . . Considering the role of sex in the very architecture of experience and behavior, the notion of psychological androgyny, with its implication of freedom from sex-related social effects on personality and behavior, is arbitrary at best" (pp. 1028–1029).

But like most psychological concepts, the concepts of sex typing and androgyny are seen as matters of degree. I, too, would agree with

[2] A detailed manual for both the original and the short BSRI will soon be available from Consulting Psychologists Press, 577 College Avenue, Palo Alto, California 94306.

the rather unexceptional position that it is not possible for an individual to be completely free from sex-related social effects, but that does not preclude the possibility that individuals may differ in the extent to which gender serves as a cognitive schema for the processing of information, a lens through which they perceive and interpret social reality. Moreover, my current research on the cognitive processes mediating sex typing and androgyny is addressed to precisely this hypothesis.

Gender as a Cognitive Schema

The distinction between male and female clearly exists "out there" in the real world as a basic and fairly primitive dichotomy. Moreover, it is a dichotomy that is important to almost all human cultures in a way that extends well beyond basic biological differences in body build and reproductive function; and, precisely because it does loom so large, it is a dichotomy that cannot be overlooked or treated as psychologically irrelevant by anyone. Indeed, the distinction between male and female is known by children as young as 1 or 2 years of age, and in most cultures, children are explicitly taught to treat the gender dichotomy with seriousness and respect from that time forward. Moreover, within most cultures, a variety of different sources can aid and abet an individual's awareness of the gender dichotomy: for example, interacting exclusively with same-sex peers and inferring that the other sex must be fundamentally different from you; observing mothers and fathers in sex-stereotyped roles as homemaker and breadwinner and inferring that the two sexes must be fundamentally different from one another; observing that mixed-sex relationships always seem to have sexual overtones and inferring that heterosexual attraction constitutes the only experience shared in common by males and females; observing that different pronouns are used in reference to the two sexes and inferring that the distinction between them must be even more important than, say, the distinction between black and white or between young and old; and so forth.

But despite the universal recognition of the gender dichotomy, it is a basic assumption underlying my current research that there are wide individual differences in the functional importance attached to it. In particular, I hypothesize that sex-typed and androgynous individuals differ from one another in how much they believe the sexes to be basically different from one another—a belief in "gender polarity"—and that the differences between these sex-role groups both in self-description and in behavior are themselves a consequence of the differences in the content of their beliefs about the two sexes. Moreover, it is suggested that not only do individuals of different sex roles differ in the extent to which they hold different beliefs and expectations about what the two sexes are like but, furthermore, these beliefs mediate both how they, as individual males and females, behave and how they interpret the behavior of male and female others as well.

At the more basic level, it is hypothesized that individuals of different sex roles differ not only in the content of their beliefs about gender differences (gender polarity) but in their cognitive structures for coding and processing information as well, structures that have variously been called frames (Minsky, 1975), scripts (Abelson, 1976), and schemata (Bartlett, 1932; Bobrow & Norman, 1975; Kelley, 1972; Markus, 1977; Stotland & Canon, 1972). The notion of a schema connotes a highly articulated and dynamic concept that organizes and guides one's processing of information by virtue of its rich associative network, its implications for cause-and-effect relationships, and its well-developed criteria for making discriminations along the relevant dimensions, implying that individuals of different sex roles should differ in the cognitive processing of gender-related information. In particular, to the extent that an individual holds different beliefs about what the two sexes are like, information related to gender and to gender differences should be more perceptually salient and more cognitively available (Nisbett & Ross, in press; Ross, 1977; Tversky & Kahneman, 1973, 1974). That is, gender differences should be more readily perceived or noted; they should be more readily stored in and retrieved from

memory; they should be more readily used as the basis of personality attributions and predictive inferences; they should be more readily put forth as a causal candidate for a variety of behavioral outcomes. In short, the gender dimension should be more cognitively available for the processing of information. Accordingly, individuals of different sex roles are seen as differing in how much they spontaneously process information about the self, about others, and about the nonsocial environment in general in gender-related terms.

In sum, then, individuals of different sex roles are not viewed here as differing primarily in terms of how much masculinity or femininity they possess, but rather, they are viewed as differing more fundamentally (a) in the content of their beliefs about what the two sexes are like and (b) in their cognitive schemata for processing gender-related information, and hence in the perceptual salience and cognitive availability of gender and gender-related concepts as dimensions for processing incoming information.

It should be noted that two studies already in the literature provide preliminary support for the hypothesis that sex-typed individuals differentiate along a gender-related dimension significantly more than androgynous individuals do. In the first study, sex-typed and androgynous subjects watched either a female or a male on videotape performing a series of neutral activities, and, using a hand-held recorder, they segmented the videotaped sequence into self-defined units that seemed natural and meaningful to them. The results revealed a significant Sex Role × Sex of Actor interaction, with sex-typed subjects differentiating between the female and male actor significantly more than androgynous subjects did (Deaux & Major, 1977). In the second study, sex-typed and androgynous subjects rated the similarity of handwritings on masculinity–femininity, and they also rated the handwritings on an absolute scale of masculinity–femininity. The results indicated both that sex-typed subjects differentiated along the dimension of masculinity–femininity significantly more than androgynous subjects did and that they weighted it more heavily in making similarity judgments (Lippa, 1977).

Androgyny as a Transitional Concept

Locksley and Colten (1979) are rightly concerned with the conceptual status of the androgyny concept. As I have noted elsewhere, I have a discomfort of my own:

> If there is a moral to the concept of psychological androgyny, it is that behavior should have no gender. But there is an irony here, for the concept of androgyny contains an inner contradiction and hence the seeds of its own destruction. Thus, as the etymology of the word implies, the concept of androgyny necessarily presupposes that the concepts of femininity and masculinity themselves have distinct and substantive content. But to the extent that the androgynous message is absorbed by the culture, the concepts of femininity and masculinity will cease to have such content and the distinctions to which they refer will blur into invisibility. Thus, when androgyny becomes a reality, the *concept* of androgyny will have been transcended. (S. L. Bem, in press)

References

Abelson, R. P. Script processing in attitude formation and decision making. In J. S. Carroll & J. W. Payne (Eds.), *Cognition and social behavior.* Hillsdale, N.J.: Erlbaum, 1976.

Anastasi, A. *Psychological testing.* New York: Macmillan, 1968.

Bartlett, F. C. *Remembering: A study in experimental and social psychology.* London: Cambridge University Press, 1932.

Bem, D. J. Constructing cross-situational consistencies in behavior: Some thoughts on Alker's critique of Mischel. *Journal of Personality,* 1972, *40,* 17–26.

Bem, D. J., & Allen, A. On predicting some of the people some of the time: The search for cross-situational consistencies in behavior. *Psychological Review,* 1974, *81,* 506–520.

Bem, S. L. The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology,* 1974, *42,* 155–162.

Bem, S. L. Sex-role adaptability: One consequence of psychological androgyny. *Journal of Personality and Social Psychology,* 1975, *31,* 634–643.

Bem, S. L. On the utility of alternative procedures for assessing psychological androgyny. *Journal of Consulting and Clinical Psychology,* 1977, *45,* 196–205.

Bem, S. L. Beyond androgyny: Some presumptuous prescriptions for a liberated sexual identity. In J. Sherman & F. Denmark (Eds.), *The future of women: Issues in psychology.* New York: Psychological Dimensions, in press.

Bem, S. L., & Lenney, E. Sex typing and the avoidance of cross-sex behavior. *Journal of Personality and Social Psychology,* 1976, *33,* 48–54.

Bem, S. L., Martyna, W., & Watson, C. Sex typing and androgyny: Further explorations of the ex-

pressive domain. *Journal of Personality and Social Psychology,* 1976, *34,* 1016–1023.

Berzins, J. I., Welling, M. A., & Wetter, R. E. A new measure of psychological androgyny based on the Personality Research Form. *Journal of Consulting and Clinical Psychology,* 1978, *46,* 126–138.

Bobrow, D. G., & Norman, D. A. Some principles of memory schemata. In D. G. Bobrow & A. Collins (Eds.), *Representation and understanding: Studies in cognitive science.* New York: Academic Press, 1975.

Campus, N. Transituational consistency as a dimension of personality. *Journal of Personality and Social Psychology,* 1974, *29,* 593–600.

Deaux, K., & Major, B. Sex-related patterns in the unit of perception. *Personality and Social Psychology Bulletin,* 1977, *3,* 297–300.

Edwards, A. L., & Ashworth, C. D. A replication study of item selection for the Bem Sex-Role Inventory. *Applied Psychological Measurement,* 1977, *1,* 501–507.

Feather, N. T. Factor structure of the Bem Sex Role Inventory: Implications for the study of masculinity, femininity, and androgyny. *Australian Journal of Psychology,* 1978, *30,* 241–254.

Gaudreau, P. Factor analysis of the Bem Sex Role Inventory. *Journal of Consulting and Clinical Psychology,* 1977, *45,* 299–302.

Ickes, W., & Barnes, R. D. Boys and girls together—and alienated: On enacting stereotyped sex roles in mixed-sex dyads. *Journal of Personality and Social Psychology,* 1978, *36,* 669–683.

Kelley, H. H. Causal schemata and the attribution process. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior.* New York: General Learning Press, 1972.

Lippa, R. Androgyny, sex typing, and the perception of masculinity–femininity in handwriting. *Journal of Research in Personality,* 1977, *11,* 21–37.

Locksley, A., & Colten, M. E. Psychological androgyny: A case of mistaken identity? *Journal of Personality and Social Psychology,* 1979, *37,* 1017–1031.

Markus, H. Self-schemata and processing information about the self. *Journal of Personality and Social Psychology,* 1977, *35,* 63–78.

Minsky, M. A framework for representing knowledge. In P. Winston (Ed.), *The psychology of computer vision.* New York: McGraw-Hill, 1975.

Nisbett, R. E., & Ross, L. *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, N.J.: Prentice-Hall, in press.

Pedhazur, E. J., & Tetenbaum, T. J. Bem Sex Role Inventory: A theoretical and methodological critique. *Journal of Personality and Social Psychology,* 1979, *37,* 996–1016.

Ross, L. The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10). New York: Academic Press, 1977.

Russell, G. The father role and its relation to masculinity, femininity, and androgyny. *Child Development,* 1978, *49,* 1174–1181.

Spence, J. T., & Helmreich, R. L. *Masculinity and femininity: Their psychological dimensions, correlates, and antecedents.* Austin: University of Texas Press, 1978.

Stotland, E., & Canon, L. K. *Social psychology: A cognitive approach.* Philadelphia, Pa.: Saunders, 1972.

Tversky, A., & Kahneman, D. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology,* 1973, *5,* 207–232.

Tversky, A., & Kahneman, D. Judgment under uncertainty: Heuristics and biases. *Science,* 1974, *185,* 1124–1131.

Walkup, H., & Abbott, R. D. Cross-validation of item selection on the Bem Sex Role Inventory. *Applied Psychological Measurement,* 1978, *2,* 63–71.

Waters, C. W., Waters, L. K., & Pincus, S. Factor analysis of masculine and feminine sex-typed items from the Bem Sex Role Inventory. *Psychological Reports,* 1977, *40,* 567–570.

Received October 2, 1978

Journal of Personality and Social Psychology
1979, Vol. 37, No. 6, 1055-1074

How Relevant is a Semantic Similarity Interpretation of Personality Ratings?

Jack Block, Daniel S. Weiss, and Avril Thorne
University of California, Berkeley

In recent years, a conceptual similarity interpretation of personality and interpersonal assessment ratings has been advanced by D'Andrade and Shweder. According to their view, such ratings are primarily understandable as linguistic artifacts having little or no connection with the real world. This position contravenes a central and widely used methodology of personality and behavior assessment. The present article critically evaluates the logic and data that have been used to support this conceptual similarity explanation and concludes that pertinent evidence is wanting. A primary finding of D'Andrade and Shweder, that judges can predict the intercorrelations among personality ratings, is shown in an experimental study to be an adventitious function of previously unrecognized patterns of redundancy in the personality variables that happen to have been used. Finally, some empirical relationships not assimilable to their position are noted together with some conceptual problems of their viewpoint.

A decade ago, D'Andrade (1965, 1974) began to advance a radical and sweeping critique of two essential techniques of personality assessment: personality ratings and personality questionnaires. Applying his own analytical logic to several sets of previously published data, D'Andrade concluded that "the correlations [among rating and inventory scales] are primarily an artifact of the rater's or the questionnaire taker's cognitive structure, and not a reflection of the real world" (1974, p. 181).

More recently, Shweder (1975, 1977a, 1977b) has extended and elaborated on D'Andrade's theme, asking the confronting question: "How relevant is an individual difference theory of personality?" (1975, p. 455). By "an individual difference theory of personality," Shweder referred to generally held conceptions of personality as epitomized, for example, by Child's (1968) definition: "More or less stable internal factors that make one person's behavior consistent from one time to another, and different from the behavior other people would manifest in comparable situations" (p. 83). Shweder answered his question negatively, boldly asserting that "an individual difference theory of personality" has been "shown" by him to be "no more than statements about how respondents [and psychologists] classify things as alike in meaning" (1975, p. 482). As Shweder noted, much of the evidence for the existence of personality parameters that shape behavior in coherent ways has depended on observation-based ratings and personality inventories. According to Shweder, when rating and inventory data are generated by psychologist-observers or by inventory takers, respondents (including psychologists)

> unwittingly substitute a theory of conceptual likeness for descriptions of behavioral co-occurrences. . . . Items alike in concept are inferred to be behaviorally characteristic of the same person even when, as is typically the case, conceptual relationships among items do not correspond to the actual behavioral relationships among items. . . . [As a result] these conceptually biased judgments create an "illusion" of underlying behavioral consistency which, although not apparent in actual behavior, deceptively validates the "individual difference" conceptualization of "personality." (1975, pp. 455-456)

This point of view, which strikes at the

This study was supported by National Institute of Mental Health Grant 5R01 MH 16080 to Jack Block.

Requests for reprints should be sent to Jack Block, Department of Psychology, University of California, Berkeley, California 94720.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3706-1055$00.75

very core of personality and behavioral assessment, has recently achieved appreciable currency, being approvingly cited in reviews by Schneider (1973) and Mischel (1968, 1969, 1973, 1977), among others. The present article is an effort to evaluate the premises, logic, and empiricism that underlie the strong conclusions of what we shall call the *conceptual similarity* position. We focus primarily on the "relevance" of personality ratings, a quite sufficient context in which to address the issues that have been raised.

The Conceptual Similarity Argument

Underlying the conceptual similarity position are several propositions and presumptions about personality assessment that require specification and scrutiny if the tenability of this conclusion is to be evaluated.

The Conceptual Similarity View of Personality and Behavior Assessment Research: Emphasis on Trait Similarity Structures

D'Andrade characterizes personality assessment as follows:

> First, one or more human observers are asked to judge one or more subjects on a number of traits of behavior; the judgments are expressed in ratings or rankings (scores) based on the observers' long-term (i.e., more than 10 minutes) memory of the subjects' behavior. Second, a single score is computed for each subject for each trait, usually by taking the mean of all the scores given to each subject for each trait. Third, to find out how the traits are related to each other, some measure of association, such as the product-moment correlation, is computed from the subjects' scores for all pairs of traits. Finally, the measures of association are analyzed to determine how the traits are organized with respect to each other . . . . The results of such analyses indicate which traits tend to go together, and the similarity structure of the trait measurements is taken as a representation of the structure or organization of the subjects' behavior. (1974, p. 161)

This impression of personality and behavioral assessment research will appear foreign to most assessment psychologists. Insofar as ratings are used as one of the primary methods of assessment, the first two steps listed by D'Andrade often apply. The rating data thus generated, however, are then generally used in a multipronged effort (together with other kinds of measures) to establish the construct validity of a concept. The goal of such effort is a dependable and coherent nomological network of relationships on which to base subsequent theorizing or prediction. The emphasis of personality research is, centrally, on the validity of the measures being used; that is, Do individuals earning different scores on an index also earn conceptually expectable different scores on entirely separate and independently formulated measures? There is a long history to this kind of effort; recent reviews of the status, accomplishments, and deficiencies of personality research are to be found in Block (1977a), Gough (1976), Hogan, DeSoto, and Solano (1977), and Stagner (1977).

The third and last steps listed in the D'Andrade (1974) conceptualization of personality research may, and often do, arise when multiple ratings or measures have been gathered and there is question or concern about the extent of redundancy of coverage of the personality or behavioral domain by the variables used. Although a certain amount of redundancy is required to ensure reliability, redundancy beyond this certain amount is wasteful. Various clustering methods, such as factor analysis, can be extremely useful in quickly and simply revealing the latent dimensionality of the set of ratings employed and as a guide to identifying rating variables that are dispensable with little, or at least acceptable, cost.

Although the results of factor analyses (and related methods) of sets of ratings are often of immediate and intrinsic interest, there are few personality psychologists today who would claim that obtained factor structures represent inexorable verities (cf. Lykken, 1971). It is now well recognized that by varying the mix of variables included in a clustering analysis, one can fundamentally alter both the number and nature of the summarizing dimensions subsequently obtained. Thus, D'Andrade's claim that "the similarity structure of the trait measurements is taken [by psychologists] as a representation of the structure or organization of the subjects' behavior" misunderstands both the analytical methods used and the views of most personality psychologists. Personality psychologists

seek or settle on useful sets of variables to be employed for the specification of the personal or behavioral qualities of the individuals to be studied. To the extent that the set of personality or behavioral variables has communal variance (i.e., is redundant), there will be a "similarity structure of the trait measurements." But such similarity structures represent neither conceptual nor empirical fixities. We believe that the focus on arbitrary and often adventitious "similarity structures" derived from ratings addresses tangential or subsidiary issues. Instead, we suggest that the essential concern of personality assessment—a concern not addressed by D'Andrade and Shweder—is the convergent and discriminant validity of personality and behavioral ratings. Moreover, we note that the validity of a personality measure is not necessarily tied to the interrelationships among a set of personality measures. We shall be returning to this point.

The Primary Demonstration of the Conceptual Similarity Position: Correspondence Between Rated Behavior and Conceptual Similarity Matrices

The personality or behavioral variables rated by observers or by peers to describe a set of subjects can be intercorrelated. We call the consequent matrix a Rated Behavior matrix. Separately, the variables that had been rated with respect to actual individuals can be scaled by judges for the degree of conceptual similarity or meaning equivalence each variable has with respect to every other variable. The resulting matrix of similarity indices we call a Conceptual Similarity matrix. The primary demonstration of the conceptual similarity position has been to show that for a number of previously published Rated Behavior matrices, the pattern of intercorrelations characterizing a Rated Behavior matrix could be well approximated by the pattern of intercorrelations characterizing the Conceptual Similarity matrix based on the same set of variables. Because of the frequently strong correspondence between these two matrices, in which the Rated Behavior matrix stems from observations of real individuals, whereas the Conceptual Similarity matrix derives from purely semantic judgments of the similarity of personality variables made without reference to palpable people, D'Andrade concluded that "with respect to personality and behavior specifications in the field of psychology . . . it is possible to confuse propositions about the world with propositions about language" (1965, p. 215). Shweder has been more vigorous in his assertions, claiming that

> Evidence generally thought to demonstrate the applicability of the "individual difference" conceptualization of "personality" [has been] shown to be equivocal by deriving the trait-factors discovered in interpersonal rating forms, questionnaire interviews and personality inventories from purely conceptual criteria without any reference to actual behavior. (1975, p. 459)

We suggest that the correspondence between a Rated Behavior matrix and a Conceptual Similarity matrix, when it exists, does not have the logical implication with respect to individual differences that the conceptual similarity position seems to claim. The reader should pause for a moment to verify that in no way whatsoever is it possible to proceed from the specific semantic similarity judgments on which the Conceptual Similarity matrix is based to a specification of just which individuals are rated high or low on a particular rating dimension. For example, to know that judges evaluate the personality rating variables "Introverted" and "Devaluates himself" as conceptually highly similar (Shweder, 1975, Table 1) is to be entirely uninformed about who is rated Introverted or as Devaluat(ing) himself or whether Introverted and Devaluates himself, as separate or conjoined variables, are important and useful characteristics in terms of which to classify people. We shall return later to this distinction between *classifying personality variables* as opposed to *classifying individuals in terms of personality variables.*

The Secondary Demonstration of the Conceptual Similarity Position: Absence of Correspondence Between Rated Behavior and Actual Behavior Matrices

From the beginning, D'Andrade recognized the confounded nature of the primary demonstration:

> It is possible that the so-called psychological traits

dealt with in this paper exist both as components in terms used to describe the external world and in the external world as well. Such an isomorphism, if it exists, might be the result of the external world affecting first the discriminations made by speakers of a language, who then eventually develop a semantic structure within the language to encode these discriminations. (1965, pp. 227–228)

Thus, the isomorphism hypothesis that "the semantic similarity of trait terms corresponds to the way these traits actually go together" (D'Andrade, 1974, p. 162) was confounded with the more dramatic systematic-distortion hypothesis "that traits the observer considers similar will be recalled as applying to the same person, even when this is not the case" (D'Andrade, 1974, p. 161).

To speak to the issue of confounding, D'Andrade (and Shweder) sought an additional demonstration that would distinguish between the two competing hypotheses. They reasoned that if the systematic distortion or conceptual similarity hypothesis was correct, a matrix of intercorrelations among what they called "actual behaviors" (which we will call the Actual Behavior matrix) should not correspond to the Rated Behavior or Conceptual Similarity matrices.

If the observer's memory-based ratings showed a very different pattern of correlations from that found for the data based on the actual behavior of the subjects (but a pattern similar to judgments of semantic similarity), it would be reasonable to reject the isomorphic hypothesis and to consider the systematic-distortion hypothesis supported. (D'Andrade, 1974, p. 162)

D'Andrade (1974) has reported on two studies in the existing psychological literature that he deemed to have Actual Behavior and Rated Behavior matrices suitable for unconfounding purposes; Shweder (1975) has reported a third study relating Actual Behavior and Rated Behavior patterns of intercorrelations to each other and to a Conceptual Similarity pattern. In all three sets of analyses, a strong conclusion is reached affirming the systematic-distortion or conceptual similarity interpretation:

Rated behavior is almost entirely under the influence of pre-existing conceptual schemes and corresponds to actual behavior relationships only to the extent pre-existing conceptual schemes happen to partially coincide with actual behavior. (Shweder, 1975, p. 462)

Later in this article, we shall suggest a number of reasons why the three studies located by D'Andrade and Shweder fail to meet their special hypothesis-testing requirements.

The Implications of the Conceptual Similarity Position

The critical import of the conceptual similarity argument is powerful, blunt, and devastating to personality psychology and other behavioral science fields relying on some form of ratings.

If the correlations on which these studies rest are primarily an artifact of the rater's or the questionnaire taker's cognitive structure, and not a reflection of the real world, there is little or no evidence that human behavior can be described by large multibehavior units. What remains is a world in which human behavior is to be described in terms of specific behaviors occurring in specific situations, as Mischel and others have argued. (D'Andrade, 1974, p. 181)

D'Andrade went on to suggest that "it is not necessary to group behavior into clusters, or traits, or dimensions to be able to give an economical description of an individual's behaviors" (1974, p. 184). Instead, he claimed that "a remarkably predictive description can be obtained" (p. 185) if one simply predicts the most frequent behavior previously observed in each of a number of different behavior settings. No data are offered with regard to the predictive efficacy of this behavior-specific approach. Whatever the predictive accuracy of this actuarial orientation may prove to be, it is clear that it abandons the elegance and economy of a theory of behavior. Many psychologists will not be interested in a psychology reduced to, and aspiring to no more than, an endless graphing of frequency distributions for behavior-category systems. Were the logic and empiricism surrounding the behavior-specific approach compelling, this reductionistic but assumption-free approach to "good prediction" would prevail. It is our contention, however, that the argument and evidence brought together to buttress the conceptual similarity interpretation of personality ratings as existing solely "in the eyes of the beholder" are

deeply flawed, wrongly focused, and greatly embarrassed by a host of empirical findings unassimilable to the systematic-distortion position.

The remainder of this article will more closely consider the evidence and reasoning underlying the conceptual similarity interpretation of the processes underlying personality assessment. We shall first evaluate the Actual Behavior – Rated Behavior matrix analyses intended to unconfound the isomorphic and systematic distortion hypotheses. Next, we go on to illustrate the conditions that are sufficient and perhaps necessary to create or to prevent correspondence between Rated Behavior and Conceptual Similarity matrices. Finally, we list a variety of connections between personality ratings and indisputably real world outcomes that cannot be understood in artifactual terms. In the course of our countering evaluation, we shall burden the reader with matters of analytic detail, psychometrics, and principles of inference. We do not apologize for the closeness of our analysis; too often in psychology, crucial experimental points and analytical considerations go unnoticed. If the reader is to form well-based opinions on the important issues involved, he or she will have to become actively involved in the specifics of the evaluative process.

Unconfounding the Isomorphic and Systematic-Distortion Hypotheses: Some Problems in the Analyses

To harden the case for a conceptual similarity interpretation of personality ratings, demonstrations were sought that

> Verbal data collecting procedures, such as interpersonal ratings, self-report inventories, and retrospective questionnaires, . . . cannot be systematically supported by detailed and immediately recorded observational data. (Shweder, 1975, pp. 459–460)

It was reasoned that

> Pattern comparison [of an Actual Behavior matrix with Rated Behavior matrices] is needed here because *it is not the validity of specific traits or categories which is in question* [emphasis added], but the validity of the correlations between traits. Ideally, for recording the actual behavior of the subjects, a mechanical device should be constructed to count frequencies of different kinds of behavior. Unfortunately . . . a mechanical measuring instrument is not practical at present for any judgment more complex than *is making noise* versus *is not making noise*. The closest approximation to a mechanical device appears to be a trained observer using a simple coding scheme to record a subject's behavior as it occurs. The immediacy of the observer's assessment and the simplicity of the coding decisions should, it is hoped, protect against systematic distortion of the type thought to take place in long-term (i.e., more than ten minutes) memory. (D'Andrade, 1974, p. 162)

The two studies reworked by D'Andrade for his Actual Behavior – Rated Behavior matrix comparisons are an investigation by Borgatta, Cottrell, and Mann (1958) and the doctoral dissertation of Mann (1959). Shweder has added a comparable analysis of a study by Newcomb (1929). It will be helpful to briefly characterize these studies.

The Borgatta et al. (1958) research used 47 graduates divided into five small groups. Each group met for ten 2-hour sessions to discuss "democratic leadership." During the ninth and tenth sessions, a *single* observer using the Interaction Process Analysis (IPA) category system (Bales, 1951) coded the ongoing interaction. Subsequently, for each member of the group, the rate of each Bales category was calculated. These rates served as the Actual Behavior scores for subsequent analysis. After the ninth session, each group member ranked all the members of his group with respect to 40 traits, 6 of which *approximated* the Bales categories. Subsequently, for each member, the average ranking (adjusted for the slight differences in group sizes) on each of the 6 traits approximating the Bales categories was calculated. These average rankings served as the Rated Behavior scores for subsequent analysis.

The Mann (1959) study used 100 fraternity undergraduates, each of whom was assigned to two different and nonoverlapping five-man groups. In one of these 50-minute groups, the group worked on a specific task; in the other kind of group, the group members attempted to formulate some fraternity policies. During each session, a *single* observer used a version of the Bales interaction categories to code the ongoing interaction. For each member of the group, the total

number of "acts" was calculated. For each of seven Bales IPA categories, the subject's act frequency was divided by his total number of acts to derive a percentage. These percentages served as the Actual Behavior scores for subsequent analysis. After a group session, each member rated the other four on 16 characteristics, 7 of which *approximated* the Bales-Mann interaction categories. Subsequently, for each subject, the sum of the ratings assigned by the other four members of his group on each of the characteristics was calculated. These pooled ratings served as one form of Rated Behavior scores for subsequent analysis. In addition, the interaction-observing coder at the end of each session was asked to rate directly for each subject the percentage of the subject's behavior in each of the Bales-Mann categories. These estimated percentages served as a second form of Rated Behavior scores in subsequent analysis.

The Newcomb (1929) study evaluated 51 problem boys (a group of 27 preadolescents and a group of 24 adolescents) attending a month-long summer camp. For each group, camp consisted of six tentfuls of boys, each tent supervised by a different counselor. Sometime during each day, for each boy, his tent counselor completed a behavior record with respect to 26 behavioral questions Newcomb asked because of their presumed relevance to extraversion–introversion. For each question, the counselor selected one of four alternatives provided by Newcomb, precoded as "positive" (i.e., implying extraversion) or "negative" (i.e., implying introversion). In subsequent tabulation for each boy, for each of the behavior questions the number of days on which a positive or a negative alternative had been recorded was divided by the total number of camp days for which a report had been submitted. These 26 proportions, from *single* and *uncalibrated* observers, served as the Actual Behavior scores for subsequent analysis. At the end of the camp period, a rating form on *approximately* the same 26 behavior situations was completed for each boy by his tent counselor and by five other counselors who had frequently observed each boy. The six ratings for each boy were then combined into one consensus rating. The 26 pooled ratings for each boy served as the Rated Behavior scores for subsequent analysis.

D'Andrade (1974) and Shweder (1975) suggest that the three studies just described permit them to "explicitly separate . . . the structure of 'social impressions' from the structure of behavioral instances" (Shweder, 1975, p. 479). We suggest, alternatively, that the reasoning and evidence adduced by D'Andrade and Shweder as support for this "separation" lack power and pertinence. Some of our concerns follow.

The categories or items or dimensions of behavior measured by the Actual Behavior indices are often different from those rated by observers. An analysis seeking to interpret the differences between the similarity structures derived from an Actual Behavior and a Rated Behavior matrix in conceptual similarity terms requires that the sets of variables being contrasted must, of course, be identical. The only data difference should be that in the Actual Behavior case, the variables are scored from "actual behavior," whereas in the Rated Behavior case, the variables are scored on the basis of retrospective judgments and inferences.

However, in none of the three studies used to demonstrate the difference between Actual Behavior and Rated Behavior matrices does the required equivalence obtain between the Actual Behavior and Rated Behavior sets of variables. Consider the following examples of failure of definitional comparability.

In the Borgatta et al. (1958) study, the Actual Behavior variable, IPA Category 1, is defined as "Shows solidarity, raises others' status, jokes, gives help, reward." The corresponding Rated Behavior variable is defined as "Shows solidarity and friendliness." Such facets of interaction as raising others' status, joking, and giving help and reward, included in the Actual Behavior version of this variable, are not included in the Rated Behavior version. The Actual Behavior variable, IPA Category 2, is defined as "Shows tension release, shows satisfaction, laughs." The corresponding Rated Behavior variable is defined as "Is responsive to laughter." Such facets of interaction as showing tension release and showing satisfaction, included in

the Actual Behavior version, are omitted from the Rated Behavior version. The Actual Behavior variable, IPA Category 4, is defined as "Gives suggestion, direction, implying autonomy for other." The corresponding Rated Behavior variable is defined as "Makes the most suggestions." The important qualifier, "implying autonomy for other," present in the Actual Behavior version, is absent from the Rated Behavior version. The remaining Actual Behavior – Rated Behavior correspondences pose similar problems of comparability.

In the Mann (1959) study, various Bales categories were merged. The first Actual Behavior variable, Mann's Category 1, subsumes showing solidarity (excluding all joking behavior) and showing agreement. The Rated Behavior version of this variable inquired "how much [the group member being rated] tended to agree with what others had said." No mention of solidarity is contained in the Rated Behavior variable definition, a most important omission.

Shweder's (1975) report of the Newcomb (1929) study claims that the same 26 items of behavior were evaluated by the daily records and by the end-of-camp ratings. In fact, 15 of the 26 items lack Actual Behavior – Rated Behavior identity. For example, the daily record item, "Was he fond of swimming?" becomes the rating item, "Did he spend his swimming period periods in the water, actively moving about?" The daily record item, "How much of the day did he spend doing things that required little or no action?" becomes the rating item, "Was he actively moving about most of the day?" Sometimes these differences seem slight but, as is well known, slight differences in phrasing often can have great effect.

The absence of exact equivalence between the definitions of the Actual Behavior and the Rated Behavior variable sets is regrettable; this consideration alone renders problematic the significance of subsequent Actual Behavior and Rated Behavior matrix comparisons that D'Andrade and Shweder have interpreted as evidence for the essential invalidity of "rated behaviors."

Emphasis solely on the correlations between traits ignores consideration of the validity of specific traits. We believe that most psychologists will indeed wish to know the specific validity of traits or categories—data deemed irrelevant by D'Andrade and Shweder. A moment of reflection will indicate to the reader that if the traits or categories being evaluated are validly measured, then the correlations between the traits or categories must be valid. However, the converse of this relationship does not apply; correlations between traits can exist and even be valid even though each trait is entirely invalid as a basis for ordering individuals.[1] Therefore, it is of interest to compare how similarly subjects are ordered by "rated behavior" scores and by "actual behavior" scores posited as the criterion for validity. In the Borgatta et al. (1958) study, the mean correlation between corresponding Rated Behavior and Actual Behavior variables is .25, with a range from .14 through .38. In the Mann (1959) study, the mean correlation of corresponding Rated Behavior and Actual Behavior variables is, for the Rated Behavior variables based on group member ratings, .33, with a range from .10 to .55. For the Rated Behavior variables based on ratings by the Actual Behavior coder, the mean Rated Behavior – Actual Behavior correlation is .41, with a range from .18 to .65. In the Newcomb (1929) study, the mean Rated Behavior – Actual Behavior correlation was .49, with a range from .16 to .73, for one group of boys, and .43, with a range from .02 to .68, for the second group of boys.

All of these figures warrant correction for attenuation due to the unreliability of the Rated Behavior and Actual Behavior measures (Block, 1963, 1964). However, unreliability of measurement is not noted as a factor operating to diminish the correspondence between Rated Behavior scores and Ac-

[1] As a simple physical illustration, consider a measure of height and a measure of weight and the consequent correlation between height and weight. If, for the sample of subjects involved, the paired scores for height and weight were to be randomly reassigned, the correlation between height and weight would be unchanged, but the validity of height and the validity of weight can be expected to be zero.

tual Behavior scores. Because the three studies provide no (or technically unsound) reliability coefficients, we cannot estimate what the relationships between the Rated Behavior and Actual Behavior variables would be if they were psychometrically improved. For many of the pairings, however, we can expect the correlation to increase appreciably.[2] Nevertheless, even as they stand, and recognizing the definitional discrepancies already noted, many of the Rated Behavior – Actual Behavior correlations cannot be considered low, especially when some of the additional attenuating factors to be introduced below are recognized.

The use of frequency counts of behavior is not a sufficient means of operationalizing complex psychological concepts. Conceptual similarity advocates recommend, "for recording the actual behavior of the subjects, . . . count frequencies of different kinds of behavior . . . using a simple coding scheme" (D'Andrade, 1974, p. 162). The several assumptions underlying this epistemological position are bothersome on several grounds.

1. D'Andrade and Shweder write as if "actual behavior" exists; it does not. Behavior in the raw has an infinite variety of facets and pattern possibilities and therefore can only be studied by selection. The act of selection, however done, represents a constructive and theoretical assertion about the world that wins its justification by the (ultimately personal) nomological network with which it subsequently can be surrounded. Kaplan (1964) has written tellingly regarding the philosophical issues in making the decision as to what to observe. More recently, Neisser (1976) has summarized various lines of evidence from cognitive psychology that indicate the human organism is an active, schematizing but also schemata-modifying being who only selectively connects with unencompassable "reality."

It should also be noted that the problems surrounding the notion of "actual behavior" cannot be escaped by shifting to a focus on "objective behavior." So-called "objective" measures of behavior of course can only reflect earlier decisions, sometimes not fully considered, as to where attention should be addressed. A conceptually inappropriate decision as to what aspect of behavior to select as an index of a psychological dimension cannot be redeemed by the subsequent impartiality of the coding of the irrelevant behavior. The scientific merit of "objective behavior" depends, finally, on the psychological incisiveness of what is being figured from the ground, not on the mechanicalness of recording.

What D'Andrade and Shweder appear to mean, then, by "actual behavior" are frequency counts or proportions summarizing arbitrary and narrow selections from the infinitely broad stream of behavior, observed by an individual who need make no inferences and who records the singled-out behaviors immediately. Such counts or proportions are viewed as an ultimate or at least sufficient criterion of truth—as being the benchmark against which all other data can be referenced and tested. However, we suggest that large problems assail even this more restricted view.

2. It is assumed that the Actual Behavior data are simple and direct in two senses: No pertinent behaviors are omitted, and no inference is required of the recording observer. Both of these assumptions are questionable. Interaction Process Analysis, used to provide Actual Behavior data in both of the studies offered by D'Andrade (1965, 1974), has been found to have many problems and is no longer viewed with the enthusiasm that surrounded it when first introduced (see Longabaugh, 1963; O'Dell, 1968; Waxler & Mishler, 1966). IPA coding is a taxing and inundating task: The solitary observer records "acts" continuously, typically at a rate of 10 to 20 scores per minute (Bales, 1968), observing *all* the members of the group for 50 minutes (in the Mann, 1959, study) or for 120 minutes (in the Borgatta et al., 1958, study), coding the onrushing stream of "acts" of group members into 12 supposedly sufficient, mutually exclusive, and by no means simple categories (7 categories in the Mann study) according to the particular

[2] For example, in the Mann (1959) study, a single 50-minute session is a doubtful basis on which to predict reliable Actual Behavior or Rated Behavior scores on five people on multiple characteristics.

"actor" involved. The task requires superhuman information-processing capacities; it is not surprising that flesh-and-blood IPA coders have not fared well when evaluated. Thus, whereas one IPA coder might record 10 acts during a particular minute, another coder might record 20. Because of the number of members in a group, the simultaneity of many "acts", and the differential and oscillating perceptions of observers, various group members receive slighted or unreliable recordings of their "acts." Over the interaction period, observers are often unable to sustain consistency in their codings. The IPA categories are widely criticized as being psychologically heterogeneous and requiring appreciable deliberation, awareness of context, and inference on the part of the observer. The reader should contemplate spending an hour or two watching 5 or 10 individuals, making decisions every 3 to 6 seconds, differentiating only between, for example, IPA Category 1 (shows solidarity, raises others' status, jokes, gives help, rewards), Category 2 (shows tension release, shows satisfaction, laughs), and Category 3 (agrees, shows passive acceptance, understands, concurs, complies). Is an act involving humor a "showing of solidarity" or a "showing of tension release"? Is a remark a "reward" (Category 1) or a "showing of satisfaction" (Category 2)? Certainly, Interaction Process Analysis can provide useful data of a kind. But the technique is highly complicated and is subject to many kinds of problems or vitiating influences often not recognized in the early IPA studies now used unquestioningly by D'Andrade. We suggest that few psychologists will award IPA-based data deferential status as a criterion.

The Newcomb (1929) "behavior records" offered by Shweder (1975) as "actual behavior" also cannot be viewed as involving a sufficiently inclusive and noninferential set of coding rules. Rather, the recording counselor was compelled to make delayed and highly complex judgments with respect only to the categories afforded by Newcomb; behavioral observations not codable into Newcomb's categories could not be registered. By way of illustration, consider Behavior Record Situation 1: "Did [the boy] show confidence in his own abilities?" Newcomb defined four "degrees of response to this situation"; (a) "Boasted loudly of greater abilities than he had"; (b) "Spoke confidently of abilities he really had"; (c) "Expressed lack of confidence in own abilities"; and (d) "Hesitated even to try his ability" (p. 21). It is immediately obvious, when the full text of the categories available to the counselors is referenced, that the superordinate questions (regarding "confidence," "initiative," etc.) asked of them required appreciable psychological inference before the "behavior record" could be completed. Such characteristics as "speaking confidently of abilities one really has" or being "loyal to the leader" are not so behaviorally denotable as to require little human judgment. Also, only one counselor completed the behavior record for each boy, and the extent to which another counselor would have completed that record in the same way is nowhere assessed.

We intend no disparagement of the Newcomb study of a half century ago, a study that for its time was in many ways exemplary. We do suggest that because of the presence of appreciable psychological inference in the behavior records, the sharp distinction necessary between "actual behavior" and "rated behavior" is, instead, blurred.

3. In the IPA studies cited by D'Andrade (1965, 1974) and in the Newcomb study reused by Shweder (1975), "actual behavior" was finally indexed in one of three ways: rate of emission of certain categories of behavior, act percentages, or the percentage of days on which a categorized behavior had been noted. All of these indices of "actual behavior" involve some form of averaging over behavioral instances. Behavior averaging is a very useful technique, but the indices so derived are insensitive to behavior context and the relative salience of any particular instance of coded behavior. All occurrences of a coded act are equally weighted in the "actual behavior" averaging. Thus, in Newcomb's study, a boy who began the camp session as shyly noninteractive and ended up as comfortably gregarious could well earn the same percentage score on the *interaction* dimension as another boy who began camp as aggressively interactive and

became a rejected isolate by the end. The behavior percentage scores used to reference "actual behavior" are logically unable to index such change. On the other hand, the "rated behaviors" in the Newcomb study have the possibility of expressing such changes as may have occurred. Making their ratings at the end of the camp session, the raters had the opportunity to express their integration of all of their understandings of each boy, including such contextual or sequential or salience effects as they believed to be pertinent. In short, useful though the averaging basis underlying the "actual behavior" scores may generally be, such scores may not be able to capture important information and recognitions codable by raters who use long-term memory. Besides the possibility of "systematic distortion" emphasized by D'Andrade (and Shweder) when raters invoke long-term memory, there is also the possibility that raters will be able to discern relationship, meanings, and trends that were denied identifiability in the highly summarizing "actual behavior" averages. Again, we suggest that most personality psychologists will not accept such averages uncritically as a criterion of "actual behavior."

Extraneous influences distorting correlation coefficients and their patterning are not considered. Many factors influence individual correlation coefficients and, thus, the subsequently organized matrix of correlation coefficients. Sometimes these coefficient-influencing factors are not of great importance vis-à-vis the broad questions being asked of the data. But when conclusions hinge on the particular values of particular pattern-making or pattern-breaking coefficients, these considerations become crucial to evaluate and to control. Yet, when D'Andrade and Shweder contrast matrices, they ignore the many psychometric and statistical considerations that influence the observed differences in correlational patterning.

1. The distributions of the Actual Behavior scores and the Rated Behavior scores are not considered. They are often likely to be highly skewed or not unimodal, particularly the IPA rates or percentages representing "actual behavior." Correlations between variables are importantly influenced (and always lowered) by differences in the shapes of the distributions being related (Carroll, 1961).

2. The logic by which IPA "actual behavior" scores are generated for each subject necessarily creates an appreciable degree of *negative* correlation among IPA "actual behavior" scores.[3] The degree of negativity increases as the number of IPA categories decreases; especially affected is the Mann (1959) 7-category IPA system. Although the entailed negative relation among IPA "actual behavior" scores does not uniformly depress the intercorrelations among such IPA scores, the average intercorrelation among these IPA scores will be moved toward negativity by this influence. The interpretation by D'Andrade (1974) of the generally lower intercorrelations among his Actual Behavior matrices as compared to his Rated Behavior matrices does not recognize this lowering effect on the Actual Behavior coefficients.

3. The Actual Behavior data used were derived from only one observer-recorder; four of the five sets of Rated Behavior data were based on pooled judgments from a number of observers. Ceteris paribus, data from a single observer are less dependable than data derived by pooling the data from a number of observers. D'Andrade and Shweder did not consider that to an extent unassessable within the data they used, the discrepancies between their Actual Behavior and Rated Behavior scores and consequent matrices are a function of this source of unreliability in the Actual Behavior data.

4. D'Andrade's IPA analyses involve six or seven IPA categories or variables. Six or seven variables generate, respectively, 15 or 21 coefficients of correlation. The correlation of these Actual Behavior coefficients with corresponding Rated Behavior coefficients yields intermatrix correlations based on *n*s of 15 or 21, extremely small "sample" sizes.

---

[3] Because an individual's IPA category scores are calculated with reference to his total number of acts, the resulting scores are like slices of a pie—as one slice becomes large, the remaining slices are constrained to be small. Guilford (1952) has warned of the dangers of misinterpreting correlations among such reciprocally dependent scores.

Furthermore, *these 15 or 21 "observations" are not independent of each other;* the correlation between Categories X and Y and the correlation between Categories X and Z will constrain the correlation between Categories Y and Z. The distribution of nonindependent observations will tend to be strange and unstable, reacting strongly to changes in the constitution of the set of six or seven categories that happen to have been used. Correlations based on nonindependent observations are not correlations in the usual sense of the term, since the interdependency of observations violates a fundamental assumption of the statistical logic underlying the correlation coefficient. One can no longer interpret such fluky coefficients in terms of amount of variance explained or significance level. The correlation between two correlation matrices has a certain descriptive value, but the intuitive appreciation of these coefficients requires the further recognition that these coefficients are indeed of an unspecifiable but certainly bizarre metric.

5. The direction of scoring or rating a psychological variable is usually arbitrary; one can algebraically reflect a variable without changing its conceptual import. Instead of assigning high numbers to indicate "extroversion," we could as well assign high numbers to mean "absence of extroversion." This long-established recognition has severe implications for the "correlation" between two matrices.

Changes in the direction in which a variable is scored reverse the signs of the correlations of that variable with other variables and will fundamentally influence the "correlation" subsequently computed between these revised matrices (Tellegen, 1965). In the D'Andrade and Shweder analyses, bipolarities tended to characterize their redundancies of measurement. Had the variables defining one end of their dimensions been defined instead in reflected terms, the subsequent correlation matrix would have had fewer negative entries and the "correlation" between correlation matrices would have been importantly smaller. Thus, in yet another way, the results of these analyses are shaped by adventitious aspects of the data.

6. The preceding psychometric recitative has listed seriatim a number of concerns about the D'Andrade and Shweder numerical analyses. Their influence, in the aggregate and in interaction, are beyond ready evaluation, but can be expected to be appreciable.

What can be concluded from the D'Andrade and Shweder reanalyses seeking to unconfound the conceptual similarity interpretation from the isomorphism view? D'Andrade reported a correlation of .34 between the Actual Behavior and Rated Behavior matrices in the Borgatta et al. (1958) study; for the Mann (1959) study, he reported Actual Behavior – Rated Behavior matrix intercorrelations averaging .02 and .24 for the pooled judgment and individual observer Rated Behavior matrices, respectively. Shweder (1975) reports Actual Behavior – Rated Behavior matrix intercorrelations of .51 and .38. Given the psychometric, definitional, conceptual, and methodological problems affecting the comparison of the Actual Behavior and Rated Behavior matrices, we suggested that it is impossible to view these "coefficients" as compelling of any interpretation at all. Instead of being impressed that these "coefficients" are so low, one can as well be surprised that they are so high.

Controlling the Correspondence of Rated Behavior and Conceptual Similarity Structures

The primary demonstration of the conceptual similarity position has been to show that a variety of previously published Rated Behavior matrices or analyses of Rated Behavior matrices are in strong correspondence with Conceptual Similarity matrices or analyses of Conceptual Similarity matrices that are entirely separately generated. D'Andrade and Shweder view the observed correspondence between the Rated Behavior and Conceptual Similarity matrices they have examined as an invariant and widely generalizable datum demanding recognition in their own preferred terms. In disagreement, we suggest that the correspondences noted are a previously unrecognized function of the high redundancy and the structure of that redundancy that just happened to be present

in the particular variable sets D'Andrade and Shweder sampled.

By varying the amount and structure of variable redundancy, thus influencing the extent and pattern of intercorrelations among variables, we have been able to manipulate the subsequent degree of correspondence between Rated Behavior and Conceptual Similarity matrices. Our logical demonstration, together with the recognition of the kinds of redundancy present in the variable sets evaluated by D'Andrade and Shweder, diminishes the implicativeness of their findings.

Method

A Demonstration of the Workings of Redundancy

For our manipulational purposes, a set of personality variables was required from which subsets could be selected that met specified redundancy designs. The California Q-set (CQ-set; Block, 1961, 1971), a broadly ranging set of personality-descriptive items, was used because of its extensive use in personality assessment and because of its multidimensional characteristics. The CQ-set had been used as a basic personality description procedure in a longitudinal study of personality development (Block, 1971). In this study, the personality characteristics of 76 men and 84 women were assessed by means of the CQ-set at three time periods: junior high school, senior high school, and adulthood (when the subjects were in their mid-thirties). Assessment for each time period was by (usually) three clinical psychologists, an entirely different set of psychologists being used for each time interval. Each psychologist formulated his or her description of a subject entirely independently of the judgments by the other psychologists. Furthermore, the information made available for a particular time period was entirely independent of the information made available for the other time periods. For each time period, the multiple CQ-set formulations of a subject's personality were averaged; it is these averaged ratings that serve as our Rated Behavior scores. Over the three time periods, there were 90 common personality variables, equivalently evaluated. The reader is referred to the original report for further details and for information on the various "quality controls" used during the prolonged process of generating these data.

The Rated Behavior data from the middle time period, senior high school, were selected and, for the sexes combined ($N = 160$), the 90 CQ variables were intercorrelated and factor analyzed. The principal components solution was rotated to a varimax criterion, yielding 15 orthogonal factors. The resulting factor loading matrix provided the basis on which selection of stimulus items was made.

Three sets of 12 items each were selected. The first set of CQ items was chosen on the basis of their high loadings, either positive or negative, on the first factor. The 12 items, 6 positive and 6 negative, had a mean first factor loading, disregarding sign, of .73 (all calculations averaging correlations used Fisher's r-to-z transformation). These 12 CQ items represent a *unidimensional* and *bipolar redundancy* and will hereafter be called the Unidimensional, Bipolar Redundancy variable set. The 66 intercorrelations among these 12 variables constitute a Unidimensional, Bipolar Redundancy Rated Behavior matrix. For each of the three time periods, and for each of the sexes, the Unidimensional, Bipolar Redundancy Rated Behavior matrices were calculated.

A second set of 12 CQ items was chosen to represent the first, second, and fourth orthogonal factors. For each factor, four CQ items, two loaded highly positively and two loaded highly negatively, were selected. The average factor loading of these 12 items (vis-à-vis their respective factors and disregarding sign) is .75. These 12 CQ items represent a *multidimensional* and *bipolar redundancy* and will hereafter be called the Multidimensional, Bipolar Redundancy variable set. The 66 intercorrelations among these 12 variables constitute a Multidimensional, Bipolar Redundancy Rated Behavior matrix. The six Multidimensional, Bipolar Redundancy Rated Behavior matrices (derived from the data on three time periods and both sexes) were calculated.

The third set of 12 CQ items was chosen so that each item represented 1 of 12 orthogonal factors. Some of the later factors extracted are defined by only a few CQ items. In selecting an item to represent a factor, some attention was paid to the item's loadings on the other 11 factors so as to preserve item orthogonality but, because of the restrictions on possible item choices, the items finally selected to represent the 12 factors can be said to only *approximate orthogonality*. Each item had a factor loading on its respective factor of at least .62, the mean of these factor loadings being .78. The mean loading of the items representing the last 11 factors with respect to the first factor was .16. We shall call this third set of CQ items the Approximately Orthogonal variable set. The 66 intercorrelations among these 12 variables constitute an Approximately Orthogonal Rated Behavior matrix. The six Approximately Orthogonal Rated Behavior matrices were calculated.

The three sets of variables selected are presented in Table 1.

Within each variable set, each CQ item was paired with every other item, the 66 pairings being typed onto index cards. Each deck of 66 cards was given to conceptual similarity judges with the following instructions:

The purpose of this research is to obtain judg-

Table 1
California Q-Set Items Constituting the Three Variable Sets

Items constituting the Unidimensional, Bipolar Redundancy set
4. Is a talkative individual.
17. Has a rapid personal tempo; behaves and acts quickly.
40. Is facially and/or gesturally expressive.
49. Behaves in an assertive fashion in interpersonal situations.
50. Tends toward undercontrol of needs and impulses; unable to delay gratification.
89. Is self-dramatizing; histrionic.
22. Tends toward overcontrol of needs and impulses; binds tensions excessively; delays gratification unnecessarily.
37. Is vulnerable to real or fancied threat, is generally fearful.
39. Reluctant to commit self to any definite course of action; tends to delay or avoid action. (Uncharacteristic end indicates quick to act.)
45. Aloof, keeps people at a distance; avoids close interpersonal relationships.
72. Tends to ruminate and have persistent, preoccupying thoughts (either pathological or creative).
87. Is emotionally bland; has flattened affect.
Items constituting the Multidimensional, Bipolar Redundancy set
4. Is a talkative individual.
17. Has a rapid personal tempo; behaves and acts quickly.
45. Aloof, keeps people at a distance; avoids close interpersonal relationships.
87. Is emotionally bland; has flattened affect.
24. Shows condescending behavior in relations with others.
33. Is subtly negativistic; tends to undermine and obstruct or sabotage.
5. Behaves in a giving way toward others.
15. Behaves in a sympathetic or considerate manner.
42. Has a brittle ego-defense system; has a small reserve of integration; would be disorganized and maladaptive when under stress or trauma.
52. Is self-defeating.
23. Is productive; gets things done.
67. Is consciously unaware of self-concern; feels satisfied with self.
Items constituting the Approximately Orthogonal set
7. Appears to have a high degree of intellectual capacity (whether actualized or not). (Originality is not necessarily assumed.)
12. Is thin-skinned; vulnerable to anything that can be construed as criticism or an interpersonal slight.
17. Has a rapid personal tempo; behaves and acts quickly.
24. Shows condescending behavior in relations with others.
43. Engages in personal fantasy, daydreams, and speculations.
56. Is concerned with own body and the adequacy of its physiological functioning. (Body cathexis.)
62. Enjoys aesthetic impressions; is aesthetically reactive.
68. Has a clear-cut, internally consistent personality. (Amount of information available before sorting is not intended here.)
73. Interested in members of the opposite sex. (At opposite end, implies absence of such interest.)
74. Is physically attractive. (The cultural criterion is to be applied here.)
85. Tends to offer advice.
86. Values own independence and autonomy.

ments of the similarity between self-descriptions. Each pair of descriptions is on a card. To familiarize yourself with the task, spread the cards out and skim through them. Then, taking the cards in any order you wish, read the two descriptions on a card and decide how *conceptually alike* are the two self-descriptions. Make each judgment according to the following scale, placing each card in the pile corresponding to your judgment: 7 = extremely similar; 6 = moderately

similar; 5 = mildly similar; 4 = unrelated, neither similar nor dissimilar; 3 = mildly dissimilar; 2 = moderately dissimilar; 1 = extremely dissimilar.

These instructions were adapted from those previously employed by D'Andrade and Shweder.

For the Unidimensional, Bipolar Redundancy variable set and the Multidimensional, Bipolar Redundancy variable set, 10 judges each made similarity judgments. For the Approximately Orthogonal set, 20 judges were used. No judge rated more than one variable set. The judges were undergraduates in a course on the psychology of personality and were uninformed as to both the nature and the purpose of the study. Two of the 20 judges of the Approximately Orthogonal variable set did not correctly complete the task, reducing this final total to 18.

Results

The 66 judgments of conceptual similarity by a judge is his or her Conceptual Similarity matrix. When the individual Conceptual Similarity matrices of the Unidimensional, Bipolar Redundancy judges are correlated with the six Unidimensional, Bipolar Redundancy Rated Behavior matrices, the average Conceptual Similarity – Rated Behavior intermatrix "correlation" is .63, ranging from .18 to .93. When the individual Conceptual Similarity matrices of the Multidimensional, Bipolar Redundancy judges are correlated with the six Multidimensional, Bipolar Redundancy Rated Behavior matrices, the average Conceptual Similarity – Rated Behavior intermatrix "correlation" is .50, ranging from −.11 to .75. When the individual Conceptual Similarity matrices of the Approximately Orthogonal judges are correlated with the six Approximately Orthogonal Rated Behavior matrices, the average Conceptual Similarity – Rated Behavior intermatrix "correlation" is .21, ranging from −.13 to .46.[4] As expected, the mean Conceptual Similarity – Rated Behavior "correlations" of the Unidimensional, Bipolar Redundancy and Multidimensional, Bipolar Redundancy groups are not significantly different ($t = .90$, $p < .38$). When the Unidimensional, Bipolar Redundancy and Multidimensional, Bipolar Redundancy judging groups are combined and contrasted with the Approximately Orthogonal judging group, the difference between their respective mean Conceptual Similarity – Rated Behavior intermatrix "correlations" is highly significant ($t = 5.17$, $p < .00001$). Thus these data reveal, strikingly, that Conceptual Similarity – Rated Behavior intermatrix correspondence is high when bipolar redundancy is high and is low when bipolar redundancy is low. Judges appear to be rather faithful reflectors of the pattern of redundancy empirically existing within the set of variables they are evaluating for conceptual similarity.

As a related finding, the average intermatrix "correlation" among the Conceptual Similarity matrices contributed by the 10 Unidimensional, Bipolar Redundancy judges was .41. The average "intercorrelation" among the Conceptual Similarity matrices of the 10 Multidimensional, Bipolar Redundancy judges was .36. The average "intercorrelation" among the Conceptual Similarity matrices of the 18 Approximately Orthogonal judges was .21. Thus, it appears that the ability of judges to agree among themselves in their judgments of conceptual similarity is a clear function of the pattern of redundancy present among the variables being evaluated.

Redundancy in the Variable Sets Evaluated by D'Andrade and Shweder

The pertinence of our demonstration of the influence of redundancy on the "correlation" between Rated Behavior and Conceptual Similarity matrices depends on the extent to which the amount and pattern of redundancy, previously unrecognized as operative, characterizes the variable sets previously found to be "retrievable" by Conceptual Similarity judgments.

Inspecting the variable sets that have been used, it is clear, we suggest, that appreciable and highly patterned redundancy does indeed characterize them.

The 20 personality terms of Norman (1963), further analyzed by D'Andrade (1965), were a carefully refined and well-

[4] It should be noted that the generally small but consistent loadings of the Approximately Orthogonal variables on the first factor contribute to the mean Conceptual Similarity – Rated Behavior intermatrix correlation of .21.

studied variable set intended to provide redundant specification of an orthogonal five-factor structure deemed conceptually suitable for encompassing peer-nomination data. Our Multidimensional, Bipolar variable set was designed to emulate the redundancy design built into the Norman terms.

The six rated variables in the Borgatta et al. study were included in a factor analysis (1958, Table 3). Four of the variables loaded highly on their first interpersonal factor; the remaining two loaded highly on their second factor. Our Multidimensional, Bipolar variable set approximately represents the redundancy characteristics of the Borgatta et al. set of rated variables.

The six (or seven) rated variables in the Mann (1959) study were not submitted to a factor analysis but, since they parallel closely the variables used by Borgatta et al., it is probably safe to assume that the pattern of redundancy characteristic of the earlier study also characterizes the later study.

The 26 variables used by Newcomb (1929) were not and cannot now be studied empirically with regard to their redundancy characteristics. However, because Newcomb selected his variables for their presumed relevance to the single broad dimension of extroversion–introversion, we suggest that appreciable redundancy of the unidimensional, bipolar type characterizes the Newcomb variable set.

The three other variable sets whose matrices are reported by Shweder (1975, 1977a) to be "retrievable" via conceptual similarity judgments—those of Bales (1970), of Sears, Maccoby, and Levin (1957), and of Block (1965)—all were factor analyzed, as reported in their initial publications, and all manifested appreciable redundancy of the multidimensional, bipolar, or unidimensional, bipolar type, according to the investigators' measurement intentions.

Thus, bipolar redundancy within the variable sets appears to be a strong and constant feature of the D'Andrade and Shweder analyses. This bipolar redundancy is generally antonymic (and synonymic) in nature, thus offering a semantic basis to judges for estimating, if only approximately, the nonzero correlations among various pairs of variables in the set. To the extent that Rated Behavior variables are measured validly but not with antonymic (and synonymic) redundancy, the possibility of correspondence ("retrievability") between a Rated Behavior and a Conceptual Similarity matrix tends to disappear because the "similarity structure" of each matrix tends to disappear. The external validity of the variables in the variable set, either singly or as subsets, is in no way contingent on the "retrievability" or the absence of "retrievability" of a Rated Behavior matrix by a Conceptual Similarity matrix. To this last concern—the predictive and concurrent validity of personality ratings—we now turn.

Some Empirical Relationships Currently Beyond the Reach of the Conceptual Similarity Position

For assessment psychologists, a frustrating aspect of the current controversy has been the narrowly delineated arena in which the argument has been waged. In this section we introduce some relationships of a kind that personality psychologists find reinforcing of their belief in reliable individual differences with regard to personality parameters of consequence in the real world. We believe the kinds of across-time or across-domain relationships to be reported are either unassimilable to the conceptual similarity interpretation or else are encompassable only by a string of conjectures so complicated and tenuous as to deprive the conceptual similarity view of interest until such time as the necessary empirical support becomes available. Our strategy for affirming the usefulness and even the necessity of the conceptual–empirical enterprise of personality requires even the necessity of the conceptual–emphasis on *classifications of personality variables*. We focus instead on the construct validity of *classifications of individuals with respect to personality variables*. Space limitations permit only brief citation of some of the available data confronting the conceptual similarity position.

The six Approximately Orthogonal Rated

Behavior matrices could not be well "retrieved" by the Approximately Orthogonal Conceptual Similarity matrix. By the reasoning of D'Andrade and Shweder, it follows therefore that "an illusion of . . . behavioral consistency" (Shweder, 1975, p. 456) was *not* created by the conceptual similarity among the personality variables involved simply because there was not appreciable "conceptual likeness" among the variables being evaluated. If there is behavioral consistency nevertheless, such consistency is not illusory.

The "behavioral consistency view" or "individual difference theory of personality," as defined by Shweder (and as rejected by him), assumes that an individual's behavior will tend to be "consistent from one time to another, and different from the behavior other people would manifest" (Child, 1968, p. 83). A direct test of behavioral consistency is the correlation of a Rated Behavior variable as rated at one time with the same Rated Behavior variable as independently rated at another time.

For the 12 Approximately Orthogonal Rated Behavior variables, across-time correlations are available, for the sexes separately, connecting the junior high school (JHS) and senior high school (SHS) ratings (a period of 3 years), connecting the SHS and Adulthood ratings (a period of about 20 years), and connecting the JHS and Adulthood ratings (a period of about 23 years). It should be recalled again that these ratings were based on entirely independent sets of data and entirely independent sets of judges.

For the male and female samples from JHS to SHS, the 12 Approximately Orthogonal rated variables correlate, on the average, .42 and .40, respectively. If these correlations are adjusted for their attenuation due to unreliability, the mean correlations rise to .68 and .66. From SHS to Adulthood, the average correlations are .27 and .30 for the two sexes, respectively. Adjusted for attenuation, these mean correlations become .41 and .49. From JHS to Adulthood, the average correlations are .24 and .28 for the two sexes, respectively. Adjusted for attenuation, these mean correlations are estimated to be .36 and .46. It should be noted that the 12 variables for which across-time person-ordering consistency is reported were *not* selected for consideration because of the across-time consistency they manifested. Many more such across-time correlations exist, the full set being reported by Block (1971). It should also be recognized that these across-time correlations are further and appreciably lowered by such developmental changes and transformations as take place from early to late adolescence and from late adolescence to the mid-thirties.

All things considered, we suggest that these across-time correlations are substantial evidence for the existence of sufficient behavioral consistency in sufficient people over sufficient time to warrant maintenence of "the individual difference theory of personality." A close and empirically supported explanation of these and related data in conceptual similarity or other attributional terms has yet to appear.

In seeking to dispense with personality ratings and personality inventories on the basis of their conceptual similarity arguments and analyses, D'Andrade and Shweder appear to have presumed the essential invalidity of personality assessment procedures. However, there exists a large and diverse personality assessment literature attesting to construct-valid connections between assessment measures and the real world that has been overlooked by D'Andrade and Shweder (e.g., Block, 1957; Block, 1971, pp. 159–168, 189–202, 228–238; Block, Jennings, Harvey, & Simpson, 1964; Clausen, 1968; Helson, 1971; Jones, 1968, 1971; Kety, 1974; MacKinnon, 1962; Manheimer & Mellinger, 1967; Peskin, 1973; Robins, 1966; Rosenthal, 1970). A priori, it would appear difficult for a conceptual similarity viewpoint to explain the existence of relationships such as these, sampled from literally hundreds that could have been offered.

The Need to Deepen, Clarify, and Substantiate the Meaning of Conceptual Similarity

The conceptual similarity position would replace the assumption that personality parameters (traits, dispositions) importantly

underlie personality ratings with a linguistic explanation. Leaving aside the empirical relationships that the conceptual similarity interpretation must attempt to encompass, we note that this proffered explanation does not resolve the problem of understanding; it merely transfers or translates the problem. We must still face up to the thorny question of the nature and basis of similarity judgments. The study of similarity is a profoundly complex area of conceptual and empirical inquiry (Tversky, 1977); the invocation of "conceptual similarity" as an explanation without immediate close consideration and articulation of the processes by which similarity judgments are made is not especially helpful. D'Andrade was circumspect in his final claims, acknowledging that because of obscurities surrounding the way in which similarities are developed and recognized,

> It is somewhat ironic that the problem of not knowing what is happening when people are used as measuring instruments, . . . given as the reason why [personality or interactional ratings] produce invalid results, returns to plague [his] formulation about exactly how these invalid results come about. (1974, p. 180)

Shweder was less cautious in his view. He argued that judges make similarity judgments based on the degree to which contiguous attitudes go together in making up what he called a behavioral type, which he treated as a learned cultural construct. Shweder went on vigorously:

> What is disputed is that the *categories* (in our terms, personality characteristics) into which people sort themselves or others can be induced from experience. Cultural *constructs* are not empirical generalizations. (Shweder, 1977b, p. 938)

It would be useful to see the argument and empiricism in support of this strong but undetailed position. Our own view is that although cultural constructs need not be evolved from living in the world, they very often are. Individuals frequently independently generate convergent terms to express their recognition about the people they encounter. To the extent that language is an efficient means of communicating calibrated understandings, individuals will or can come to agree in their usage of terms to describe or integrate or summarize the implications of the behavior of other people (cf. Norman & Goldberg, 1966). The developmental sequence by which children evolve understandings of other people has received some study (Peevers & Secord, 1973) and appears to express the *pragmatic* relevance that descriptions of others in terms of personality characteristics take on for individuals as they interact with the people in their world. Such individually evolved and yet consensual structurings of interpersonal experience provide the basis, we suggest, for many learned cultural constructs. Indeed, with respect to certain personality characteristics or constellations rooted in the common human biology, very different cultures may evolve essentially equivalent cultural constructs encompassing various forms of psychopathology. In a splendid essay, Murphy (1976) brought together evidence from many and diverse societies, primitive and advanced, testifying to a larger consensuality in the cultural constructs of psychopathology that societies separately have experientially educed and codified.

Finally, we note that although the effects attributable to manipulation of the redundancy and structure of the variables being rated are large, the range of correlations between an individual's Conceptual Similarity matrix and the Rated Behavior matrix is interestingly wide. Moreover, the phenomenon previously noted (Block, 1977b) was again observed: A judge's degree of consensuality in conceptual similarity judging, even for the Approximately Orthogonal set, was related to his ability to "retrieve" the Approximately Orthogonal Rated Behavior matrix ($r = .46$, $p < .06$).

Separately, we will be studying the possible implications of these individual differences. There are preliminary indications that these differences are reliable and can provide additional perspective on the processes involved in the effort to develop, conceptualize, integrate, and apply one's psychological understanding of people. Meanwhile, we suggest that the inherent intricacy characterizing both human judgment and the ongoing stream of interpersonal behavior makes it unlikely that so simple a principle as conceptual simi-

larity will provide an appreciable explanation of so complex a process as personality assessment.

References

Bales, R. F. *Interaction process analysis: A method for the study of small groups.* Reading, Mass.: Addison-Wesley, 1951.

Bales, R. F. Interaction process analysis. *International Encyclopedia of the Social Sciences,* 1968, *7,* 465–471.

Bales, R. F. *Personality and interpersonal behavior.* New York: Holt, Rinehart & Winston, 1970.

Block, J. A study of affective responsiveness in a lie-detection situation. *Journal of Abnormal and Social Psychology,* 1957, *55,* 11–15.

Block, J. *The Q-sort method in personality assessment and psychiatric research.* Springfield, Ill.: Charles C Thomas, 1961.

Block, J. The equivalence of measures and the correction for attenuation. *Psychological Bulletin,* 1963, *60,* 152–156.

Block, J. Recognizing attenuation effects in the strategy of research. *Psychological Bulletin,* 1964, *62,* 214–217.

Block, J. *The challenge of response sets.* New York: Appleton-Century-Crofts, 1965.

Block, J. *Lives through time.* Berkeley, Calif.: Bancroft Books, 1971.

Block, J. Advancing the science of personality: Paradigmatic shift or improving the quality of research? In D. Magnusson & N. S. Endler (Eds.), *Psychology at the crossroads: Current issues in interactional psychology.* Hillsdale, N.J.: Erlbaum (Wiley), 1977. (a)

Block, J. An illusory interpretation of the first factor of the MMPI: A reply to Shweder. *Journal of Consulting and Clinical Psychology,* 1977, *45,* 930–935. (b)

Block, J. H., Jennings, P. H., Harvey, E., & Simpson, E. Interaction between allergic potential and psychopathology in childhood asthma. *Psychosomatic Medicine,* 1964, *26,* 307–320.

Borgatta, E. F., Cottrell, L. S., & Mann, J. H. The spectrum of individual interaction characteristics: An interdimensional analysis. *Psychological Reports,* 1958, *4,* 279–319.

Carroll, J. B. The nature of the data, or how to choose a correlation coefficient. *Psychometrika,* 1961, *26,* 347–372.

Child, I. L. Personality in culture. In E. F. Borgatta & W. W. Lambert (Eds.), *Handbook of personality theory and research.* Chicago: Rand McNally, 1968.

Clausen, J. A. Adolescent antecedents of cigarette smoking: Data from the Oakland growth study. *Social Science and Medicine,* 1968, *1,* 357–382.

D'Andrade, R. G. Trait psychology and componential analysis. *American Anthropologist,* 1965, *67,* 215–228.

D'Andrade, R. G. Memory and the assessment of behavior. In H. Blalock (Ed.), *Measurement in the social sciences.* Chicago: Aldine, 1974.

Gough, H. G. Personality and personality assessment. In M. D. Dunnette (Ed.), *Handbook of industrial and organizational psychology.* Chicago: Rand McNally, 1976.

Guilford, J. P. When not to factor analyze. *Psychological Bulletin,* 1952, *49,* 26–37.

Helson, R. Women mathematicians and the creative personality. *Journal of Consulting and Clinical Psychology,* 1971, *36,* 210–220.

Hogan, R., DeSoto, C. B., & Solano, C. Traits, tests, and personality research. *American Psychologist,* 1977, *32,* 255–264.

Jones, M. C. Personality correlates and antecedents of drinking patterns in adult males. *Journal of Consulting and Clinical Psychology,* 1968, *32,* 2–12.

Jones, M. C. Personality antecedents and correlates of drinking patterns in women. *Journal of Consulting and Clinical Psychology,* 1971, *36,* 61–69.

Kaplan, A. *The conduct of inquiry.* San Francisco, Calif.: Chandler, 1964.

Kety, S. S. From rationalization to reason. *American Journal of Psychiatry,* 1974, *131,* 957–963.

Longabaugh, R. A category system for coding interpersonal behavior as social exchange. *Sociometry,* 1963, *26,* 319–344.

Lykken, D. T. Multiple factor analysis and personality research. *Journal of Experimental Research in Personality,* 1971, *5,* 161–170.

MacKinnon, D. W. The personality correlates of creativity: A study of American architects. In G. S. Nielsen (Ed.), *Proceedings of the 14th International Congress of Applied Psychology, Copenhagen, 1961* (Vol. 2). Copenhagen: Munksgaard, 1962.

Manheimer, D. I., & Mellinger, G. D. Personality characteristics of the child accident repeater. *Child Development,* 1967, *38,* 491–513.

Mann, R. D. *The relation between personality characteristics and individual performance in small groups.* Unpublished doctoral dissertation, University of Michigan, 1959.

Mischel, W. *Personality and assessment.* New York: Wiley, 1968.

Mischel, W. Continuity and change in personality. *American Psychologist,* 1969, *24,* 1012–1018.

Mischel, W. Toward a cognitive social learning reconceptualization of personality. *Psychological Review,* 1973, *80,* 252–283.

Mischel, W. The interaction of person and situation. In D. Magnusson & N. S. Endler (Eds.), *Psychology at the crossroads: Current issues in interactional psychology.* Hillsdale, N.J.: Erlbaum (Wiley), 1977.

Murphy, J. M. Psychiatric labeling in cross-cultural perspective. *Science,* 1976, *191,* 1019–1028.

Neisser, U. *Cognition and reality.* San Francisco, Calif.: Freeman, 1976.

Newcomb, T. M. *The consistency of certain extrovert–introvert behavior patterns in 51 problem boys.* Contributions to Education No. 382. New York: Teachers College, Columbia University, 1929.

Norman, W. T. Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer-nomination personality ratings. *Journal of Abnormal and Social Psychology*, 1963, *66*, 574–584.

Norman, W. T., & Goldberg, L. R. Raters, ratees, and randomness in personality structure. *Journal of Personality and Social Psychology*, 1966, *4*, 681–691.

O'Dell, J. W. Group size and emotional interaction. *Journal of Personality and Social Psychology*, 1968, *8*, 75–78.

Peevers, B. H., & Secord, P. F. Developmental changes in attribution of descriptive concepts to persons. *Journal of Personality and Social Psychology*, 1973, *27*, 120–129.

Peskin, H. Influence of the developmental schedule of puberty on learning and ego functioning. *Journal of Youth and Adolescence*, 1973, *2*, 273–290.

Robins, L. *Deviant children grown up*. Baltimore, Md.: Williams & Wilkins, 1966.

Rosenthal, D. *Genetic theory and abnormal behavior*. New York: McGraw-Hill, 1970.

Schneider, D. J. Implicit personality theory: A review. *Psychological Bulletin*, 1973, *79*, 294–309.

Sears, R. R., Maccoby, E. E., & Levin, H. *Patterns of child rearing*. New York: Harper & Row, 1957.

Shweder, R. A. How relevant is an individual difference theory of personality? *Journal of Personality*, 1975, *43*, 455–484.

Shweder, R. A. Illusory correlations and the MMPI controversy. *Journal of Consulting and Clinical Psychology*, 1977, *45*, 917–924. (a)

Shweder, R. A. Illusory correlation and the MMPI controversy: Reply to some of the allusions and elusions in Block's and Edwards' commentaries. *Journal of Consulting and Clinical Psychology*, 1977, *45*, 936–940. (b)

Shweder, R. A., & D'Andrade, R. G. Accurate reflection or systematic distortion? A reply to Block, Weiss, and Thorne. *Journal of Personality and Social Psychology*, 1979, *37*, 1075–1084.

Stagner, R. On the reality and relevance of traits. *Journal of General Psychology*, 1977, *96*, 185–207.

Tellegen, A. Direction of measurement: A source of misinterpretation. *Psychological Bulletin*, 1965, *63*, 233–243.

Tversky, A. Features of similarity. *Psychological Review*, 1977, *84*, 327–353.

Waxler, N. E., & Mishler, E. G. Scoring and reliability problems in interaction process analysis: A methodological note. *Sociometry*, 1966, *29*, 28–40.

Received May 12, 1978 ■

Postscript

1. If personality ratings demonstrate conceptually expected concurrent and predictive validity, the semantic "retrieval" of their intercorrelations by Shweder and D'Andrade (1979) takes on a very different implication and significance. We cited a sampling of evidence, from a variety of sources, that ratings can generate replicable external relationships that support their empirical validity. Shweder and D'Andrade state that evidence of empirical validity is "beside the point" (p. 1079) in evaluating the question of the presence or absence of conceptual factors in judges; we do not think so.

2. Interested readers will simply have to judge for themselves, by referring to the original sources, whether the behavior categories worded differently from the rating dimensions "still maintain equivalent meaning," (p. 1078) as Shweder and D'Andrade declare.

3. Because we are without the space for proper psychometric explanation, we can only reiterate that the studies reported by Shweder and D'Andrade "provide no (or technically unsound) reliability coefficients."

4. Shweder and D'Andrade remain impressed by the generality of their finding that appreciable intermatrix correspondence between what they called "actual behaviors" and "rated behaviors" is not observed. We suggest that *the common effect of the many problems besetting their analyses is to increase unreliability and invalidity of measurement and thus to attenuate the possibility of finding strong congruence between Actual Behavior and Rated Behavior matrices*. We agree with Shweder and D'Andrade (1979) that "it is the consistency with which a given result is obtained across variations in method that ultimately supports or disconfirms a hypothesis" (p. 1080). But "the given result" must be a *positive* result, not repeated failures to reject the null hypothesis. The motivated reader should look again at our catalogue of the problems surrounding the comparison of Actual Behavior and Rated Behavior matrices before deciding whether the generally poor intermatrix correspondences observed by Shweder and D'Andrade are deserving of substantive interpretation or are more simply understandable as due to comparing the noncomparable.

5. In our study, we argued that "retrieval" of the matrix of correlations among Rated Behaviors by a matrix of relations among Conceptual Similarity judgments was a function of the particular mix of personality variables employed, namely, their bipolar psychological redundancy. Shweder and D'Andrade (1979) suggest that our reduction of the variance of the distribution of the intervariable correlations creates an "artifactual effect" (p. 1077). But, equally, the high intermatrix correlations previously reported by Shweder and D'Andrade can be viewed as an

"artifactual effect" of extended variances of the distribution of intervariable correlations. Properly, arguments should not arise about artifactual effects on intermatrix congruences; the interdependent intervariable correlations from which they are computed have no "natural" population to which they can be referred so that one can sensibly talk of artificiality or representativeness of their distribution. What is important is that *we have been able to provide a psychological explanation, in terms of redundancy and the proper measurement goals of investigators, of how these different extremes of "artificiality" come about and have their effect.* In particular, high redundancy, both antonymic and synonymic, often has been deliberately introduced by investigators for good reasons (to balance direction of wording effects, to achieve psychometric respectability of subsequent composite scores, etc.). To make such redundancy subtle or indirect would be confusing to the psychologists using these rating scales. It is therefore not surprising that conceptual similarity judges are able to latch onto the structure underlying cleanly separated clusters of variables when highly and simply organized variable sets are employed. But such congruences are essentially adventitious; they have no implications, in and of themselves, for the validity and usefulness of well-based and well-encoded personality ratings.

Journal of Personality and Social Psychology
1979, Vol. 37, No. 6, 1075-1084

Accurate Reflection or Systematic Distortion? A Reply to Block, Weiss, and Thorne

Richard A. Shweder
Committee on Human Development
University of Chicago

Roy G. D'Andrade
Department of Anthropology
University of California, San Diego

The "systematic distortion" hypothesis, advocated by D'Andrade and Shweder, states that correlational structures derived from memory-based personality ratings are primarily a product of the conceptual affiliations among rating categories rather than a reflection of the empirical correlational structure of behavior. This hypothesis is clarified with reference to the Block, Weiss, and Thorne critique. The "bipolar redundancy" effect, presented by Block, Weiss, and Thorne as an alternative explanation for correspondence between the correlations derived from memory-based ratings and judgments of conceptual affiliation, is found to be an artifact of reduced variance in matrices of uncorrelated traits. Block, Weiss, and Thorne argue that D'Andrade and Shweder failed to find correspondences between correlations for memory-based ratings and immediate scorings because of methodological flaws in their studies. This argument is reviewed, and it is concluded that the systematic distortion hypothesis remains well supported and that the evidence for the existence of covarying, multibehavior personality traits established by the correlational analysis of memory-based ratings remains dubious. Finally, an attempt is made to clarify the different uses of the term *trait* and to identify the different types of consistency in behavior that are relevant to the controversy.

Most classifications of individual differences in the personality literature are classifications of response patterns on personality rating forms, inventories, and questionnaire interviews. The pattern of correlations among variables in most of these classifications can be reproduced or replicated by asking a small number of respondents to judge the conceptual similarity or dissimilarity of the variables on the rating form, inventory, or questionnaire. To date, results from the following analyses of rating data have been successfully reproduced:

1. Factor-analytic classification of personality adjectives, as given in Norman (1963). (See D'Andrade, 1965; Mulaik, 1964.)

2. Leary grid organization of interpersonal behavior, as given in LaForge and Suzcek (1955). (See D'Andrade, 1965.)

3. Factor-analytic classification of personality and interpersonal behavior, as given in Bales (1970). (See Shweder, 1972, 1975.)

4. Factor-analytic classification of maternal personality, as given in Sears, Maccoby, and Levin (1957). (See Shweder, 1975.)

5. Correlation matrices for Bales Interaction Process Analysis categories used as rating scales, as given in Borgatta, Cottrell, and Mann (1958) and in Mann (1959). (See D'Andrade, 1974.)

6. Partial correlation matrix for observers' ratings of extraversion–introversion in boys in summer camp, as given in Newcomb (1929). (See Shweder, 1975, 1977a.)

7. The Alpha factor from the MMPI, as given in Block (1965). (See Shweder, 1977a, 1977b.)

8. The common factor structure of the

Requests for reprints should be sent to Richard A. Shweder, Committee on Human Development, University of Chicago, 5730 South Woodlawn Avenue, Chicago, Illinois 60637, or Roy G. D'Andrade, Department of Anthropology, University of California at San Diego, La Jolla, California 92093.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3706-1075$00.75

Murray Needs from five personality tests, as given in Fiske (1973) and in Huba and Hamilton (1976). (See Ebbesen & Allen, Note 1.)

9. Syndrome clusters from the Brief Psychiatric Rating Scales, as given in Overall, Hollister, and Pichot (1967). (See Shweder & D'Andrade, Note 2.)

In these studies a variety of techniques for obtaining similarity judgments have been used, ranging from pile sorting to triads testing. The phrasing of the instructions for these judgments has also varied from an emphasis on "co-occurrence" to an emphasis on "likeness in meaning." Similarity judgments appear to be robust, in that small differences in technique do not seem to affect the correspondence found between conceptual similarity judgments and memory-based personality ratings.

At least two hypotheses can be constructed to account for the fact that conceptual similarity judgments reproduce the correlational structure of variables in personality ratings. The first is the "accurate reflection" hypothesis, which asserts that ordinary folk learn or develop "implicit personality theories" that summarize and preserve the empirical covariation of behavior across individual differences in conduct. According to this hypothesis, people use empirically valid implicit personality theories in making conceptual similarity judgments, thereby accurately reporting the intercorrelation of behaviors (Jackson, Chan, & Stricker, in press; Passini & Norman, 1966).

The second hypothesis is the "systematic distortion" hypothesis, which asserts that the correspondence between preexisting ideas of what is like what and memory-based rating correlations occurs because, under difficult memory conditions, people infer what happened from their general model of what the world is like, and because conceptually affiliated memory items are easier to retrieve (Mandler, 1970). According to the systematic distortion hypothesis, the schemata held by most people tend to be inaccurate with respect to how behaviors covary, confusing "what is like what" with "what goes with what." Hence, memory for events contains a systematic bias, in that things that are conceptually similar are recalled as if they covaried. L. J. Chapman (1967) has called this effect "illusory correlation."

A test of the accurate reflection hypothesis versus the systematic distortion hypothesis can be made by obtaining both observational evidence and memory-based ratings for a set of behavioral variables. By observational evidence we mean reasonably objective records made at the time of observation. Such records are called here "immediate scorings." (Block, Weiss, and Thorne, 1979, refer to this type of data as "Actual Behavior matrices.") To the extent that the correlations among variables derived from memory-based ratings are not like the correlations among variables derived from immediate scorings, but are like the pattern of conceptual similarity judgments, the accurate reflection hypothesis is disconfirmed, and the systematic distortion hypothesis supported. (See Shweder, 1977c, Note 3, for other ways to test the accurate reflection and systematic distortion hypotheses.)

In previous papers (D'Andrade, 1973, 1974; Shweder, 1975, 1977a) we examined published sets of data in which immediate scorings and memory-based ratings for the same variables were available. In all cases the pattern of intercorrelations for memory-based ratings was like the pattern of conceptual similarity judgments but only weakly related to the pattern of intercorrelations for the immediate scorings. The evidence supported the systematic distortion hypothesis, and we concluded that personality classifications derived from memory-based data could not be trusted.

Block, Weiss, and Thorne (1979) assert that the systematic distortion hypothesis is not adequately supported. Two general arguments are advanced: the "bipolar redundancy" argument and a series of methodological criticisms.

The Bipolar Redundancy Argument

Block, Weiss, and Thorne (1979) claim that the correspondence between conceptual similarity matrices and correlation matrices for memory-based ratings is "a previously unrecognized function of the high redundancy and

the structure of that redundancy that just happened to be present in the particular variable sets D'Andrade and Shweder sampled" (pp. 1065–1066). They attempt to show this by demonstrating that by varying the degree of bipolar redundancy among variables, they can affect the degree of correspondence between conceptual similarity matrices and correlation matrices for memory-based behavior ratings. Using the California *Q*-set variables and a factor analysis based on ratings by clinical psychologists of a sample of 160 subjects, they selected three sets of 12 items each. For the first set, 6 items with high positive loadings and 6 items with high negative loadings from a single factor were selected; for the second set, 2 items with high positive loadings and 2 items with high negative loadings from three orthogonal factors were selected; for the third set, single items with high loadings from each of 12 approximately orthogonal factors were selected.

For each of these three sets, conceptual similarity judgments were made by a small sample of respondents. For the first set, the mean correlation between the individual conceptual similarity matrices and the rated behavior matrix was found to be .63. For the second set, the mean correlation was found to be .50. But for the third set, which is made up of items from orthogonal factors, the correlation was found to be only .21.

Based on the fact that the mean intermatrix correlation for the orthogonal items is significantly less than the mean intermatrix correlations for items selected from bipolar factors, Block, Weiss, and Thorne (1979) conclude that "Conceptual Similarity – Rated Behavior intermatrix correspondence is high when bipolar redundancy is high and is low when bipolar redundancy is low" (p. 1068) and that they have therefore controlled the degree of correspondence between conceptual similarity and memory-based rating correlations by varying the degree of bipolar redundancy.

We believe that Block, Weiss, and Thorne are incorrect in this claim, since the artifactual effect of the reduced variance within the matrix of orthogonal items has been overlooked. Using Block, Weiss, and Thorne's original data, kindly supplied to us by the authors, and comparing the three data sets, we found standard deviations of .53 and .37 in the rated behavior correlation matrices for the first and second sets, in contrast to a standard deviation of only .15 for the third set. The variability of the rated behavior correlations in the third set has been much reduced by the selection process, and this in turn has reduced the possible covariance between the matrices.

It is also crucial to note that contrary to Block, Weiss, and Thorne's claim, *respondent judgments about the conceptual similarity of the items in the third set do in fact correspond to the level of the correlations in the rated behavior matrix.* Our reanalysis of Block, Weiss, and Thorne's data reveals that the mean conceptual similarity score for items in the third set is 4.3 on a scale for which 1 is defined as "extremely dissimilar," 4 as "unrelated," and 7 as "extremely similar." The mean of the correlations in the rated behavior matrix is .12 disregarding sign. Thus, in a set of almost orthogonal items, conceptual similarity judgments indicate that the items are almost unrelated.

The problem with using correlations to measure the relation between matrices in this instance can be illustrated by considering the extreme case in which all the items are correlated .00 with each other, and all the conceptual similarity judgments are at 4.0, indicating that the items are unrelated. Using correlations as the only measure of correspondence, one would assume that there was no relation between these matrices, even though the judgments and the correlations match perfectly. We conclude that the correspondence between conceptual similarity matrices and correlation matrices for memory-based ratings is not a function of adventitious redundancy.

There is an important methodological problem with the redundancy hypothesis, namely, that the independent variable is operationally the same as one of the dependent variables. By "redundancy" Block, Weiss, and Thorne mean the "communal variance" shared by a set of variables. But communal variance is a function of the intercorrelations between the variables. Therefore, in trying to control for redundancy, they are in effect controlling for

the intercorrelations between variables, which is also one part of the relation they are trying to predict. As a result, in reducing redundancy they necessarily reduce the variance in the intercorrelations between the ratings and, thereby, the possible covariance between matrices.

It is interesting that the redundancy hypothesis asserts part of what the systematic distortion hypothesis asserts: that there is a correspondence between conceptual similarity judgments and the intercorrelations among memory-based behavior ratings (since redundancy is a function of the intercorrelation among variables); however, it reverses the direction of causation and thus becomes a version of the accurate reflection hypothesis.

The Methodological Critique

Block, Weiss, and Thorne (1979) present a series of methodological criticisms concerning the analyses by D'Andrade (1974) and Shweder (1975) of data taken from studies by Newcomb (1929), Borgatta et al. (1958), and Mann (1959). What follows is a list of responses to these criticisms.

1. "The categories or items or dimensions of behavior measured by the Actual Behavior indices are often different from those rated by observers" (Block, Weiss, & Thorne, 1979, p. 1060). This is an important point, since the data were analyzed as if they were from a memory experiment in which the task is to try to remember the relative frequencies of various behaviors across different actors. To the extent that the memory-based ratings were made on categories not equivalent in meaning to the categories used in the immediate scorings, lack of correspondence between the two sets of data becomes artifactual.

Since our analyses were carried out on already-published data, we had no control over how the labeling was done. In our judgment, the categories that do not have identical wording still maintain equivalent meaning. However, it is possible to make a more stringent test by using only categories that have identical wording. We have recomputed the intermatrix correlations for the Mann data using only items that are identically phrased (Categories 2, 3, 4, and 7). The intermatrix correlations between the immediately scored
havior matrices and memory-based rat
matrices for the Mann data are .15, −
−.27, and .03, respectively ($M = -.03$), w
the nonidentical items removed, compared
intermatrix correlations of .07, .20, −.03, a
.27, respectively ($M = .13$), for the data
originally given. Thus removing the items t
were not identically phrased did not incre
(in fact decreased slightly) the corresp
dence between the actual behavior and
memory-based rated behavior matrices.

Since the issue about maintaining equivale
definitions in both conditions is important
evaluating the evidence for or against t
systematic distortion hypothesis, it should
mentioned that relevant data exist in an u
published study (Shweder & D'Andrade, No
2). Block, Weiss, and Thorne did not ha
the results of this study available to them
the time they wrote their critique, nor do t
results of this new study diminish the pert
nence of their concern about this issue. Briefl
this study is similar in design to the studi
already reviewed, except that (a) the cat
gories of behavior were taken from natur
language terms for social interaction (e.g
"advises," "criticizes," etc.), (b) the immed
ate scoring categories and the memory-base
rating categories are identical, and (c) th
stimulus material consisted of videotape
natural family interaction, thereby permittin
immediate scoring by a number of observer
The intermatrix correlations are quite simil
to the results found in the other cases: Th
correlation between the immediately score
behavior matrix and the memory-based be
havior rating matrix is .23; the correlatio
between the memory-based behavior ratin
matrix and the conceptual similarity matri
is .75; and the correlation between the im
mediately scored behavior matrix and the con
ceptual similarity matrix is .00.

In summary, on the basis of the recompute
figures and the new study, we conclude tha
the low degree of correspondence we hav
consistently found between memory-base
rated behavior correlation matrices and im
mediately scored behavior correlation matrice
is not a product of wording changes in be
havioral categories.

2. "Emphasis solely on the correlations between traits ignores consideration of the validity of specific traits" (Block, Weiss, & Thorne, 1979, p. 1061). This is true, but beside the point. The reason for the emphasis on the correlations between traits is that it is these correlations that test the systematic distortion and accurate reflection hypotheses. Of course, if the correspondence between the memory-based data and the immediately scored data were very high, there would not be enough error variance for systematic distortion to take place. However, the correlations between specific immediate scorings and memory-based behavior ratings are in the .3 range. Even rigorous correction for attenuation would not put the level of validity so high that there would be insufficient error variance to permit the effect of systematic distortion to occur.

Related to the issue of validity, Block, Weiss, and Thorne (1979) raise questions about the reliability of the immediate scorings. They assert that "because the three studies provide no (or technically unsound) reliability coefficients, we cannot estimate what the relationships between the Rated Behavior and the Actual Behavior variables would be if they were psychometrically improved" (p. 1062). For the Newcomb (1929) study, an odd–even day reliability check on behavior percentages yielded a mean reliability coefficient of .78 over 26 variables. In the Borgatta et al. (1958) study, reliability was based on comparing the profiles of the three groups across two sessions, with the results presented graphically. From visual inspection there appears to be high consistency in the results for each group across sessions. In the Mann (1959) study, Mann did the scoring alone. However, he had worked with R. F. Bales for 3 years and had attained a scoring reliability of approximately .90 on the majority of the categories, with Bales as the criterion. However, since the reliabilities are not given on a category-by-category basis, Block, Weiss, and Thorne are correct in pointing out that correction for attenuation cannot be made. Nevertheless, the figures indicate that the immediate scorings have reasonable mean reliability in all three studies.

3. "The use of frequency counts of behavior is not a sufficient means of operationalizing complex psychological concepts" (Block, Weiss, & Thorne, 1979, p. 1062). Again, this is true but beside the point. This criticism would be appropriate were we trying to develop a personality assessment instrument. Our main concern in these studies is not to develop personality assessment techniques but to test hypotheses about systematic distortion in memory. In trying to anchor one set of ratings in what can be observed and counted, we have used studies that used relatively simple and direct methods. If we did not have simple and direct measures, how could we know whether memory distortion had occurred?

This same general point holds for the related criticism by Block, Weiss, and Thorne that the categories of Interaction Process Analysis (IPA) are not the most effective means of describing social interaction and (on the other side of the fence) the criticism that the Newcomb categories require considerable psychological inference. Difficulties in applying particular category systems would be likely to lower the validity of the ratings, but would not introduce systematic distortion.

Block, Weiss, and Thorne (1979) also question the use of averages as a way of summarizing behavior. They say, "Useful though the averaging basis underlying the 'actual behavior' scores may generally be, such scores may not be able to capture important information and recognitions codable by raters who use long-term memory" (p. 1064). We do not deny these criticisms of averages, but we do deny the relevance of these criticisms to our research. Repeating the point made above, such criticisms would apply if we were attempting to assess the personality of the subjects rated in these studies. Given that we are trying to test hypotheses about memory, we hold that these criticisms do not apply, since the methods used did not artifactually introduce systematic distortion. Were it the case that the memory-based raters captured important information not to be found in the immediate scorings, this would affect the correlations between specific categories across conditions but would not affect the relation between the correlation matrices.

4. "Extraneous influences distorting correlation coefficients and their patterning are not considered" (Block, Weiss, & Thorne, 1979, p. 1064). First, Block, Weiss, and Thorne make the point that the IPA rates are likely to be highly skewed or not unimodal, which will tend to lower correlations. Second, where the behavior counts are transformed into percentages, some degree of negativity is introduced into the correlation matrix. Third, in four of the five sets of data, the memory-based behavior rating data were based on pooled judgments from a number of observers, but the immediately scored behavior data were obtained from a single observer, so that the discrepancies may be a function of the "volatility" of the immediate scorings.

If we had merely presented one test of the systematic distortion hypothesis that had all of these problems, our results would be equivocal. However, we have presented a number of different tests of the systematic distortion hypothesis on different types of data, collected by different investigators, using different methods. One advantage of using a number of different tests is that criticisms of the data or of the procedures in one case can be matched with a case in which the criticism does not apply. We can then see if the same results obtain. Thus the Newcomb (1929) data do not have distributional difficulties of the sort that might have affected the Mann (1959) data, and these data are also not affected by the negativity problem. Further, in the Mann study there are two sets of data—one in which there is no discrepancy between the number of subjects used in the memory-based behavior rating condition and the number of subjects used in the immediately scored behavior condition, and one in which there is. No difference is observed in the results. Thus, to each of Block, Weiss, and Thorne's three criticisms there are one or more tests to which the criticism does not apply, yet the same pattern of findings occurs. Almost every study has some methodological weak point. We believe it is the consistency with which a given result is obtained across variations in method that ultimately supports or disconfirms a hypothesis.

Also included under the heading of "extraneous influences" by Block, Weiss, a Thorne are two other criticisms. The fi concerns the use of correlations to measu the degree of relatedness between two matric of correlations. Block, Weiss, and Thor point out that because of the lack of ind pendence in the "observations," the norm interpretations of the correlation coefficie with regard to amount of variance explain and levels of significance are inappropria To date we have not found a better way compare these types of correlation matrice However, even under these conditions, the co relation coefficient maintains its descripti value, yielding an interpretable figure, whi indicates how well one can predict from o to the other set of measurements. Since have no alternative, and since we are primari interested in obtaining a comparative figu for the *relative* degree of association betwe various sets of measurements, we have pe force relied on the product-moment a Spearman's rank-order correlation coefficie

The second issue concerns the effect of flecting variables in a matrix of correlatio thereby changing the degree of association b tween matrices. Since some psychological va ables permit such reflections, it is possib that some of the variables could have be reflected, which would possibly have chang the relationship between matrices. Howeve this criticism applies only to some of the stu ies that attempt to reproduce the structur found in various personality assessment i struments. None of the variables we have us in comparing immediately scored behavi matrices with memory-based behavior rati matrices were of the dimensional or bipol types that would permit reflection, and this particular problem does not affect our a tempts to test the systematic distortion h pothesis.

Last on the list of methodological criticism Block, Weiss, and Thorne (1979) raise t question of whether the correlations fou between the immediately scored behavior m trices and memory-based behavior rating m trices are so high that it is impossible to "u confound the conceptual similarity interpret tion from the isomorphism view" (p. 1065 They say, "Instead of being impressed th

these 'coefficients' are so low, one can as well be surprised that they are so high" (p. 1065). However, the test of the systematic distortion versus accurate reflection hypothesis requires the comparison of *three* correlations: (a) the correlation between the memory-based behavior rating matrix and the conceptual similarity matrix, (b) the correlation between the memory-based behavior rating matrix and the immediately scored behavior matrix, and (c) the correlation between the immediately scored behavior matrix and the conceptual similarity matrix.

Across seven different tests we find the following: (a) The mean correlation between memory-based behavior rating matrices and immediately scored behavior matrices is .25; (b) the mean correlation between memory-based behavior rating matrices and conceptual similarity matrices is .75; and (c) the mean correlation between the immediately scored behavior matrices and conceptual similarity matrices is .26.

Comparing the three relationships, we observe that the relation between the matrices for memory-based ratings and the matrices for immediate scorings is weaker than the relation between the matrices for memory-based ratings and the conceptual similarity matrices. This result is not due to some effect of the relation between the immediate scoring matrices and the conceptual similarity matrices, since these are also only weakly related. It is this pattern of correlations that supports the systematic distortion hypothesis, not our impression of whether any one of these relationships is surprisingly strong or weak.

Theoretical Issues

In the final section of their article, Block, Weiss, and Thorne take up a series of general theoretical issues concerning the conceptual similarity position. One of the issues they raise concerns the nature and the basis of similarity judgments. "The study of similarity is a profoundly complex area of conceptual and empirical inquiry (Tversky, 1977); the invocation of 'conceptual similarity' as an explanation without immediate close consideration and articulation of the processes by which similarity judgments are made is not especially helpful" (Block, Weiss, & Thorne, 1979, p. 1071). The study of similarity judgments has been a neglected area. A simple model, informally presented, dealing with only the denotative aspects of meaning, is given in Romney and D'Andrade (1964). It is our current view that similarity judgments are based on the relative number of predications common to the two terms being rated; such predications we believe refer to a wide variety of things, including the sets to which both items belong, parts which are common to both items, common effects of both items, and so on. The study of similarity can be divided into two parts: The first is to determine how information is processed in making a judgment; the second is to determine what informational content is being processed. Multidimensional scaling of similarity judgments has proven to be an effective way of investigating the second problem (Romney & D'Andrade, 1964; Shepard, 1974). It seems likely that with the stimulus of Tversky's model there will soon be an increase in research on the first problem.

Another general issue taken up by Block, Weiss, and Thorne (1979) is the degree to which cultural belief systems can be assumed to be veridical. They point out that although constructs need not be empirically accurate, they often are. Further, they argue that there is good reason to expect cultures and individuals to develop both reliable and valid ways of describing human personality (p. 1071).

As anthropologists interested in the relation between culture and cognition, we support the generalization that cultural belief systems are sometimes empirically accurate. However, having some experience with the strong commitment people often have to beliefs that are empirically inaccurate, we feel that the other side of the proposition also needs stating; that is, cultural belief systems can be quite inaccurate and yet quite compelling. In this instance, the question is whether people have personality traits consisting of covarying behaviors. The fact that this is believed to be true in our culture does not constitute evidence for either its truth or its falsity.

From a more psychologically oriented per-

spective, Block, Weiss, and Thorne (1979) assert: "In seeking to dispense with personality ratings and personality inventories on the basis of their conceptual similarity arguments and analyses, D'Andrade and Shweder appear to have presumed the essential invalidity of personality assessment procedures" (p. 1070). Arguing for the essential validity of personality assessment procedures, Block, Weiss, and Thorne introduce "some relationships of a kind that personality psychologists find reinforcing of their belief in reliable individual differences with regard to personality parameters of consequence in the real world" (p. 1069). As examples of relationships of this kind they point to across-time consistencies in trait ratings and the ability of personality assessment measures to discriminate between criterion groups. Such relationships, they assert, can be neither reproduced nor explained by the conceptual similarity position.

It is true that such relationships, to the extent they did exist, could not be reproduced or explained by reference to conceptual similarity. This is because the conceptual similarity argument has no bearing on the issue of perceived longitudinal consistency or on the asserted ability of personality measures to discriminate between criterion groups.

What becomes apparent in examining the final section of Block, Weiss, and Thorne's article is that there are (at least) two uses of the term *trait* and the related phrase *individual differences*. The conceptual similarity position is part of an attack on the hypothesis that people have global traits and on the related hypothesis that one can determine such traits through the analysis of individual differences. The defense given by Block, Weiss, and Thorne concerns the predictive utility of traits—but in a different sense of the word—and the possibility of determining such traits through the analysis of individual differences—in a different sense of the phrase.

Originally, the notion of a trait appears to have contained four distinct kinds of consistency: consisting across similar behaviors (an honest person does not lie, does not steal, and does not cheat); consistency across situations (an honest person does not cheat when taking a math test and does not cheat when playing a ring toss game); consistency across time (an honest person is likely to have been honest yesterday and will be honest tomorrow); and consistency within a framework of psychological functioning (honesty is theoretically and empirically related to self-control and identification with parental figures). (See McClelland, 1951, 210–234.)

The kind of trait consistency that is relevant to the conceptual similarity position is consistency across similar behaviors. The classical approach to the problem of determining consistency across similar behaviors is correlational analysis, in which the correlations are computed across persons, as typified by the work of Eysenck and Cattell. This approach assumes that individual differences arise from differences between persons in the amount of a trait, and so the co-variation of behaviors across persons can be used to determine which behaviors belong to which traits (Cattell, 1946).

It has been our position that evidence for the existence of this type of trait—this type of consistency across similar behaviors—is rather dubious. Most of the supporting data are from memory-based ratings, questionnaires, and inventories. (Correlational data from self-report tests can be considered "objective evidence" in one sense, but such data do not establish that the behaviors reported on actually co-occur as rated.) We believe that memory-based data are so strongly affected by conceptually based systematic distortion that they are primarily evidence for the existence of a powerful cultural belief system.

However, the definition of *trait* given above does not correspond to the use of the term by Block, Weiss, and Thorne (1979). They say:

> Personality psychologists seek or settle on useful sets of variables to be employed for the specification of the personal or behavioral qualities of the individuals to be studied. To the extent that the set of personality or behavioral variables has communal variance (i.e., is redundant), there will be a "similarity structure of the trait measurements." But such similarity structures represent neither conceptual nor empirical fixities. (p. 1056–1057)

Thus, for Block, Weiss, and Thorne, the emphasis is on useful ratings, which usually involve constructing a composite index across

a number of measures. The different measures to be placed in a composite index do not have to be internally correlated or homogeneous; they may even lack face validity, as long as the composite index displays significant relationships with an external criterion.

With respect to current personality assessment work as described by Block, Weiss, and Thorne, it appears that two of the four types of consistency originally contained in the notion of a trait have been dropped or suspended: consistency across similar behaviors and consistency across different situations. If true, this is an interesting shift. We have been arguing that one of these two suspended types of consistency has little empirical warrant, and we are encouraged to believe that this position no longer appears heretical.

Reference Notes

1. Ebbesen, E. B., & Allen, R. B. *Further evidence concerning Fiske's question: "Can personality constructs ever be validated?"* Unpublished manuscript, 1977. (Available from E. B. Ebbesen, Department of Psychology, University of California at San Diego, La Jolla, California 92037.)
2. Shweder, R. A., & D'Andrade, R. G. *Personality or culture?* Unpublished manuscript, 1978. (Available from R. A. Shweder, Committee on Human Development, University of Chicago, Chicago, Illinois 60637, or R. G. D'Andrade, Department of Anthropology, University of California at San Diego, La Jolla, California 92093.)
3. Shweder, R. A. *Fact and artifact in personality assessment: The influence of conceptual schemata on individual difference judgments.* Paper presented at the 86th annual convention of the American Psychological Association, Toronto, Canada, August 1978. (Available from R. A. Shweder, Committee on Human Development, University of Chicago, Chicago, Illinois 60637.)

References

Bales, R. F. *Personality and interpersonal behavior.* New York: Holt, Rinehart & Winston, 1970.

Block, J. *The challenge of response sets.* New York: Appleton-Century-Crofts, 1965.

Block, J., Weiss, D. S., & Thorne, A. How relevant is a semantic similarity interpretation of personality ratings? *Journal of Personality and Social Psychology,* 1979, *37,* 1055–1074.

Borgatta, E. F., Cottrell, L. S., & Mann, J. H. The spectrum of individual interaction characteristics: An interdimensional analysis. *Psychological Reports,* 1958, *4,* 279–319.

Cattell, R. D. *Description and measurement of personality.* New York: World Book, 1946.

Chapman, L. J. Illusory correlation in observational report. *Journal of Verbal Learning and Verbal Behavior,* 1967, *6,* 151–155.

D'Andrade, R. G. Trait psychology and componential analysis. *American Anthropologist,* 1965, *67,* 215–228.

D'Andrade, R. G. Cultural constructions of reality. In L. Nader & T. W. Maretzki (Eds.), *Cultural illness and health.* Washington, D.C.: American Anthropological Association, 1973.

D'Andrade, R. G. Memory and the assessment of behavior. In T. Blalock (Ed.), *Measurement in the social sciences.* Chicago: Aldine-Atherton, 1974.

Fiske, D. W. Can a personality construct be validated empirically? *Psychological Bulletin,* 1973, *80,* 89–92.

Huba, G. J., & Hamilton, D. L. On the generality of trait relationships: Some analyses based on Fiske's paper. *Psychological Bulletin,* 1976, *83,* 868–875.

Jackson, D. N., Chan, D. W., & Stricker, L. J. Implicit personality theory: Is it illusory? *Journal of Personality,* in press.

LaForge, R., & Suczek, R. F. The interpersonal dimension of personality: III. An interpersonal checklist. *Journal of Personality,* 1955, *24,* 94–112.

Mandler, G. Words, lists and categories: An experimental view of organized memory references. In J. L. Cowan (Ed.), *Studies in thought and language.* Tucson: University of Arizona Press, 1970.

Mann, R. D. *The relation between personality characteristics and individual performance in small groups.* Unpublished doctoral dissertation, University of Michigan, 1959.

McClelland, D. C. *Personality.* New York: Sloane Associates, 1951.

Mulaik, S. A. Are personality factors raters' conceptual factors? *Journal of Consulting Psychology,* 1964, *28,* 506–511.

Newcomb, T. M. The consistency of certain extravert–introvert behavior patterns in 51 problem boys. Contributions to Education No. 382. New York: Teachers College, Columbia University, 1929.

Norman, W. T. Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology,* 1963, *67,* 574–583.

Overall, J. E., Hollister, L. E., & Pichot, P. Major psychiatric disorders: A four-dimensional model. *Archives of General Psychiatry,* 1967, *16,* 146–151.

Passini, F. T., & Norman, W. T. A universal conception of personality structure? *Journal of Personality and Social Psychology,* 1966, *4,* 44–49.

Romney, A. K., & D'Andrade, R. G. Cognitive aspects of English kin terms. *American Anthropologist,* 1964, *66,* 146–170.

Sears, R. R., Maccoby, E. E., & Levin, H. *Patterns of child rearing.* New York: Harper & Row, 1957.

Shepard, R. N. Representation of structure in similarity data: Problems and prospects. *Psychometrika,* 1974, *39,* 373–421.

Shweder, R. A. Semantic structures and personality assessment (Doctoral dissertation, Harvard University, 1972). *Dissertation Abstracts International*, 1972, *33*, 2452B. (University Microfilms No. 72-29,584)

Shweder, R. A. How relevant is an individual difference theory of personality? *Journal of Personality*, 1975, *43*, 455–484.

Shweder, R. A. Illusory correlation and the MMPI controversy. *Journal of Consulting and Clinical Psychology*, 1977, *45*, 917–924. (a)

Shweder, R. A. Illusory correlation and the MMPI controversy: Reply to some of the allusions and elusions in Block's and Edwards' commentaries. *Journal of Consulting and Clinical Psychology*, 1977, *45*, 936–940. (b)

Shweder, R. A. Likeness and likelihood in everyday thought: Magical thinking in judgments about personality. *Current Anthropology*, 1977, *18*, 637–648. (c)

Tversky, A. Features of similarity. *Psychological Review*, 1977, *84*, 327–353.

Received December 24, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 6, 1085–1089

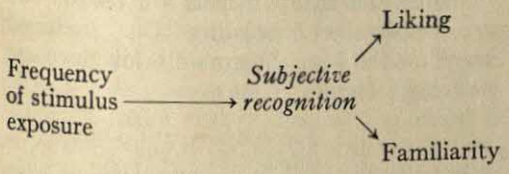
Exposure Effects May Not Depend on Stimulus Recognition

Richard L. Moreland
University of Pittsburgh

Robert B. Zajonc
University of Michigan

Birnbaum and Mellers have presented an alternative theoretical model for Moreland and Zajonc's data on the role of stimulus recognition in the mere exposure phenomenon. They contend that the exposure effects observed in the authors' research could have been produced by a single underlying factor called *subjective recognition*. Birnbaum and Mellers' model is quite plausible, and follows the traditional view that the improvement in attitudes associated with repeated stimulus exposure arises largely from increasing stimulus familiarity. In order to investigate how well their one-factor model, as well as the authors' own two-factor model, actually fits the data, a series of linear structural equation analyses was carried out. The results revealed that the two-factor model provided a significantly better degree of fit. Moreland and Zajonc's original conclusions thus received additional support—stimulus recognition still does not appear to be a necessary condition for the occurrence of exposure effects.

Recently, we reported on the results of two experiments that sought to determine whether stimulus recognition was a necessary condition for the occurrence of exposure effects (Moreland & Zajonc, 1977). Our analyses of the data from those two experiments suggested that it was not. Birnbaum and Mellers (1979b), however, have since offered an alternative model for our data. Their model takes the following form,



in which familiarity and liking ratings are viewed as imperfect measures (with uncorrelated errors) of "subjective recognition"—a hypothetical mediating factor that depends on the frequency of stimulus exposure.

Birnbaum and Mellers evaluated their model by examining the pattern of correlations among the three major variables (frequency of stimulus exposure, liking, and subjective familiarity) in each of our experiments. In both cases, the correlations were consistent with a one-factor model of the data. Birnbaum and Mellers concluded, therefore, that stimulus recognition may have indeed mediated the observed exposure effects.

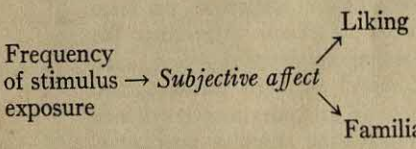
This new theoretical model provides a plausible explanation for our results. There are at least two important issues that must be considered, however, before any final conclusions can be drawn about the role of stimulus recognition in producing exposure effects. The first of those issues involves the nature of the proposed mediating factor. Birnbaum and Mellers (1979b) showed that a one-factor model was consistent with our data, and in accordance with the null hypothesis being tested, they called that factor "subjective recognition." It should be noted, however, that the true nature of such a factor remains an open issue. All that Birnbaum and Mellers really showed was that *some* single factor *could* have had an impact on our results. They

This research was supported by National Science Foundation Grant GS-35495. The authors would like to thank Frank Andrews, David Kenny, Laura Klem, and Henry Law for their advice on statistical issues, and Mary Grace Moore for her help with the data analyses.

Requests for reprints should be sent to Richard L. Moreland, Department of Psychology, University of Pittsburgh, Pittsburgh, Pennsylvania 15260.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3706-1085$00.75

did not prove, nor did they claim to prove, that those results were in fact influenced by stimulus recognition. It could be argued, for example, that the exposure effects observed in our experiments were actually mediated by feelings of "subjective affect," as indicated in the model shown below:



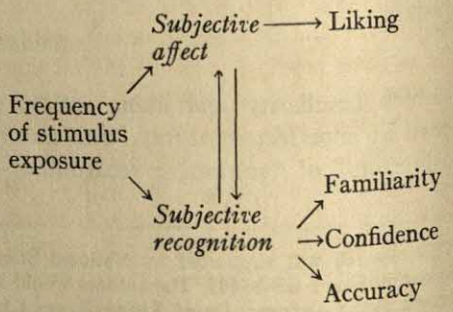
This affective model, which we described briefly in our earlier article, does not require any conscious form of stimulus recognition for the occurrence of exposure effects. Nevertheless, it is a model that Birnbaum and Mellers would have to accept as a possible explanation for our results, since it is in complete accordance with their own analyses of the data. A valid decision regarding the nature of their hypothetical mediating factor cannot be made on the basis of the available evidence. In particular, these is no good empirical reason to believe that the exposure effects observed in our two experiments were mediated by stimulus recognition.

A second important issue, quite distinct from the first, involves the ability of *any* one-factor model to account for our results. Birnbaum and Mellers showed that their own model was at least consistent with our data. That does not mean, however, that their model must necessarily be accepted as the best explanation for our results, or that more complex theoretical models must now be rejected on any grounds other than parsimony. Clearly, it would be both interesting and informative to specify several potential models of exposure effects and then investigate the degree to which each of them actually fits the observed data.

In order to examine our data more closely, we decided to perform some linear structural equation analyses. Unlike sociologists, econometricians, and political scientists, relatively few experimental social psychologists have taken advantage of these powerful analytical techniques, which can be used to evaluate a variety of fairly complex causal models within a single data set (cf. Jöreskog & Sörbom, 1977; Long, 1976). We used the LISREL III program (Jöreskog & Sörbom, 1976) to calculate maximum likelihood estimates for several causal models of our data. This program distinguishes between latent variables (constructs) and their observed indicators (measures), and allows the researcher to specify particular causal models for the variables under investigation. By estimating any unknown coefficients in the system of linear structural equations for a particular model, the program reveals the pattern of relations among its latent variables, differentiating causal effects from unexplained variance in each case. An unstandardized maximum likelihood solution, as well as a standardized solution, is generated for each causal model that is specified.

Using the LISREL III program, we were able to evaluate the degree to which several different models fit our data. For the present purposes, we will only describe the results of analyses performed on the data from our second experiment.[1] The program's input for those analyses was the matrix of intercorrelations among five variables: frequency of stimulus exposure, liking, subjective familiarity, recognition confidence, and recognition accuracy. The actual correlations are shown in Table 1.

Among the many models we tested, two were especially interesting. Our preferred model for the data, illustrated below, involved two major factors:



[1] Similar results were found in the data from both experiments. The data from the second experiment were more suitable for LISREL III analyses, however, because (a) more subjects were tested, (b) separate measures of the various recognition variables were obtained, and (c) recognition and liking measures were obtained from the same individuals.

Table 1
Matrix of Intercorrelations for Experiment 2

Variable	1	2	3	4	5
1. Frequency of stimulus exposure	1.00				
2. Liking	.663	1.00			
3. Subjective familiarity	.579	.533	1.00		
4. Recognition confidence	.340	.252	.291	1.00	
5. Recognition accuracy	.413	.302	.555	.417	1.00

Note. Frequency of stimulus exposure is the log plus one of the actual exposure frequency. All of these correlations were calculated using the residual data set (cf. Moreland & Zajonc, 1977, p. 195).

Briefly, we felt that repeated exposure to a novel stimulus could produce feelings of *both* recognition and affect, each of which might in turn have some influence of its own on the other. Liking ratings were assumed to be primarily affective in nature, while responses on the various recognition measures were thought to share a more cognitive origin.

Applying this two-factor model to the data from our second experiment, we obtained the standardized solution shown in Figure 1. Latent variables are shown there in ellipses, while indicators of those variables are shown in boxes. The path coefficients linking ellipses with boxes represent the estimated validities with which the latent variables were measured by their respective predictors. Path coefficients linking latent variables to each other represent the estimated strength of particular causal relationships. Unexplained variance in the latent variables (V_1 and V_2) and measurement error in their indicators (E_1 to E_5) are also estimated. Note that a few of the model's parameters (shown in parentheses) were set equal to some a priori value in the maximum likelihood solution, so that the variance in all of the latent variables could be identified.[2]

The results provided some interesting information about our data. The chain of paths linking stimulus exposure to subjective recognition, and subjective recognition to subjective affect, indicated that some of the exposure effects observed in our research may have indeed been mediated by subjective recognition. However, there was also a relatively strong and direct path linking stimulus exposure to subjective affect, suggesting that subjective recognition was not an essential factor in producing our results. The overall goodness of fit for our model was $\chi^2(5) = 38.99$.[3]

The standardized solution that we obtained when Birnbaum and Mellers' (1979b) one-factor model was applied to the same data set

[2] Since stimulus exposure and subjective affect each had only one indicator, the program could not successfully identify their variance. For that reason, we specified (in the maximum likelihood solution) both the validity and the error coefficients for those two indicators. Frequency of stimulus exposure was assumed to have a validity of 1.00 (and an error of .00), as it represented a presumably perfect experimental manipulation of stimulus exposure. The validity coefficient for liking was set equal to .87 (and its error coefficient to .24) on the basis of a reliability estimate obtained in our original research (cf. Moreland & Zajonc, 1977, pp. 195 and 197). In order for the program to identify the variance in subjective recognition, we followed an accepted standard practice, and set the validity coefficient for subjective familiarity to 1.00, again in the maximum likelihood solution. It should be noted that parameters whose values were specified in the maximum likelihood solution sometimes took on slightly different values in the standardized solution. These differences occurred whenever the estimated variability of a latent variable was less than 1.00, and were a natural by-product of the standardization procedure (cf. Jöreskog & Sörbom, 1976, p. 24).

We also specified that the two path coefficients between subjective affect and subjective recognition had to be equal to each other in the maximum likelihood solution. This was equivalent to specifying the value of a single parameter, and therefore added a single degree of freedom to the chi-square. Without this additional constraint, which is often adopted when testing models involving reciprocal causation (cf. Wiley, 1973; Wright, 1960), our model would have been underidentified.

[3] The degrees of freedom for this chi-square are the number of cells in the correlation matrix (including the diagonals) minus the number of "free" parameters. The chi-square measuring goodness of fit was fairly large for all of the models that we tested. Jöreskog (1969) has suggested, however, that the chi-square statistic may not, in itself, provide the most useful information about how well a model

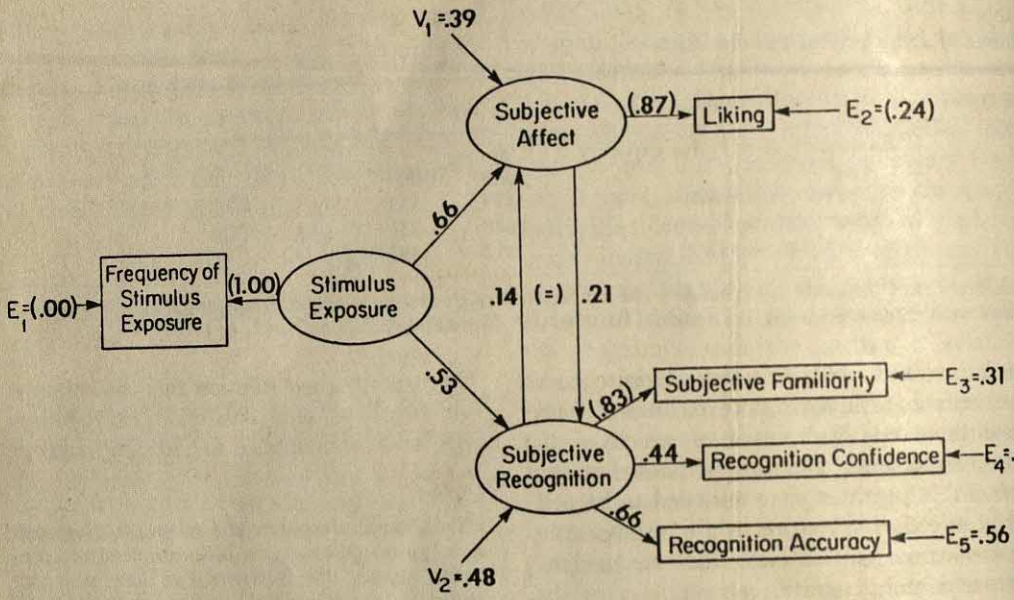


Figure 1. Moreland and Zajonc's two-factor model of exposure effects (standardized solution). (V = variable; E = error.)

is illustrated in Figure 2. The same parameter constraints described earlier were again used. In general, the path coefficients for this model seemed reasonable, although it was somewhat surprising that liking measured subjective recognition as well as subjective familiarity did, and even better than recognition confidence or recognition accuracy were able to do.[4] If our data were indeed influenced by subjective recognition, as Birnbaum and Mellers' model indicates, then variables meant to measure recognition processes should have been most closely associated with that factor. The overall goodness of fit for Birnbaum and Mellers' model was $\chi^2(6) = 83.58$.

To compare the accuracy of these two models, we first took Jöreskog's (1969) suggestion and looked to see which one had the lower ratio between its chi-square and degrees of freedom. According to that criterion, our own model of the data (38.99:5 = 7.80) was clearly better than the one proposed by Birnbaum and Mellers (83.58:6 = 13.93). Since their model was actually nested within ours, however, a statistical test could be made by subtracting the chi-square and degrees of freedom for one model from those of the other (cf. Long, 1976). The results showed that our model provided a significantly better fit to the data, $\chi^2(1) = 44.59$, $p < .01$.

In summary, we see no reason to alter our earlier conclusions about the role of stimulus

fits the data. Instead, he recommends looking at the ratio between the chi-square and its degrees of freedom as another means of evaluating the degree of fit (cf. Jöreskog, 1969, p. 201).

[4] In the accompanying article, Birnbaum and Mellers (1979a) have argued that at least two of the recognition measures (confidence and accuracy) were related to stimulus exposure in a nonlinear manner and thus could not be expected to load more heavily on subjective recognition. In order to test this argument, we repeated the two LISREL III analyses reported here, but allowed for the possibility of correlated errors between recognition confidence and recognition accuracy. This was accomplished by introducing a new latent variable into each analysis, one that was unrelated to any of the other latent variables in the model and was measured only by the confidence and accuracy of the subjects' recognition responses. Both of the revised models fit the data better than their predecessors did, indicating that shared nonlinearity among the recognition measures was indeed a factor in our research. Nevertheless, even after the possibility of correlated errors between recognition confidence and recognition accuracy was taken into account, our own model of the data was still more accurate than that proposed by Birnbaum and Mellers.

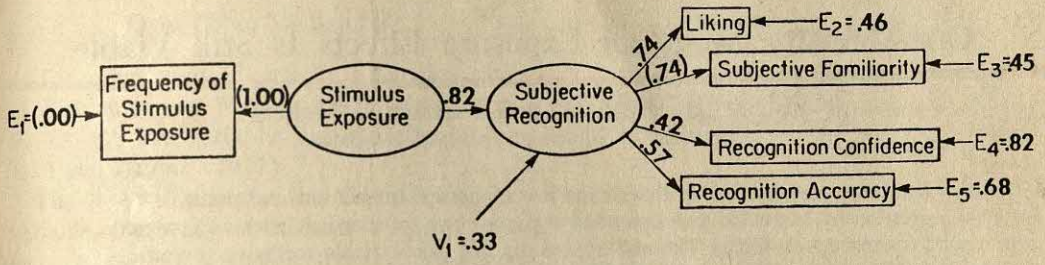


Figure 2. Birnbaum and Mellers' one-factor model of exposure effects (standardized solution). (V = variable; E = error.)

recognition in producing exposure effects. Birnbaum and Mellers (1979b) proved that a one-factor model of our data cannot be rejected. The nature of their hypothetical factor, however, remains a mystery—in particular, there is no evidence to show that it necessarily involved any form of stimulus recognition. Moreover, the fact that a one-factor model (of any sort) was consistent with our data says little about the relative ability of such a model to account for our results. In fact, when we reanalyzed our data using linear structural equation analyses, the results showed that our own two-factor model provided a significantly better fit. This finding, along with the results of other research (cf. Matlin, 1971; Wilson, 1975, 1979), strongly suggests that stimulus recognition is not a necessary condition for the occurrence of exposure effects.

References

Birnbaum, M. H., & Mellers, B. A. One-mediator model of exposure effects is still viable. *Journal of Personality and Social Psychology,* 1979, *37,* 1090–1096. (a)

Birnbaum, M. H., & Mellers, B. A. Stimulus recognition may mediate exposure effects. *Journal of Personality and Social Psychology,* 1979, *37,* 391–394. (b)

Jöreskog, K. G. A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika,* 1969, *34,* 183–202.

Jöreskog, K. G., & Sörbom, D. *LISREL III: Estimation of linear structural equation systems by maximum likelihood methods.* Chicago: National Educational Resources, 1976.

Jöreskog, K. G., & Sörbom, D. Statistical models and methods for analysis of longitudinal data. In D. J. Aigner & A. S. Goldberger (Eds.), *Latent variables in socioeconomic models.* Amsterdam: North Holland, 1977.

Long, J. S. Estimation and hypothesis testing in linear models containing measurement error. *Sociological Methods and Research,* 1976, *5,* 157–203.

Matlin, M. W. Response competition, recognition, and affect. *Journal of Personality and Social Psychology,* 1971, *19,* 295–300.

Moreland, R. L., & Zajonc, R. B. Is stimulus recognition a necessary condition for the occurrence of exposure effects? *Journal of Personality and Social Psychology,* 1977, *35,* 191–199.

Wiley, D. E. The identification problem for structural equation models with unmeasured variables. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences.* New York: Seminar Press, 1973.

Wilson, W. R. *The unobtrusive induction of positive attitudes.* Unpublished doctoral dissertation, University of Michigan, 1975.

Wilson, W. R. Feeling more than we can know: Exposure effects without learning. *Journal of Personality and Social Psychology,* 1979, *37,* 811–821.

Wright, S. The treatment of reciprocal interaction, with or without lag, in path analysis. *Biometrica,* 1960, *16,* 423–445.

Received June 16, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 6, 1090–1096

One-Mediator Model of Exposure Effects Is Still Viable

Michael H. Birnbaum and Barbara A. Mellers
University of Illinois at Urbana-Champaign

Birnbaum and Mellers criticized the use of partial correlation and multiple regression by Moreland and Zajonc to argue for two independent effects of stimulus exposure on liking. The null hypothesis that one variable mediates the effect of the independent variable on the dependent variables was not tested by their analyses. In response, Moreland and Zajonc reanalyzed their data, using structural equations analysis, and replied that there is evidence to support their previous conclusions. However, the present article shows that the small residuals from the one-mediator model may be due to shared nonlinearity (correlated errors) in three of the dependent variables. This simpler interpretation achieves a better fit to the Moreland and Zajonc data than the two-mediator model they advocated. Since the null hypothesis of one mediator is still viable, the burden of proof rests on those who contend that there is more to the exposure effect than stimulus recognition.

Moreland and Zajonc (1977) replicated the exposure effect, that is, the finding that stimuli that are presented with greater frequency will be rated more favorably. They also asked their subjects to rate their familiarity with the stimuli, and they found a significant partial correlation between liking and exposure frequency when rated familiarity was partialed out. This partial correlation (and related regression analyses) led them to conclude that there is an additional, "independent" effect of exposure frequency on liking that is not mediated by stimulus recognition.

Birnbaum and Mellers (1979) questioned the use of partial correlation and regression in this argument. They pointed out that the null hypothesis that a single variable (e.g., recognition) mediates the effect of the independent variable on *both* dependent variables predicts that both partials should have the same sign as the original correlations. Only if the dependent variable measuring recognition is assumed to be perfectly correlated with the mediator does the regression analysis test the null hypothesis. Otherwise, lack of perfect validity of the dependent variable vitiates the analysis.

Birnbaum and Mellers (1979) presented path models both for the null hypothesis of one mediator and for the alternative hypothesis that a second mediator is required. Implications of the models and methods to distinguish them were described. These methods were applied to the three major variables in Moreland and Zajonc (1977) with the result that no evidence was found to require the hypothesis of two mediators over the simpler hypothesis of one.

In response to these arguments, Moreland and Zajonc (1979) reanalyzed their data, using structural equations analysis. They argued that when all five variables are considered, a case can be made for the existence of an extra path from stimulus frequency to liking, in addition to the path via subjective recognition. However, the present article will show that the residuals from the one-mediator model, fit to all five variables, are small in magnitude and are not of the form anticipated by the Moreland and Zajonc model. Instead, they can be explained by the simpler hypothesis that some of the dependent variables are nonlinearly related to the single mediator.

Requests for reprints should be sent to Michael H. Birnbaum, Department of Psychology, University of Illinois, Champaign, Illinois 61820.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3706-1090$00.75

One or Two Mediators?

The diagram in Figure 1 is an extension of Birnbaum and Mellers (1979, Expression 9) for all five variables reported by Moreland and Zajonc (1977).

The diagram indicates that the independent variable, frequency of stimulus exposure, affects a mediator, "subjective recognition," which in turn influences the four dependent variables, rated affect (liking), rated familiarity, rated recognition confidence, and recognition accuracy. Moreland and Zajonc (1977, 1979) argue that there is an additional, independent effect of stimulus exposure on liking, presumably via a second mediator, labeled "subjective affect" in Figure 1.

The one-mediator model is a special case of Figure 1 with $b = 0$. The model favored by Moreland and Zajonc (1979, Figure 1), although written in more complex form, is also a special case of Figure 1, with $c_1 = c_2 = c_3 = 0$.

One-Mediator Model

Table 1 shows the residual correlations that remain after fitting a linear one-mediator model ($b = 0$, $c_1 = c_2 = c_3 = 0$) to the correlations for both experiments of Moreland and Zajonc (1977; see Table 1 of Birnbaum & Mellers, 1979). Models were fit to maximum likelihood criterion by means of ACOVS and LISREL III (Jöreskog, 1974; Jöreskog, Gruvaeus, & van Thillo, 1970; Jöreskog & Sörbom, 1976). In general, the residual correlations in Table 1 are small in magnitude. In particular, the largest deviation does not occur for the residual correlation between liking (affect) and frequency (only .03 and .06 for Experiments 1 and 2), which might be expected to be large if there is an extra causal path from frequency to liking. Instead, the largest residual for both experiments is for the correlation between recognition accuracy and recognition confidence (.08 and .18).

One can reasonably ask if the residual correlations in Table 1, which (for Experiment 2) are based on 10 ratings from each of 40 subjects, are of sufficient magnitude to reject the one-mediator model. Moreland and Zajonc (1979) contend that the lack of fit for this

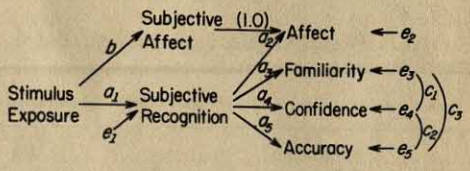


Figure 1. Path diagram representing theories of stimulus exposure. (The coefficient a_1 represents the correlation between stimulus exposure and a mediator, labeled "subjective recognition," which in turn is correlated with the four dependent variables with coefficients of a_2, a_3, a_4, and a_5. The errors, e_1 through e_5 represent residuals, analogous to uniquenesses in factor analysis. The coefficient b represents the effect of stimulus exposure on liking apart from its effect via the first mediator. The correlations c_1, c_2, and c_3 allow for correlated errors, which could be produced by shared nonlinearity of the dependent variables measuring recognition. The issue at hand is whether the data of Moreland and Zajonc (1977) provide sound evidence to reject the null hypothesis that $b = 0$.)

model is evidence of a second mediator. It will be shown below that this pattern of small residual correlations can be better fit by a one-mediator model, which allows the errors in the recognition measures to be correlated (i.e., the cs in Figure 1 are not fixed to zero). Shared nonlinearity of the recognition measures would predict correlated errors.

Nonlinearity in Moreland and Zajonc Data

Figure 2A plots the means of the four dependent variables as a function of the independent variable, log stimulus frequency, as reported in Table 3 of Moreland and Zajonc (1977). The ordinate has been recalibrated linearly so that all four dependent variables can be shown simultaneously. The one-mediator model and the two-mediator model both imply that the curves should all be linear. Should two variables share common nonlinearity, the residual correlation in Table 1 is expected to be positive. All three recognition measures appear to be nonlinearly related to frequency of stimulus exposure. Furthermore, recognition confidence and recognition accuracy appear to share a common cubic trend in relation to the other curves. A similar pattern was evident for Experiment 1. Thus, the two variables that show the largest residual from the one-mediator model (Table

Table 1
Residuals From One-Mediator Linear Model

	Frequency	Liking	Familiarity	Confidence	Accuracy
Frequency		.03	.01	.02	−.07
Affect	.06		−.01	.02	−.06
Familiarity	−.03	−.02		−.06	.04
Confidence	−.01	−.06	−.02		.08
Accuracy	−.05	−.11	.13	.18	

Note. Each entry is the residual correlation after fitting a one-mediator model to the correlations obtained by Moreland and Zajonc (1977). Values above the diagonal are for Experiment 1; values below the diagonal are for Experiment 2. Liking and affect are corresponding measures for Experiments 1 and 2, respectively. The question under consideration is whether correlations of this pattern and magnitude, based on 10 judgments from each of 40 subjects (Experiment 2), warrant rejection of the one-mediator model. Note that the largest residuals involve accuracy, which is nonlinearly related to frequency (Figure 2).

1) are the two that are most nonlinearly related to the others (Figure 2A).

Nonlinearity Can Produce Correlated Residuals

To illustrate how shared nonlinearity can produce deviations from a one-mediator model, a hypothetical data set was constructed from the following one-mediator model:

$$\begin{aligned} F &= X + e_1; \\ Y_1 &= F + e_2; \\ Y_2 &= F + .33F^2 + e_3; \\ Y_3 &= F + .33F^2 + e_4; \\ Y_4 &= F + .33F^2 + e_5; \end{aligned} \qquad (1)$$

where X is the manipulated independent variable (analogous to stimulus exposure in Figure 1); Y_1 through Y_4 are the measured, dependent variables; F is the mediating factor (analogous to subjective recognition); the addition of $.33F^2$ produces nonlinear relationships between Y_2 through Y_4 and F; and e_1 through e_5 are mutually uncorrelated error terms. The value .33 for the coefficient of F^2 was chosen so that the degree of nonlinearity would be roughly comparable to that in the data of Moreland and Zajonc (1977), as in Figure 2A. To generate the 160 hypothetical cases, 5 values of X, (−2, −1, 0, 1, 2) were factorially combined with 2 values of e_1 (−1, 1) to produce 10 values of F. These values of F were factorially combined with two levels of each error ($e_2 = -1.1, 1.1$; $e_3 = -4, 4$; $e_4 = -3, 3$; $e_5 = -2, 2$) to generate the values of the dependent variables, Y_1, Y_2, Y_3, and Y_4, according to Equation 1. The mean values for Y for each level of X are shown in Figure 2B, plotted for comparison with Figure 2A. For the hypothetical example, variables Y_2, Y_3, and Y_4 share a common nonlinear (quadratic) relationship with the other two variables. If the data were generated without the F^2 terms, the curves would be linear; therefore, the linear one-mediator model (with $b = c_1 = c_2 = c_3$ fixed to zero in Figure 1) would fit perfectly.

The correlation matrix, generated from the

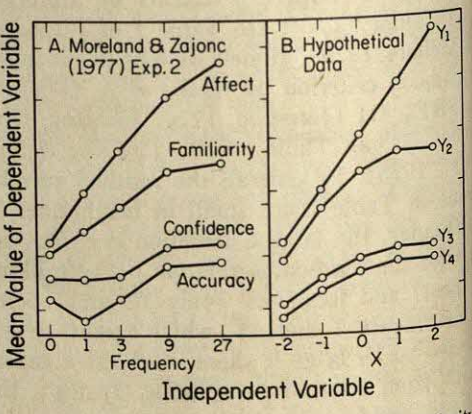


Figure 2. A. Mean values of affect, rated familiarity, recognition confidence, and recognition accuracy as a function of exposure frequency [log (f + 1)]. (From Moreland & Zajonc, 1977, Table 3. Ordinate separately calibrated for each variable.) B. Mean values of dependent variables, Y_1, Y_2, Y_3, and Y_4, as a function of independent variable. (Hypothetical data were generated from a one-mediator model with Y_2, Y_3, and Y_4 nonlinearly related to the mediating variable, as shown in Equation 1.)

one-mediator model of Equation 1 (with some dependent variables nonlinearly related to the mediator), is shown in the upper portion of Table 2. The correlation matrix was fit by means of LISREL to a linear one-mediator model (b, c_1, c_2, and c_3 in Figure 1 were fixed to zero). The residual correlations are shown below the diagonal in Table 2. Note that the largest discrepancies in Table 2 are among variables Y_2, Y_3, and Y_4, which have nonlinearity (i.e., quadratic term, F^2, in Equation 1) in common. The residual correlation between X and Y_1 is also positive, since the best-fit single factor falls between F and F^2. This pattern of residuals is similar to that in Table 1.

A one-mediator model can be fit to the hypothetical data perfectly if the residuals for Y_2, Y_3, and Y_4 are allowed to be correlated. These correlations represent the shared nonlinearity of these variables. The model in Figure 1 thus fits the hypothetical data perfectly with b set to zero, by allowing correlations c_1, c_2, and c_3 to be nonzero (the values of c can be further constrained to satisfy Equation 10 of Birnbaum & Mellers, 1979).

This example illustrates that even when data are constructed from a single-mediator model, if some variables are nonlinearly related, deviations from the model can appear unless the model allows correlations among the residuals for nonlinear variables.

Table 2
Hypothetical Example for One-Mediator Nonlinear Model

	X	Y_1	Y_2	Y_3	Y_4
X		.69	.32	.39	.49
Y_1	.02		.32	.40	.51
Y_2	−.02	−.02		.26	.37
Y_3	−.03	−.02	.04		.40
Y_4	−.02	−.02	.06	.08	

Note. X = independent variable; Y_1 through Y_4 = dependent variables. Correlations are above diagonal; residuals from one-mediator linear model are below diagonal. Note similarity of residuals to those in Table 1.

Correlated Residual Model Fits Moreland and Zajonc Data

The model that fits the hypothetical data (Figure 2B) also achieves a good fit to the Moreland and Zajonc (1977) data (Figure 2A). It is a special case of a one-mediator model with nonzero correlations among the errors in the recognition variables. The correlations among the residuals (c_1, c_2, and c_3) may be given a response-bias, or "method," interpretation, since this model does not postulate any additional paths from stimulus frequency to any of the variables, besides the one via the recognition mediator (b is set to zero). The residual correlations from this one mediator model are quite small, the largest deviation being .04 for both experiments. The residual correlations between exposure frequency and liking are −.02 and .00 for Experiments 1 and 2, respectively.

Thus, the residual correlations from the linear one-mediator model may be attributed to shared nonlinearity of the recognition measures. This interpretation is consistent with the data shown in Figure 2A, with the pattern of residuals in Table 1 (in which the largest deviations involve recognition accuracy), and with the fact that the one-mediator model with correlated errors provides a good fit to the correlations. In sum, the one-mediator model remains a viable interpretation of the Moreland and Zajonc (1977) data if it is allowed that the recognition measures are nonlinearly related to the other variables.

Two-Mediator Model Not Required

The two-mediator model favored by Moreland and Zajonc (1979, Figure 1) yields the theoretical correlation matrix shown in the lower triangle of Table 3. The model is equivalent to Figure 1 of the present article with $c_1 = c_2 = c_3$ set to zero. Table 3 shows that if $b = 0$, each correlation can be represented as the product of correlations with the mediating variable, subjective recognition. However, with $b > 0$, the correlations between Y_1 (liking) and the others increase, especially the correlation between X and Y_1.

The hypothetical correlations generated from a two-mediator model in the upper portion of Table 3 show that in principle, it is possible to distinguish the two-mediator model from the one-mediator model. Note that the

Table 3
Theoretical Values for Two-Mediator Model and Hypothetical Example

	X	Y_1	Y_2	Y_3	Y_4
X		.75	.40	.40	.40
Y_1	$a_1a_2 + b$		.60	.60	.60
Y_2	a_1a_3	$a_2a_3 + ba_1a_3$		.64	.64
Y_3	a_1a_4	$a_2a_4 + ba_1a_4$	a_3a_4		.64
Y_4	a_1a_5	$a_2a_5 + ba_1a_5$	a_3a_5	a_4a_5	

Note. X = independent variable; Y_1 through Y_4 = dependent variables. Theoretical values are below diagonal. Hypothetical values above diagonal are based on $a_1 = a_2 = .5$, $a_3 = a_4 = a_5 = .8$, and $b = .5$.

hypothetical correlations of frequency, liking, and familiarity violate Equation 10 of Birnbaum and Mellers (1979), since $.75/.40 = 1.88$, which exceeds the limit set by $1/.60 = 1.67$.[1]

Table 4 shows parameter estimates and an index of fit for the various models under consideration, applied to data for Experiment 2 of Moreland and Zajonc (1977). To obtain the predicted correlation between two variables, multiply the coefficients along each path and sum these products over all distinct paths connecting the two variables in Figure 1. For example, for the one-mediator model, the predicted correlation between stimulus exposure and rated affect is a_1a_2, or .61. For the two-mediator model, the predicted value would be $a_1a_2 + b$, or .66.

The two-mediator model shown in Figure 1 of Moreland and Zajonc (1979) is equivalent to the model in Figure 1, with $c_1 = c_2 = c_3 = 0$. Because the models are mathematically identical, the index of fit in Table 4 ("χ^2") is the same as reported by Moreland and Zajonc (1979).[2] The largest residual for the two-mediator model for Experiment 2 is .13, between recognition confidence and recognition accuracy, a slight improvement over the corresponding value of .18 for the one-mediator model.[3]

The one-mediator model with correlated errors provides a better fit to the data than the two-mediator model, although it does not postulate an additional path from exposure to liking. The largest residual for this model is only .04 for both experiments. The value of χ^2 is less than one sixth as large as that for the two-mediator model. In addition, permitting b to be nonzero does not improve the fit, once correlations are allowed among the errors in the recognition measures (see last column of Table 4).

In sum, the analyses provide no compelling reason to reject the null hypothesis that b in Figure 1 is zero. Therefore, the data do not demonstrate the existence of another path from stimulus exposure to liking apart from

[1] The numerical example in Table 3 also illustrates that the two-mediator model predicts a larger correlation between X and Y_1 than would be predicted from the one-mediator model. According to the one-mediator model

$$\hat{\rho}_{XY_1} = \frac{(a_1a_3)(a_2a_3)(a_4a_5)}{(a_3a_4)(a_3a_5)} = \frac{\rho_{XY_2}\rho_{Y_1Y_2}\rho_{Y_3Y_4}}{\rho_{Y_2Y_3}\rho_{Y_2Y_4}}, \quad (2)$$

which reduces to a_1a_2. In contrast, the two-mediator theory with $b > 0$ predicts that the observed correlation should exceed the value of $\hat{\rho}_{XY_1}$ from Equation 2. For the values used in the hypothetical example of Table 3, Equation 2 (the one-mediator model) yields $\hat{\rho}_{XY_1} = (.40)(.60)(.64)/(.64)(.64) = .375$, which is only half as large as the table (two-mediator model) value of $\rho_{XY_1} = .75$. When Equation 2 is applied to the correlations of Moreland and Zajonc (1977), the values of $\hat{\rho}_{XY_1}$ are .48 and .81 for Experiments 1 and 2. The obtained values, which by the foregoing should have been *larger* than these values of $\hat{\rho}_{XY_1}$ according to the two-mediator theory, were only .42 and .66.

[2] The model comparisons presented by Moreland and Zajonc (1979) are based on chi-square values that may have been inflated by the assumption that by partialing out the main effect of subjects, the sample of 40 subjects produced 400 independent observations. However, repeated observations from the same subjects should not be treated as if they were independent. Since the computed chi-square is directly proportional to the assumed number of independent observations in each correlation, the "χ^2" values reported should not be used for statistical inferences. The "χ^2" values reported here are calculated using the same value (400) for the sample size, only to provide an index that can be compared with the accompanying article.

[3] Moreland and Zajonc (1979) fixed the coefficient of familiarity in both of their analyses even though it was actually estimated from their data (see their Footnote 2). As a consequence of this procedure the correct *df* for both models are 1 less than they report.

Table 4
Estimates of Parameters for Model Comparisons

Variable	Model			
	One mediator	Two mediator	One mediator correlated errors	Two mediator correlated errors
a_1	.82	.69	.86	.86
a_2	.74	.26	.77	.72
a_3	.74	.83	.68	.68
a_4	.42	.44	.37	.38
a_5	.57	.66	.45	.45
b	(0)	.48	(0)	.04
c_1	(0)	(0)	.04	.03
c_2	(0)	(0)	.25	.25
c_3	(0)	(0)	.25	.24
df	5	4	2	1
"χ^2"	83.58	38.99	6.22	6.22

Note. Based on data of Experiment 2 of Moreland and Zajonc (1977). Each entry represents estimated parameter as described in Figure 1, with restrictions imposed by the models. Values shown in parentheses are fixed. Degrees of freedom (df) are 10 minus the number of estimated parameters. The values of e_i are redundant; for example for the one-mediator model, e_i can be computed from $1 - a_i^2$. The "χ^2" is inflated and should not be used for statistical tests, but represents an index of fit that can be compared with values reported in the accompanying article.

an effect that could be mediated by stimulus recognition.[4]

It is always preferable to choose among theories on the basis of qualitative differences in predictions for a variety of experimental manipulations rather than on the basis of small differences in an index of fit. Perhaps evidence for a second mediator would occur if a new variable could be found that could potentially reverse the exposure effect, so that repeated exposure could either increase or decrease liking (while increasing recognition), depending on the value of the new variable. A future study may provide sound evidence that stimulus recognition does not mediate the exposure effect. Until such evidence is presented, however, it is premature to assert that there are two independent effects of stimulus exposure on liking.

Conclusions

These analyses do not disprove the possible existence of unconscious affect, learning without awareness, or subception. They do not disprove the favored interpretation of Moreland and Zajonc (1979), though the one-mediator model with correlated errors fits better than the two-mediator model without correlated errors (Table 4). However, the analyses do show that the null hypothesis of one mediator remains a viable description of the data of Moreland and Zajonc (1977), who argued that there are two independent effects. Perhaps the null hypothesis should be favored by skeptics and the burden of proof be laid upon those who would refute it.

[4] Moreland and Zajonc (1979) argued that if a single mediator explains their (1977) data, the mediator may be named "affect" rather than "recognition." The name of a mediator is a matter of definition that cannot be refuted by experiment. However, any discussion of the proper label for a single mediator is tangential to the original argument of Moreland and Zajonc (1977, p. 193). In that article they stated that evidence of two independent effects of exposure on liking was required to demonstrate that stimulus recognition is not necessary to the exposure effect.

References

Birnbaum, M. H., & Mellers, B. A. Stimulus recognition may mediate exposure effects. *Journal of Personality and Social Psychology*, 1979, *37*, 391–394.

Jöreskog, K. G. Analyzing psychological data by structural analysis of covariance matrices. In D

H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 2). San Francisco: Freeman, 1974.

Jöreskog, K. G., Gruvaeus, G. T., & van Thillo, M. *ACOVS—A general computer program for analysis of covariance structures. Research Bulletin 70–15*. Princeton, N.J.: Educational Testing Service, 1970.

Jöreskog, K. G., & Sörbom, D. *LISREL III: Estimation of linear structural equation systems by maximum likelihood methods*. Chicago: National Educational Resources, 1976.

Moreland, R. L., & Zajonc, R. B. Is stimulus recognition a necessary condition for the occurrence of exposure effects? *Journal of Personality and Social Psychology*, 1977, *35*, 191–199.

Moreland, R. L., & Zajonc, R. B. Exposure effects may not depend on stimulus recognition. *Journal of Personality and Social Psychology*, 1979, *37*, 1085–1089.

Received December 27, 1978 ■

Ad Hoc Reviewers

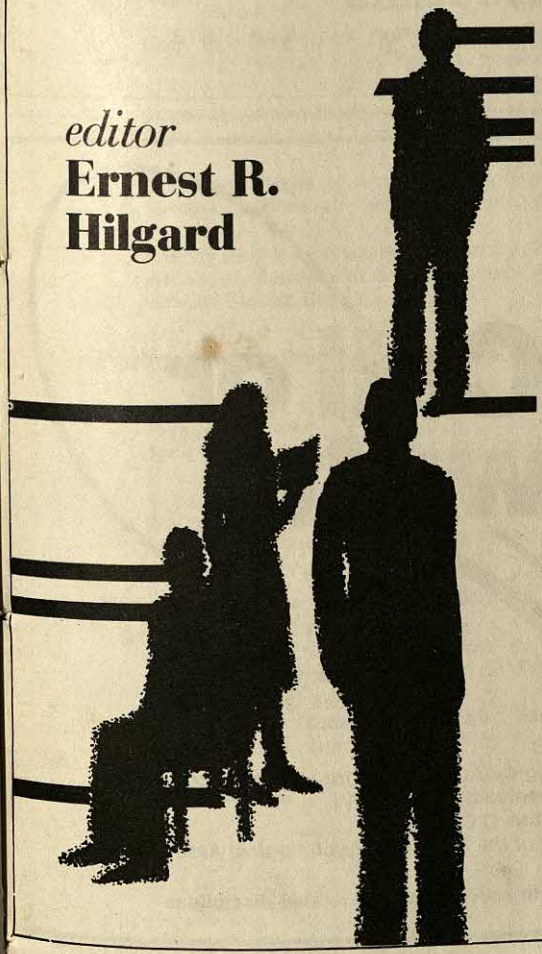
The Editor acknowledges with thanks the advice and consultation provided by the following persons, who served recently as ad hoc reviewers of manuscripts. (The number of manuscripts reviewed is in parentheses, if more than one.)

Robert Abelson
John R. Anderson
Norman Anderson
Robert S. Baron
Sandra L. Bem
Michael Birnbaum (4)
Marilyn Brewer
Philip Brickman (2)
Ross W. Buck (2)
Lynn E. Bush
Allan Buss (2)
John Carroll (2)
David W. Chan
George Cicala
Victor Cline
Harris Cooper
James H. Davis
Kay Deaux
Ebbe Ebbesen
Susan Eisen (2)
Robert Folger
Irene Frieze
David Funder
G. Gerbner
Rick Gibbons (2)
Wallace Hall
Richard J. Harris (3)
Elaine Hatfield
Chester Insko
Carroll Izard
Robert L. Kahn
David Kenny
Vladimir Konečni (2)
Gerald Koocher
Robert Kraut
Arie Kruglanski
Michael Kuhlman
Lawrence Littig
Guy E. Lometti
Jacqueline Macaulay
Neil Malamuth
Melvin Manis (2)
Richard E. Petty (2)
Dean Pruitt
Dennis Regan (2)
Robert Rosenthal
Michael Scheier
Brendan Rule
Barry Schlenker
Harold Sigall
Dean K. Simonton
Shelley Taylor
Harry C. Triandis
Gary Wells (3)
Shirley Weitz
Paul Werner
Ladd Wheeler
Jerry Wiggins (2)
Timothy D. Wilson
Robert Zajonc

American Psychology in Historical Perspective

1892-1977

editor
Ernest R. Hilgard



Now available in one book — 21 APA presidential addresses in their entirety from some of the most important names in American psychology. Included are classic pieces from James, Cattell, Dewey, Thorndike, Woodworth, and Watson. You'll also find Harlow's "The Nature of Love" and Miller's "Analytical Studies of Drive and Reward."

This book is truly a milestone for historical psychology. It provides both a fascinating chronology of the presidents of the American Psychological Association from 1892 to 1977 and a valuable collection of significant essays.

This new APA publication traces the development of American psychology over four broad periods in psychology's history: the first 25 years (1892-1916), the years of the two world wars (1917-1945), the 20 years after World War II (1946-1967), and the recent past (1968-1977).

For each of these periods, the editor, Ernest R. Hilgard, provides a brief summary of the thinking in psychology at the time, biographies of all the APA presidents with abstracts of their presidential addresses, and the selected presidential addresses in full.

American Psychology in Historical Perspective *may be ordered in hard cover for $18 or in soft cover for $15 by writing to: American Psychological Association, Order Dept., 1200 Seventeenth Street, NW, Washington, D.C. 20036*

Please include full remittance for orders of $25 or less.

Sensitivity to Nonverbal Communication

The PONS Test

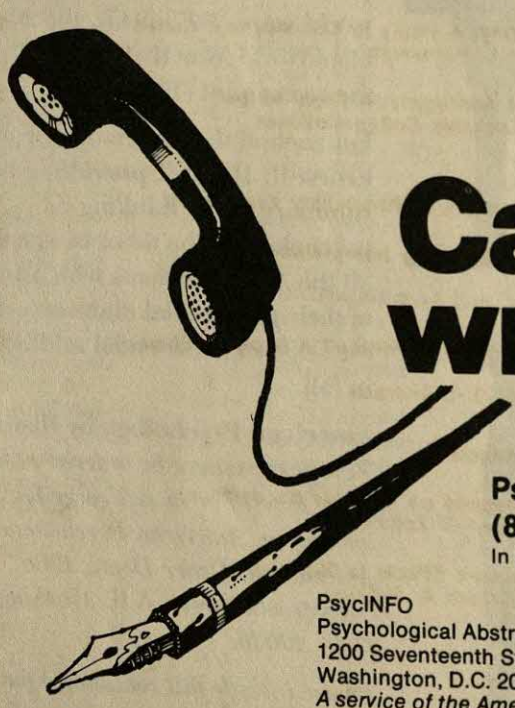
Robert Rosenthal, Judith A. Hall, M. Robin Di Matteo, Peter L. Rogers, and Dane Archer

The Profile of Nonverbal Sensitivity (PONS) is the first scientifically validated test for assessing the ability to decode nonverbal cues in various channels of communication. This book describes the PONS test and reports on the results of its widespread use. It has been administered to groups of children, teenagers, married couples, therapists, physicians, actors, mental patients, and others—a total of more than 7,000 people in a dozen different countries.

Based on their research with the PONS test, the test developers answer some new and some classic questions, including: Are women more accurate decoders than men? Are members of certain professions (e.g., actors, psychologists) exceptionally good at decoding emotions? Is nonverbal decoding correlated with general intelligence or cognitive skills? Are people accurate judges of their own nonverbal sensitivity? Does sensitivity predict success—for instance, are better decoders more popular or more effective in their work? $20.00

Johns Hopkins

The Johns Hopkins University Press, Baltimore, Maryland 21218



Call or write!

PsycINFO Action Line
(800) 336-4980
In Virginia call (202) 833-5908

PsycINFO
Psychological Abstracts Information Services
1200 Seventeenth Street, N.W.
Washington, D.C. 20036
A service of the American Psychological Association

PsycINFO, the database of record in psychology and related disciplines

(Continued from front cover)

Special Section on Interpretive Controversies in Personality and Social Psychology

Editor's Foreword 930

Analysis of Variance Models in Balance Theory

A Reply to Norman H. Anderson's Critique of the Subject–Verb–Object Approach to Social Cognition 931
Harry F. Gollob

Indeterminate Theory: Reply to Gollob 950
Norman H. Anderson

Cognitive Mediators of Emotion

Negative Emotional Biasing of Unexplained Arousal 953
Christina Maslach

Affective Consequences of Inadequately Explained Physiological Arousal 970
Gary D. Marshall and Philip G. Zimbardo

Comments on the Maslach and Marshall–Zimbardo Experiments 989
Stanley Schachter and Jerome E. Singer

Androgyny

Bem Sex Role Inventory: A Theoretical and Methodological Critique 996
Elazar J. Pedhazur and Toby J. Tetenbaum

Psychological Androgyny: A Case of Mistaken Identity? 1017
Anne Locksley and Mary Ellen Colten

The Many Faces of Androgyny: A Reply to Locksley and Colten 1032
Janet T. Spence and Robert L. Helmreich

Theory and Measurement of Androgyny: A Reply to the Pedhazur–Tetenbaum and Locksley–Colten Critiques 1047
Sandra Lipsitz Bem

Empirical Versus Conceptual Structure in Personality Assessment

How Relevant Is a Semantic Similarity Interpretation of Personality Ratings? 1055
Jack Block, Daniel S. Weiss, and Avril Thorne

Accurate Reflection or Systematic Distortion? A Reply to Block, Weiss, and Thorne 1075
Richard A. Shweder and Roy G. D'Andrade

Mediators of "Mere Exposure" Effects

Exposure Effects May Not Depend on Stimulus Recognition 1085
Richard L. Moreland and Robert B. Zajonc

One-Mediator Model of Exposure Effects Is Still Viable 1090
Michael H. Birnbaum and Barbara A. Mellers

Ad Hoc Reviewers 1096

Willo P. White, editor

Resources in Environment and Behavior

NEW FROM APA IN 1979

An invaluable sourcebook for students, instructors, and researchers in the new field of environment and behavior. This solid reference includes:

- Overview and history of this emerging field
- Graduate programs—both formal and informal
- Teaching innovations introduced in the United States, Canada, and Great Britain
- Funding sources
- Career opportunities
- Directory of key individuals currently working in the field
- Annotated bibliography
- Listing of relevant journals

Resources in Environment and Behavior is available from the American Psychological Association for $10 (soft cover only). To order, make your check payable to APA.

American Psychological Association
Order Department
1200 17th Street, NW
Washington, DC 20036

Please include full remittance for orders of $25 or less.

3/2/80 R.C

JOURNAL OF

Personality AND Social Psychology

VOLUME 37 NUMBER 7 · JULY 1979

The Stability of Behavior: I. On Predicting Most of the People Much of the Time 1097
Seymour Epstein

Psychological Differentiation: Current Status 1127
Herman A. Witkin, Donald R. Goodenough, and Philip K. Oltman

Self As Actor, Active Observer, and Passive Observer: Implications for Causal Attributions 1146
John D. Cunningham, Philip A. Starr, and David E. Kanouse

Self-Esteem and Education: Sex and Cohort Comparisons Among High School Seniors 1153
Patrick M. O'Malley and Jerald G. Bachman

Deindividuation, Self-Awareness, and Disinhibition 1160
Ed Diener

Effects of Difficulty and Diagnosticity on Choice Among Tasks in Relation to Achievement Motivation and Perceived Ability 1172
Ursula Buckert, Wulf-Uwe Meyer, and Heinz-Dieter Schmalt

Role of Foreseen, Foreseeable, and Unforeseeable Behavioral Consequences in the Arousal of Cognitive Dissonance 1179
George R. Goethals, Joel Cooper, and Anahita Naficy

Focus of Attention, Chronic Expectancy, and Responses to a Feared Stimulus 1186
Charles S. Carver, Paul H. Blaney, and Michael F. Scheier

Sharing Secrets: Disclosure and Discretion in Dyads and Triads 1196
Ralph B. Taylor, Clinton B. De Soto, and Robert Lieb

Expressive Tendencies and Physiological Response to Stress 1204
Clifford I. Notarius and Robert W. Levenson

The Cognition–Emotion Process in Achievement-Related Contexts 1211
Bernard Weiner, Dan Russell, and David Lerman

Instrumentality Effects in the Assessment of Racial Differences in Self-Esteem 1221
Bernadette Gray-Little and Mark I. Appelbaum

(Continued on inside back cover)

Diary No. 122
Date 3/2/80
File No. Library
Bureau Ednl Pay. Research.

Published monthly by the
AMERICAN PSYCHOLOGICAL ASSOCIATION
ISSN 0022-3514

Editor

Clyde Hendrick, *University of Miami*

Action Editors

Irwin Altman, *University of Utah*
Charles S. Carver, *University of Miami*
Richard J. Harris, *University of New Mexico*
John H. Harvey, *Vanderbilt University*
William Ickes, *University of Wisconsin—Madison*
George Levinger, *University of Massachusetts—Amherst*
Brendan Gail Rule, *University of Alberta, Edmonton, Canada*
David R. Shaffer, *University of Georgia*
Cookie White Stephan, *New Mexico State University*
Barbara Strudler Wallston, *George Peabody College for Teachers*

Consulting Editors

Robert Steven Baron, *University of Iowa*
Paul A. Bell, *Colorado State University*
Ellen Berscheid, *University of Minnesota (Minneapolis)*
Michael H. Birnbaum, *University of Illinois at Urbana-Champaign*
Jack Brehm, *University of Kansas*
Philip Brickman, *Northwestern University*
J. Merrill Carlsmith, *Stanford University*
Paul Christopher Cozby, *California State University, Fullerton*
Rick Crandall, *Texas Christian University*
James M. Dabbs, Jr., *Georgia State University*
John M. Darley, *Princeton University*
Andrew R. Davidson, *University of Washington*
James H. Davis, *University of Illinois at Urbana-Champaign*
Kay Deaux, *Purdue University*
Marshall Dermer, *University of Wisconsin—Milwaukee*
Edward Donnerstein, *Iowa State University*
Ebbe B. Ebbesen, *University of California, San Diego*
Yakov M. Epstein, *Rutgers—The State University*
Irene Hanson Frieze, *University of Pittsburgh*
J. Martin Giesen, *Mississippi State University*
Lewis R. Goldberg, *University of Oregon*
William G. Graziano, *University of Georgia*
Alan E. Gross, *University of Maryland (College Park)*
Nancy Hirschberg, *University of Illinois at Chicago Circle*
Chester A. Insko, *University of North Carolina at Chapel Hill*
Alice M. Isen, *University of Maryland, Baltimore County*
Martin F. Kaplan, *Northern Illinois University*
Samuel S. Komorita, *University of Illinois at Urbana-Champaign*
Vladimir J. Konečni, *University of California, San Diego*
Ellen Langer, *Harvard University*
Mark R. Lepper, *Stanford University*
Darwyn E. Linder, *Arizona State University*
Dale T. Miller, *University of British Columbia, Vancouver, Canada*
Judson Mills, *University of Maryland (College Park)*
Warren T. Norman, *University of Michigan (Ann Arbor)*
Michael Ross, *University of Waterloo, Waterloo, Canada*
Barry R. Schlenker, *University of Florida*
Kelly G. Shaver, *College of William & Mary*
Phillip Shaver, *New York University*
Marvin E. Shaw, *University of Florida*
Steven J. Sherman, *Indiana University*
Harold Sigall, *University of Maryland (College Park)*
Mark Snyder, *University of Minnesota (Minneapolis)*
Melvin L. Snyder, *University of Texas at Austin*
Shelley E. Taylor, *Harvard University*
Stephen G. West, *Florida State University*
Steve Worchel, *University of Virginia*
Mark P. Zanna, *University of Waterloo, Waterloo, Canada*

Assistant to the Editor

Sally Miller

APA Journal Staff

Anita DeVivo, *Executive Editor*
Ann I. Mahoney, Manager, *Journal Production*
Marjorie M. Henry, *Production Supervisor*
Robert J. Hayward, *Advertising Representative*
Juanita Brodie, *Subscription Manager*

The Journal of Personality and Social Psychology publishes original reports of research, methodology, theory, criticism, and literature review in the fields of personality and social psychology of normal processes and populations. Submitted manuscripts should delineate an important question and provide a clear answer to it. In the case of research reports, this will often require demonstration that the reported results are generalizable beyond a single research setting. Consult the editorial in the January 1976 issue (Vol. 33, pp. 1–7) for a detailed statement of policies and instructions to authors.

Manuscripts: Submit manuscripts in quadruplicate, prepared according to the *Publication Manual of the American Psychological Association* (2nd ed.). Copies should be highly legible and will not be returned to authors unless they are original typewritten copies. Mail manuscripts according to instructions on p. 1249.

Required supplementary information: Manuscripts cannot be reviewed for publication unless the author provides supplementary information in the following categories: (a) a cover letter that indicates the relation (or the lack of relation) of the submitted manuscript to other papers by the same author(s) that are currently submitted, in press, or recently published; (b) one copy of each such related paper; (c) one copy of a statement of the procedures of the research reported in the submitted manuscript in sufficient detail to serve as a guide for conduct of an exact replication; and (d) one copy of the summary tables (e.g., correlation tables, analysis of variance summaries) for the *major* analyses reported. If it appears that the supplementary materials in categories (c) or (d) might be bulky, the author should exercise judgment and restraint in selecting what to send. It is not necessary to send supplementary material that repeats verbatim manuscript content.

After a manuscript is accepted, authors will be expected to provide assurance that their data are available in unanalyzed form for reanalyses by other investigators and that their research procedures were consistent with the American Psychological Association's published principles of research ethics, except insofar as may be detailed otherwise in the manuscript.

Copyright and Permission: All rights reserved. Written permission must be obtained from the American Psychological Association for copying or reprinting text of more than 500 words, tables, or figures. Permission is normally granted contingent upon like permission of the author, inclusion of the APA copyright notice on the first page of reproduced material, and payment of a fee of $10 per page, table, or figure. Abstracting is permitted with credit to the source. Permissions and fees are waived for authors who wish to reproduce their own material for personal use; fees only are waived for authors who wish to use their own material commercially. Permission and fees are waived for the photocopying of isolated articles for nonprofit classroom or library reserve use by instructors and educational institutions. Libraries are permitted to photocopy beyond the limits of U.S. copyright law: (1) those post-1977 articles with a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through the Copyright Clearance Center, P.O. Box 765, Schenectady, NY 12301; (2) pre-1978 articles, provided that the per-copy fee stated in the Publishers' Fee List is paid through the Copyright Clearance Center, P.O. Box 765, Schenectady, NY 12301. Address requests for reprint permission to the Permissions Office, APA, 1200 Seventeenth Street, N.W., Washington, D.C. 20036.

Subscriptions: Subscriptions are available on a calendar year basis only (January through December). Nonmember rates for 1979: $60 domestic, $64 foreign, $6 single issue. APA member rate: $20. Write to Subscription Section, APA.

Change of Address: Send change of address notice and a recent mailing label to the attention of the Subscription Section, APA, 30 days prior to the actual change of address. APA will not replace undelivered copies resulting from address changes; journals will be forwarded only if subscribers notify the local post office in writing that they will guarantee second-class forwarding postage.

Reprints: Authors may order reprints of their articles from the printer when they receive proofs.

Back Issues and Back Volumes: For information regarding back issues or back volumes write to Order Department, APA.

Microform Editions: For information regarding microform editions write to any of the following: Johnson Associates, Inc., P.O. Box 1017, Greenwich, Connecticut 06830; University Microfilms, Ann Arbor, Michigan 48106; or Princeton Microfilms, Princeton, New Jersey 08540.

Published monthly in one volume per year by the American Psychological Association, Inc., 1400 North Uhle Street, Arlington, Va. 22201. Printed in the U.S.A. Second-class postage paid at Arlington, Va., and at additional mailing offices.

JOURNAL OF
Personality
AND
Social Psychology

Volume 37 July 1979 Number 7

The Stability of Behavior: I. On Predicting Most of the People Much of the Time

Seymour Epstein
University of Massachusetts—Amherst

One of the classic debates in psychology concerns the stability of personality. With rare exception, studies that have correlated objective behavior on two occasions have obtained coefficients below .30. Not only has the direct measurement of objective behavior failed to provide evidence of stability, but self-report scales in attitude and personality inventories, as well as ratings of behavioral samples by judges (although themselves stable), have produced low correlations with objective behavior. Does this indicate, as some have suggested, that stability of behavior lies primarily in the eye of the beholder? The issue can be resolved by recognizing that most single items of behavior have a high component of error of measurement and a narrow range of generality. In four separate studies it was demonstrated that when measures of behavior are averaged over an increasing number of events, stability coefficients increase to high levels for all kinds of data, including objective behavior, self-ratings, and ratings by others, and that objective behavior can then be reliably related to self-report measures, including standard personality inventories. The observation that it is normally not possible to predict single instances of behavior, but that it is possible to predict behavior averaged over a sample of situations and/or occasions, has important implications not only for the study of personality but for psychological research in general.

A critical issue in personality theory is whether stable behavioral dispositions, or traits, exist. On the basis of everyday observation, it seems evident to most people that they do. Yet the vast bulk of psychological research fails to provide confirmatory evidence. It must be concluded that either the lay view is right and our typical methods of research are lacking, or the research findings are correct and the lay view itself is a phenomenon worthy of study. Not surprisingly, psychologists of both persuasions can be found. The debate on the stability of personality, which is one of the classic debates in psychology, has recently been given new impetus by findings from statistical procedures that have suggested a resolution in the form of what has been called "modern interactionism." It will be demonstrated in this article that modern interactionism does not resolve the issue of stability of personality, no matter what other virtues it may have. A solution proposed in this article lies in an entirely different direction, one that is so obvious that, once pointed out, it reminds one of the fairy tale of *The Emperor's New Clothing*.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3707-1097$00.75

Like the fairy tale, the implications of the solution go far beyond the immediate issue at hand.

This is not the place for a lengthy review of the debate. Excellent reviews and discussions are available in papers by Bowers (1973), Ekehammer (1974), Endler and Magnusson (1976), and Magnusson and Endler (1977). Allport (1937, 1961, 1966) can be referred to for a staunch theoretical defense of the trait position and Cattell (1965) and Eysenck (1969, 1970) for sophisticated, empirically based trait theories. The flavor of the attacks on the trait position can be represented by the following quotations spanning the period of the debate.

> Training the mind means the development of thousands of particular, independent capacities, the formation of countless particular habits, for the working of any mental capacity depends upon the concrete data with which it works. Improvement in any one mental function or activity will improve others only in so far as they possess elements common to it also. The amount of identical elements in different mental functions and the amount of general influence from special training are much less than common opinion supposes. (Thorndike, 1906, p. 248)

> Over and over, a battery of tests designed to measure traits such as persistence, or aggressiveness, or honesty yields results so unreliable and undependable (when compared with other criteria) that one is led to question the actual existence of the general trait. In numerous instances the instruments are very loosely constructed and are clearly equivocal. A survey of . . . a wide variety of frequently used tests suggests that there is a fundamental limitation common to most of them. Trait testers appear to assume that whatever they name has objective reality; many need not so much to improve their measures but to improve or change their thought regarding traits. (Lehman & Witty, 1934, p. 49)

> The generality of these (personality) measures over method and situation was still not high enough to justify perpetuating the traditional concepts of personality. The findings required abandonment of a line of research to which I had devoted ten years of my life as a psychologist. The results also required a change in beliefs about the nature of personality. This research, per se, did not say which way the conceptual shift should go, but it suggested very strongly that traditional conceptions of personality as internal behavior dispositions were inadequate and insufficient. (Peterson, 1968, p. 23)

> With the possible exception of intelligence, highly generalized behavioral consistencies have not been demonstrated, and the concept of personality traits as broad predispositions is thus untenable. (Mischel, 1968, p. 146)

> As noted earlier, there was nothing silly about the initial assumption of personologists that everything was glued together until proved otherwise. But since it has now proved otherwise, it seems only fair to give a sporting chance to the counter-assumption that nothing is glued together until proved otherwise. Instead of assuming cross-situation correlations to be +1.00, let us begin by supposing them to be 0.00 until we can explicitly construct them to be otherwise. (Bem, 1972, p. 25)

The charge that personality traits do not exist clearly strikes at the very heart of personality theory. One could well argue that if individuals do not have relatively stable behavioral dispositions that differentiate them from other individuals, then the concept of personality itself can be dispensed with. Shweder (1975), in fact, comes close to arguing just that.

> The degree of relevance of the concept "personality" constructed in this "individual difference" sense is questioned. Some of the most ubiquitous personality assessment procedures, specifically interpersonal checklists, personality inventories, and questionnaire interviews are shown to be subject to a form of systematic bias which creates the "illusion" that the diversity of individual behavioral differences can be reduced to a limited set of person-distinguishing underlying forces (traits, dispositions, scales, factors or dimensions). (p. 455)

As will be seen shortly, the claim of such bias is one of the main arguments raised against the trait position, as it can explain

---

This article and three of the studies reported in it were supported by National Institute of Mental Health Research Grant MH-01293. It was distributed in mimeographed form in May 1977.

I would like to acknowledge my indebtedness to Edward O'Brien for several of the historical quotations included in this acticle, for drawing my attention to findings in the Hartshorne and May studies that provide evidence for stability in personality, and for the quotation to that effect. The information was contained in a paper by O'Brien on stability in personality that he wrote as part of his comprehensive examination for his PhD degree at the University of Massachusetts.

I also wish to express my appreciation to James Averill, Icek Ajzen, Seymour Berger, Ronnie Janoff Bulman, Alice Eagly, Howard Leventhal, and Ervin Staub, for their helpful comments on an earlier draft of this article.

Requests for reprints should be sent to Seymour Epstein, Department of Psychology, University of Massachusetts, Amherst, Massachusetts 01003.

why there are data from self-report inventories and ratings by others in support of stability in personality but not such data from the direct, objective measurement of behavior.

Arguments For and Against the Existence of Traits

Not surprisingly, given attacks on the very foundation of personality theory such as those cited above, some lively debate has ensued. Others who have joined the fray include Alker (1972), Argyle and Little (1972), Averill (1973), Campus (1974), Cronbach (1957, 1975), Endler and Hunt (1968, 1969), Epstein (1977), Fiske (1961), Moos (1969, 1970), Olweus (1977a), Pervin (1968), Raush (1965), Sarason (1977), Schalling (1977), Vale and Vale (1969), Vernon (1964), Wachtel (1973), and Zuckerman (1977). It is not our intention to review the position of each here. For present purposes, it will suffice to list the arguments and the kind of evidence cited for and against traits.

Situationist Position

According to the situationist position, there is little stability in personality, as behavior is determined almost exclusively by situational variables. The belief that there is little stability in personality rests on three major sources of evidence. The most important is that when behavior in one situation is correlated with behavior in another situation, the correlations are so low—usually less than .30—that they have been disparagingly referred to by Mischel (1968) as "personality coefficients." Mischel (1969) further notes, "A correlation of .30 leaves us understanding less than 10% of the relevant variance. And even correlations of that magnitude are not very common and have come to be considered good in research on the consistency of any noncognitive dimension of personality" (p. 1012). At the same time there is no dearth of evidence that behavior varies markedly as a function of stimulus or situational variables.

A second source of evidence consists of findings from the apportionment of variances in analysis of variance designs. This procedure was independently introduced into the assessment of stability in personality by Raush and his colleagues (Raush, 1965; Raush, Dittman, & Taylor, 1959; Raush, Farbman, & Llewelynn, 1960) and by Endler, Hunt, and their colleagues (Endler, 1966; Endler & Hunt, 1968, 1969; Endler, Hunt, & Rosenstein, 1962). They and others since (e.g., Argyle & Little, 1972; Ekehammar & Magnusson, 1973; Magnusson, 1971; Moos, 1968, 1969, 1970; Magnusson, Note 1) have demonstrated that the variance attributable to individual differences is usually much smaller than the variance attributable to situations and to the interaction of individuals and situations.

A third source of evidence cited by situationists is that when people rate others, they tend to attribute more stability to individuals across situations than is objectively warranted (e.g., Bem & Allen, 1974; Jones & Nisbett, 1971; Mischel, 1968; Shweder, 1975). This, of course, can explain how there can be a widespread belief in the stability of personality when, in fact, there is little stability. The study of such bias in person perception has itself become a significant area of research, and has produced a quantity of explanations as to why the phenomenon of falsely perceiving stability in personality occurs. For example, it has been suggested that (a) it is emotionally satisfying to believe that behavior is predictable, particularly when it is someone else's; (b) it is simpler to classify behavior by people than by situations; (c) people have implicit personality theories that assume stability in personality, and their theories bias their perceptions; (d) a primacy effect operates to make new impressions conform to old ones; (e) the observer is always present in the situations that he or she observes in real life, thereby presenting the observer with a biased sample; (f) there is a tendency for observers to equate behavior that elicits the same emotional reactions in them; (g) there is a tendency for judges to generalize from a few attributes that are stable, such as intelligence, to others that are not; and (h) there are more terms for classifying people than for classifying situations, which leads to a bias toward attributing

behavior more often to characteristics of people than to characteristics of situations.

Trait Position

The arguments that have been advanced in favor of traits amount primarily to conjectures that if different procedures had been followed in the investigations of stability in personality, higher stability coefficients would have been found. There are also a few studies that can be cited that obtained respectable stability coefficients.

The arguments can be summarized as follows: (a) Many of the studies undertaken to assess stability in personality have been experimental studies, which are better suited to demonstrate change than stability in personality (Bowers, 1973). (b) The unit of analysis is a critical factor that has not been adequately taken into account. What appears to be instability at a phenotypic level of analysis may be stability at a genotypic level (Alker, 1972; Bowers, 1973). To use one of Bowers' examples, a woman who is continuously changing her wardrobe may be consistently fashionable. (c) The use of moderator variables would considerably increase reliability coefficients (Alker, 1972). (d) A failure to recognize that some people are more variable than others results in reporting generally low stability coefficients, rather than in noting that at least some individuals are highly stable (Alker, 1972; Bem & Allen, 1974). There is at least one study (Bem & Allen, 1974) that has demonstrated this effect. (e) In everyday life, people determine their own environment, which, in turn, helps maintain the stability of their personality. The operation of this effect is not possible in laboratory studies in which individuals are arbitrarily assigned to conditions (Bowers, 1973; Wachtel, 1973). (f) Stability in personality is mediated by an individual's cognitions and, accordingly, will only be found when ideographic procedures are used that take into account the subjective nature of perception (Alker, 1972; Bem & Allen, 1974; Bowers, 1973; Mischel, 1973). (g) Stability may be demonstrable to a greater extent in within-subject relationships than in between-subjects relationships (Alker, 1972). Carlson (1971) makes a related point in noting that although the concept of personality implies an organization of variables within an individual, among a large number of studies she reviewed, "*not a single published study attempted even minimal inquiry into the organization of personality variables within the individual*" (p. 209).

As to studies that have reported evidence for stability in personality, we will not consider investigations relying on self-report inventories, as such studies can simply demonstrate that people's beliefs about their behavior are consistent, which is a far cry from demonstrating that the behavior itself is consistent. Nor will an attempt be made to review all behavioral studies that have reported positive findings. Our purpose will be served adequately by a review of three series of carefully conducted studies. On the basis of the findings in these studies, it is possible to formulate an integrative hypothesis that can account for all the known results to date. This will be followed by a presentation of four studies that were explicitly undertaken to test the hypothesis.

Block (1971, 1977) conducted a series of studies in which the stability of a variety of personality variables was examined during different periods between childhood and adulthood. In one set of studies, records were obtained from the archives of the well-known longitudinal studies at Berkeley of the behavior of students when they were in junior and senior high school. The subjects were then intensively interviewed when they were in their mid-thirties. Judges rated each subject for the different periods on the items in the California *Q* Set (Block, 1971). To ensure that stability coefficients would not be artificially inflated by response sets, memory, or other sources of rater bias, different judges rated the subjects at the different time periods. For the period from junior to senior high school, 58% of 114 personality variables produced stability coefficients that were at least as high as .35, which is significant at the .001 level. Some of the correlations were as high as .70. For the period from senior high school to age 30 and above, 29% of 114 personality variables yielded stability coeffi-

cient of .35 or greater, with some as high as .61.

In a study of younger children by Block and Block (Block, 1977), observations were made during the children's 4th, 6th, and 8th years of life. Two or three judges, who were the children's nursery school teachers, observed the children for 3 hours a day over a period of 5–9 months and then rated each child on a modification of the California *Q* Set. Rater bias was controlled by having different teachers rate the children at different periods. The average stability coefficient for 100 *Q* items was .48, with several over .60, and a few as high as .70. When the *Q* items were grouped into broader scales by factor analysis, the mean stability coefficients rose to .56.

In a series of studies on aggression in young boys, Olweus (1973, 1974, 1977a, 1977b) had classmates rate 201 boys in the 6th grade, and again 3 years later, on a number of variables pertaining to aggression, such as tendency to start fights with peers. Ratings for each boy were averaged over 3–10 raters. Memory effects and response bias were evaluated and controlled by statistical procedures and by examining subgroups in which there was no overlap in raters. The mean stability coefficient for the 3-year period was .66. When correlations were corrected for attenuation due to rater unreliability, the stability coefficients rose to about .80. In a second similar study in which 85 13-year-old boys were the subjects, even higher stability coefficients were obtained. In other studies in the series, it was observed that peer ratings, teacher ratings, and self-ratings on a specially designed aggression inventory were all highly intercorrelated, suggesting the existence of a broad dispositional trait of aggression.

The classic series of studies by Hartshorne and May (1928, 1929) and Hartshorne, May, and Shuttleworth (1930) on honesty, often cited as evidence against the existence of stability in personality, was an enormous project that spanned half a dozen years, employed a large team of researchers, and tested a national sample of over 8,000 children. Among the behavioral items that were assessed were cheating in a classroom, cheating on a take-home examination, cheating during a game, stealing money, lying, and falsifying records of athletic performance. The average intercorrelation of 23 subtests used as part of a total character score was found to be .23, which led Hartshorne and May to conclude that honesty in any single situation has low predictive value for honesty in any other single situation. This conclusion is invariably cited as evidence against the existence of stability in personality. What is generally not known is that when Hartshorne and May combined several tests of honesty into a single score, the reliability coefficient increased to .73, and they concluded,

> Just as one test is an insufficient and unreliable measure in the case of intelligence, so one test of deception is quite incapable of measuring a subject's tendency to deceive. That is, we cannot predict from what a pupil does on one test what he will do on another. If we use ten tests of classroom deception, however, we can safely predict what a subject will do on the average whenever ten similar situations are presented. (Hartshorne & May, 1928, p. 135)

Further support for the existence of a broad trait of honesty in the Hartshorne and May data is provided by a factor analysis by Burton (1963), who observed that a general factor of honesty accounted for nearly 50% of the total variance.

Interactionist Position

According to the interactionist position, the question of which is more important, the situation or the person, is a meaningless one, as behavior is always a joint function of the person and the situation. In its applicability to the issue of stable individual differences, the interaction position can be viewed as a compromise between the trait position and the situationist position, for it acknowledges the existence of behavioral stability, but only within situational constraints.

The so-called modern interactionist position was derived from findings on apportioning variances in an analysis of variance design. As already noted, it was observed by a number of psychologists that the interaction of individuals and situations accounted for more variance than either source of variance by itself. It was consequently argued that an interactionist position should supplant both the trait and the situationist position (Bowers,

1973; Ekehammer, 1974; Endler, 1966; Endler & Hunt, 1968; Magnusson, Note 1).

As for a more general interactionist viewpoint that goes beyond statistical interaction (see Magnusson & Endler, 1977, for a thorough review of this position), it is assumed that individuals and situations are interdependent. That is, the individual's cognitions and perceptual processes, as much as the objective characteristics of the stimulus, determine the meaning of the stimulus. From this viewpoint, behavior can best be viewed as a transaction between the individual and the stimulus, each influencing the other. Since behavior never takes place in a vacuum but always occurs in a situational context, it is meaningless to talk about characteristics of an individual's behavior without specifying the situation in which the behavior occurs. To understand and predict behavior, it is, accordingly, just as necessary to have a classification system for situations as for individuals and, most important, to know how individuals interpret different kinds of situations.

Evaluation of the Arguments For and Against the Existence of Traits

Evaluation of the Situationist Position

The case of the situationists against traits rests mainly on empirical grounds. They note that the concept of traits is not unreasonable but add that it is also not supported by the facts. The strongest evidence they cite against traits consists of the low stability coefficients obtained when data are derived from direct behavioral measurement. However, it is important to recognize that in almost all cases the correlations were based on single items of behavior. Such correlations actually have little relevance for the existence of traits, as no trait theorist believes that a trait can be inferred from a single instance of behavior. A trait is a generalized tendency for a person to behave in a certain manner over a sufficient sample of events and does not imply that he or she will exhibit trait-relevant behavior in all situations or even on all occasions in the same situation. From another standpoint, single items of behavior have a high component of error of measurement, thereby limiting the possibility of replication, and a high component of situational uniqueness, thereby limiting the possibility of generalization. It is no more reasonable to assess the stability of nonintellective behavior by correlating single observations than it is to assess the stability of intellective behavior by correlating single items in an intelligence test. Thus, for statistical reasons alone, the low correlations cited as evidence against the existence of stable response dispositions can be discounted.

The view that there are traits consisting of relatively broad, stable, behavioral dispositions does not require the assumption that situations do not affect behavior. Behavior can vary significantly with situations, and there can still be an underlying consistent thread in behavior averaged over situations. Thus, demonstrations that behavior varies with situations or even that it varies more with situations than with individuals cannot be taken as evidence against the existence of traits.

Let us now consider the procedure of apportioning variances in an analysis of variance design, on the basis of which it has been claimed that there is little stability in personality. It has been falsely argued that if there were stability in personality, individual differences would necessarily account for a relatively large proportion of total variance. This argument is fallacious for two reasons. First, the proportion of variance attributable to any one factor, such as individuals, is always influenced by the range of variability represented by the other factors. Thus, if situations are selected over a wide range of variability and individuals over a narrow range, the proportion of variance for individuals will be smaller than that for situations. It is evident that depending on how one selects individuals and situations, and how the two are selected in relationship to each other, individuals, situations, or their interaction can be made relatively large or small. Second, the analysis of variance has been misused in obtaining estimates of stability, for instead of using the appropriate error term for obtaining estimates of stability coefficients, the variance for individual differences has been compared to total variance

(see Epstein, 1977; Golding, 1975; Olweus, 1977; Magnusson, Note 1, for a further development of this point on the misuse of analysis of variance). It is possible with such a procedure for individuals to be perfectly reliable over time, as indicated by a reliability coefficient of +1.00, and yet have individual differences contribute a relatively small proportion of total variance. For example, consider a situation in which all individuals come out exactly the same in a 50-yard dash and in a 500-yard dash. As the mean difference between events will be many times greater than the differences among racers within events, it is evident that the variance due to individuals will be a small proportion of the variance due to situations, let alone of the total variance.

The argument that judges sometimes attribute stability to people that is not there describes an interesting phenomenon, but cannot establish that there is no stability in behavior apart from such bias. In fact, the assumption that under most circumstances there is stability in personality can account for a bias to assume there is stability even when there is not. Further, the same factors that contribute to the perception of stability in personality could also contribute to its actual stability. If it is assumed that there is a fundamental need for people to establish orderliness and predictability in the world of their experience and that they accomplish this through their cognitions, including their habits of perception (see Epstein, 1973, Note 2; Mischel, 1973), then the same cognitions and perceptions could contribute not only to people's ratings of stability but to the stability of their actual behavior as well.

Finally, it should be recognized that the null hypothesis cannot be proven by the failure of many studies to demonstrate stability in personality. The possibility remains that with new understanding and new approaches, the conditions for demonstrating stability in objectively measured behavior will be established.

Evaluation of the Trait Position

The arguments in defense of traits are, for the most part, speculations that if things had been done differently, stability in personality might have been demonstrated. Although the proposals may be of interest in suggesting new directions for personality research, they do not constitute an adequate defense of the position that there is stability in personality. All these proposals can do is indicate that the issue should not be closed until they have been tried. Unfortunately, though some of these proposals have been tested, they have failed to fulfill their promise (Bem, 1972). In our own research, in which we compared mean within-subject correlations to between-subjects correlations for the same variables, we have generally found the former to yield smaller values.

Finally, one must agree with Bem (1972) that the issue of stability in personality cannot be resolved by disputation but only by data, "And if these separate indices permit Alker to predict behavior across situations better than +.30, Mischel will fold up his tent and steal away" (p. 18).

How is one to evaluate the findings in a few studies that have demonstrated stability in behavior in comparison to the vast number of studies that have obtained negative findings? One possibility is that when enough studies are done, a few are bound to produce significant results by chance alone. An alternative is that the few studies that produced positive findings were better conceived and conducted than the many that failed (see Block, 1977). A careful analysis of the studies that succeeded could uncover the critical conditions for demonstrating stability in behavior. It is noteworthy, in this respect, that all three series of studies that succeeded examined relatively extensive samples of behavior. In the series of studies reported by Block (1971, 1977) and by Olweus (1973, 1974, 1977a, 1977b), judges rated behavior over a relatively long period. In the Hartshorne and May studies (1928, 1929), reliability was found only when a sufficient number of behavioral items was combined into a single index. As previously noted, single items of behavior, like single items in a test, tend to have a high component of error of measurement. Thus, it may be that procedures for reducing error of measurement are critical for demonstrating stability in

personality. Further evidence in support of this conjecture is that in the studies by the Blocks and by Olweus, ratings from several judges were combined, which is one means of reducing error of measurement. In addition, the Blocks found that when they combined a number of single measures into a broader one, stability coefficients markedly increased, a finding also reported by Hartshorne and May. Thus, a possibility that must be considered is that the critical factor separating the studies that succeeded in establishing stability from those that failed to do so is the steps taken to reduce error of measurement by obtaining adequate samples of behavior in the former studies.

Evaluation of the Interactionist Position

As already noted, there have been problems with the statistical and methodological procedures employed by the interactionists. Moreover, the claim that the interaction of individuals and situations accounts for more variance than either individuals or situations alone has not been uniformly upheld (see Sarason, Smith, & Diener, 1975). In a review of a large number of studies, Sarason and his colleagues found that individuals, situations, and their interaction did not account for very much of the total variance. Rather, the predominant source of variance was error variance. Thus, which of the first three sources of variance is most important is a moot point, as often none accounts for a satisfactory amount of variance.

The more general interactionist position does not involve so much a new position as a rewakening of interest in an old position (see Ekehammer, 1974). Murray (1938), in his classic studies of personality, not only endorsed an interactionist view at a theoretical level, but developed a classification system for response tendencies and for subjectively and objectively defining stimulus situations. Even more to the point, his concept of thema identified a unit of interaction between a response disposition and a stimulus variable. There is little question but that behavior is a joint function of the person and the situation. However, this is irrelevant to the question of whether a reasonable degree of stability in individual behavior can be demonstrated when the behavior is averaged over a sufficient sample of situations. It is noteworthy, in this respect, that interactionists have been no more successful than others in breaking the .30 personality barrier. If they are to succeed in doing so, they will have to obtain a sufficient sample of people in particular situations. That is, interactionism does not replace the need to reduce error of measurement by sampling, but simply determines what it is that must be sampled, namely people with certain attributes in situations with certain attributes. It can be anticipated, of course, that to the extent to which situational as well as personality variables are taken into consideration, the amount of averaging over occasions that is required for a fixed level of temporal reliability will be reduced.

Resolution of the Three Positions and the Proposal of an Integrative Hypothesis

The trait position, the situationist position, and the interactionist position are often presented as three different approaches to the issue of stability in behavior, with the implication that only one of them can be right. Actually, all of them can be right, because they identify not three different solutions to the same problem but three different problems. The interactionist wishes to study the behavior of people with certain attributes in situations with certain attributes. The trait theorist wishes to study consistent behavioral tendencies in individuals over a sample of situations. The situationist is concerned with the general effects of situations over a sample of individuals, for example, the educator who seeks to determine the most effective method for teaching most pupils to read and the architect who seeks to determine the design for a structure that will please the most people. In other words, for certain purposes it is important to predict behavior of people with certain attributes in situations with certain attributes; for other purposes it is important to predict a person's behavior over a sample of situations; and for yet other purposes it is important to predict the influence of a specific situation on a sample of

people. Underlying all approaches is the need to consider that error of measurement is apt to be high and temporal reliability, or replicability, low when findings are derived from single observations. This is so because of the contribution of unassessable factors of "situational uniqueness" that are apt to be present on any one occasion. (Situational uniqueness will be further discussed in the second article in this series; see Epstein, Note 2.) By averaging over situations and/or occasions, error of measurement can be reduced and temporal reliability increased. Without temporal reliability, meaningful generalization cannot be established, and this applies equally to the demonstration of stable effects for individual differences, for situations, and for their interaction.

The main focus of this article is the stability of individual differences. If our assumption about reducing error of measurement through averaging over observations is correct, it should be possible to routinely break the presumed .30 personality barrier by averaging behavior over a sufficient number of events. It should be noted that this was done implicitly in the studies by the Blocks and by Olweus, as when a judge makes a single rating after observing a child on many occasions, the single rating can be viewed as an intuitive averaging.

The above analysis of what is necessary to demonstrate stability in behavior, simple as it is, is different from previous explanations that have assumed that the crucial factor is the kind of data obtained. Thus, Block (1977) noted that there is impressive evidence for the stability of personality in ratings by judges of real-life behavior (R data) and in self-reports (S data) and that there exists impressive evidence for the coherence of the two forms of data. He further observed, however, that there is no evidence for stability in personality when the data consist of the direct measurement of objective behavior (T data) and that such data are, at best, tenuously related to the other two kinds of data. Block concluded, "It is now incumbent upon us to consider . . . what strategies are likely to extend the realm of coherence so as to include as well the domain of T-data" (p. 63). Mischel (Note 3) also expressed concern over the lack of relationship between objective data and data derived from ratings by others and from self-report. He stated,

> An important test, although surely not the only one—of the utility of constructs about personality dispositions remains their ability to predict the individual's behavior in specific situations. Unless R and S data predict T data appreciably, the links between trait impressions and specific behavior-in-situations remains tenuous. (p. 4)

If our reasoning is correct, the following hypothesis and corollary should define the conditions for routinely establishing stability in all kinds of data, including ratings by others, self-ratings, and objective behavior, and for relating ratings by others and self-ratings to objective behavior.

Hypothesis

Stability can be demonstrated over a wide range of variables so long as the behavior in question is averaged over a sufficient number of occurrences. This applies equally to data derived from the direct measurement of objective behavior, from self-reports, and from ratings by others.

Corollary

Reliable relationships can be demonstrated between ratings by others and self-ratings, including standard personality inventories on the one hand and objective behavior on the other, so long as the objective behavior is sampled over an appropriate level of generality and averaged over a sufficient number of occurrences.

Empirical Testing of the Hypothesis and Corollary

A series of studies has been completed that examined the temporal reliability and in some cases the validity of data derived from self-observation, observation by others, and the direct measurement of objective behavior. In all cases, a similar procedure was followed, that is, behavior was observed on several occasions, and single observations were treated like single items on a test. More specifically, stability coefficients were first determined for

a 1-day sample by correlating each subject's scores on Day 1 with each subject's scores on Day 2. Next, coefficients were determined for a 2-day sample by correlating the mean of a subject's scores on Days 1 and 3 with the mean of the subject's scores on Days 2 and 4, and so on, until the mean of a subject's scores on all odd days was correlated with the mean of the subject's scores on all even days. This made it possible to examine split-half stability coefficients for each variable as a function of the number of observations that were averaged. Such a procedure is analogous to that which compares the reliability coefficients of tests of different lengths.[1]

Study 1: Stability of Self-Recorded Data

Method

Study 1 used data available from a previous investigation of emotions in everyday life. The selection of variables in the study was influenced by a theory of the self-concept that assumes that a person's significant postulates about self and world can be inferred from events that elicit emotions (Epstein, 1973, 1976, in press). For 1 month 14 men and 14 women university students kept records, on specially devised forms, of their most pleasant and unpleasant experience each day. At the end of each week, they met with interviewers who answered their questions and checked for errors. The number of recording days completed by subjects varied from 24 to 34. The first page of each form consisted of a blank page for describing the incident in narrative form. This was followed by a 90-item adjective checklist for recording emotions and a 66-item checklist for recording response tendencies and for noting whether they had been carried out. Emotions and impulses were scored on a 4-point scale for intensity, with a blank signifying zero intensity. Behavior carried out was scored on a 2-point scale for carried out or not carried out. Items were collapsed into broader scales by a factor analysis of the data. These broader scales constituted the data that were analyzed in the present study. As an example of a scale of emotions, worthy included the following six adjectives: adequate, competent, appreciated, respected, pleased with self, and proud. As an example of a scale of response tendencies, stimulus seeking included the following three items: to try something new and adventurous; to take risks, to gamble, to seek stimulation; and to seek thrills and excitement. Scores were obtained by averaging the responses to the items in a scale.

As an afterthought, it was decided to score the narratives for the stimulus conditions, or situational factors, responsible for evoking the emotions. Despite successive modifications of Murray's (1938) system for scoring Thematic Apperception Test protocols and intensive training of judges, it was not possible to obtain satisfactory interscorer reliability coefficients, as information necessary to make decisions was often unavailable. Rather than drop the scoring of situational factors, it was decided to retain it, with the realization that whatever results were obtained on the stability of situational factors would underestimate the true stability but would nevertheless be of some interest in indicating a lower limit of such stability.

Results

Between-subjects reliability coefficients: The stability of individual differences. In Table 1, stability coefficients for pleasant experiences are presented for the mean of all odd days compared to the mean of all even days, for Day 1 compared to Day 2, and for the last day compared to the next-to-last day. Also included are the means and standard deviations for all variables for the data averaged over all days. Table 2 presents the same data for unpleasant experiences. It can be seen in both tables that for a 1-day sample, most reliability coefficients are below .30, and this is true for the last 2 days as well as for the first 2 days, indicating that a bias toward greater stability did not develop over the course of the study. On the other hand, when the mean of all odd days is correlated with the mean of all even days, most coefficients are above .70, and some are above .90. Emotions, which were the central focus of the study and on the basis of which events were selected by the subjects, had the highest stability coefficients, with a mean of .79 for unpleasant emotions and of .88 for pleasant

[1] A related procedure is reported by Patrick, Zuckerman, and Masterson (1974; also see Zuckerman, 1976). They administered an adjective checklist to a class on 13 occasions. They then examined the correlations of responses over an increasing number of days with the sum of the responses over all days. Not surprisingly, as the number of days increased so did the correlations with total days, until a perfect correlation was obtained when the data over all days were correlated with themselves. Of greater interest is the form of the curve, which was negatively accelerated. Correlations increased rapidly over the first 3 days, after which increments became relatively small. This indicates that for data such as those examined in the study, aggregating over three occasions can effectively predict the results for a much larger number of observations.

Table 1
Between-Subjects Reliability Coefficients, Means, and Standard Deviations for All Pleasant Experience Variables in Study 1

Variable	*M*	*SD*	Correlation: All odd vs. all even days	Correlation: Day 1 vs. Day 2	Correlation: Last vs. next-to-last day
Emotion					
Happy	89.4	32.2	.92	−.03	.22
Kindly	69.1	40.7	.89	.40	.29
Calm	75.3	37.7	.88	.43	.34
Adequate	52.6	34.7	.87	.37	.44
Unified	25.3	35.4	.87	.46	.44
Energetic	60.5	29.0	.86	.49	.34
M	62.0	35.0	.88	.36	.34
Situation					
Entertainment	6.1	8.4	.78	−.10	−.05
Love and affection	22.7	15.2	.74	−.10	.47
Freedom	10.9	10.2	.73	.10	.34
Positive evaluation	20.7	14.4	.57	.18	−.10
Affiliation	29.0	17.1	.55	.01	.08
Adequacy	20.8	15.0	.52	−.17	.21
Pleasant physical stimulus	6.2	6.8	.45	.35	−.08
Relief	19.9	11.3	.35	−.19	−.03
Aesthetic stimulus	5.9	6.8	−.07	.00	.00
M	15.8	11.7	.55	.00	.09
Impulse					
Pleasure seeking	50.1	29.1	.92	.08	.48
Nurturance	78.2	43.3	.87	.13	.38
Exuberance	62.7	41.0	.82	.38	.26
Stimulus seeking	19.5	20.6	.77	−.18	.04
Problem solving	12.6	10.8	.71	.48	.21
Affiliation	46.9	25.9	.68	.36	.40
Achievement	24.6	20.1	.58	.10	.09
M	42.1	27.3	.76	.19	.27
Behavior					
Nurturance	23.6	17.7	.95	.06	.62
Exuberance	13.4	15.1	.91	.28	−.22
Pleasure	13.7	10.0	.89	−.28	.34
Affiliation	14.4	9.3	.77	.27	.61
Problem solving	4.0	3.6	.70	.23	.02
Stimulus seeking	3.6	4.7	.57	.00	.35
Achievement	6.4	5.4	.40	−.17	.25
M	11.3	9.4	.74	.06	.28

Note. Correlations of .37 and .48 are significant, respectively, at the .05 and .01 levels. $N = 28$.

emotions. It is not evident why positive emotions should produce higher reliability coefficients than negative emotions, but this finding has held up in all further studies. Perhaps subjects are more confused about their negative than their positive feelings.

As expected, the lowest reliability coefficients were obtained for situations. Yet, even here, the mean correlation was well above the .30 barrier, more than half the correlations were significant at the .01 level, and several were above .70. The finding of stable

Table 2
Between-Subjects Reliability Coefficients, Means, and Standard Deviations for All Unpleasant Experience Variables in Study 1

Variable	*M*	*SD*	Correlation: All odd vs. all even days	Correlation: Day 1 vs. Day 2	Correlation: Last vs. next-to-last day
Emotion					
Blocked	43.7	27.4	.91	.25	.45
Fragmented	36.0	25.0	.89	.32	.37
Depressed	60.9	33.5	.88	.22	.38
Tense	34.6	23.3	.82	.52	.07
Angry	70.7	32.7	.77	.26	−.04
Frightened	46.0	21.8	.72	.20	.14
Inadequate	25.9	19.8	.72	.09	.40
Tired	37.0	29.2	.62	.12	.33
M	44.5	26.6	.79	.25	.26
Situation					
Loss of love	12.8	12.4	.70	−.11	.00
Noxious stimulation	11.0	11.1	.54	−.01	−.08
Frustration	46.2	18.2	.53	.06	.01
Isolation	4.6	6.0	.50	−.04	.35
Inconsideration	15.9	11.3	.44	.08	.36
Attack	13.3	12.6	.42	.19	.08
Identification	11.5	11.5	.40	−.16	−.16
Failure	18.4	12.5	.29	−.06	.25
Immorality	6.1	6.9	.28	−.04	−.10
Accidental injury	4.4	6.2	.06	.00	−.05
M	14.4	10.9	.40	−.02	.07
Impulse					
Problem solving	26.2	18.9	.88	.64	.39
Stimulus reduction	33.5	16.6	.85	−.04	.02
Affiliation	31.4	22.4	.82	.32	−.02
Counteraction	39.0	21.0	.82	.08	.47
Physical escape	29.9	25.0	.75	.06	.44
Mental escape	49.1	24.0	.68	−.06	.12
Self-punishment	18.3	16.1	.68	−.06	.29
Tension discharge	18.1	17.6	.66	.48	−.13
Achievement	13.3	10.9	.64	−.05	.59
Aggression	32.4	29.2	.64	.43	−.12
Nurturance	19.4	22.8	.51	.24	−.08
Withdrawal	32.7	19.4	.51	−.13	.18
M	28.6	20.4	.70	.16	.18
Behavior					
Affiliation	6.5	6.7	.87	.04	.48
Problem solving	8.2	7.5	.87	.33	.67
Stimulus reduction	8.7	5.7	.82	.13	.17
Counteraction	8.6	7.5	.79	.18	.62
Mental escape	12.7	9.2	.73	.06	.28
Achievement	3.1	3.7	.65	−.16	.52
Nurturance	3.6	5.9	.60	−.08	.03
Withdrawal	5.2	5.5	.55	−.06	−.04
Aggression	1.4	2.4	.43	.56	.00
Self-punishment	1.8	2.3	.35	−.04	.34
Tension discharge	3.1	3.7	.22	.81	.04
Physical escape	1.2	1.9	−.05	.00	.00
M	5.3	5.2	.57	.15	.25

Note. Correlations of .37 and .48 are significant, respectively, at the .05 and .01 levels. $N = 28$.

individual differences for situational variables is consonant with the view that an important source of stability in personality is that people create stable environments for themselves (Bowers, 1973; Wachtel, 1973).

Figure 1 presents mean stability coefficients averaged over all variables in a category as a function of the number of days in the sample. In all cases, correlations are relatively low for a 1-day sample, rise sharply as the number of days in the sample is increased, and then approach an asymptote. Figure 1 can be used to determine how many observations are necessary to obtain different levels of reliability. It can be seen, for example, that for unpleasant emotions, a 4-day sample is required for a mean stability coefficient of about .50 and a 10-day sample for a mean coefficient of about .75.

The most obvious explanation of the increase in reliability with averaging over increasing observations is that the averaging reduces error of measurement. This explanation can be tested by examining standard deviations as a function of the number of days averaged. In Figure 2 it can be seen that for all categories, there is an initial sharp decline in standard deviation as a function of increasing observations, followed by a gradual approach to an asymptote. The curves, to a large extent, are an upside-down reflection of the curves for the correlations in Figure 1. As the correlations get larger, the standard deviations get smaller. This indicates that the variance being eliminated is within-subject variance, which from the viewpoint of the present study (where concern is with establishing stable individual differences), can be regarded as error variance.

To what extent are the findings on stability coefficients limited because the data consist of self-report statements? Such a question misses the point, which is to demonstrate that in a condition where a high level of stability can be shown to exist by adequate sampling, it would not be found if only a single event had been observed. It follows that the vast majority of studies that have failed to find stability in personality when comparing behavior on single occasions used procedures incapable of demonstrating stability. In any event, other studies, reported later, indicate that the results are not limited to self-report data.

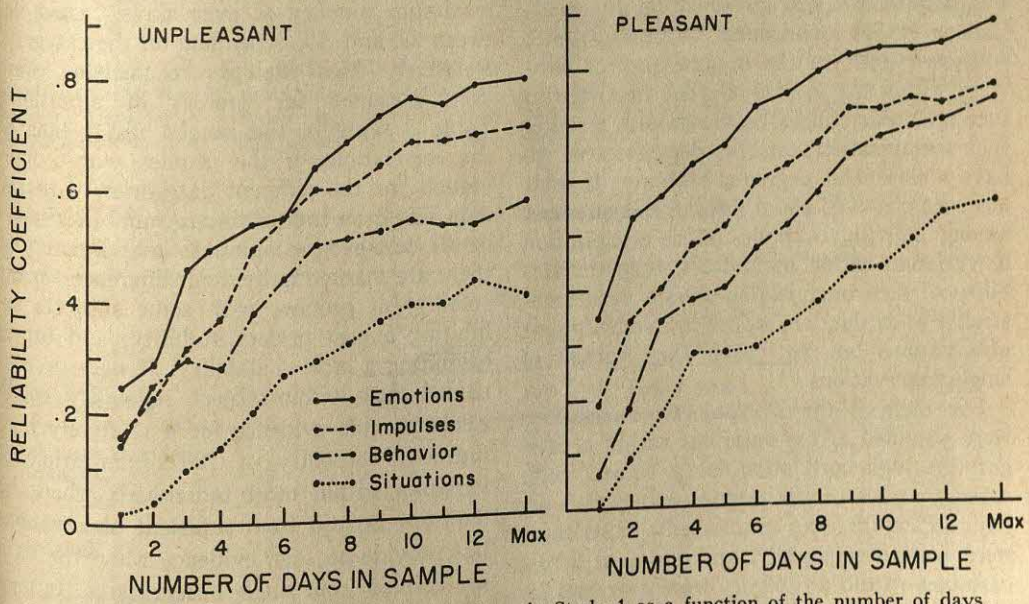


Figure 1. Between-subjects reliability coefficients in Study 1 as a function of the number of days in the odd–even samples. (Values plotted represent the mean of the correlations for the variables in a category. [From "Traits are Alive and Well" by S. Epstein. In D. Magnusson & N. S. Endler (Eds.), *Personality at the crossroads: Current issues in interactional psychology*. Copyright 1977 by Erlbaum, Hillsdale, N.J. Reprinted by permission.])

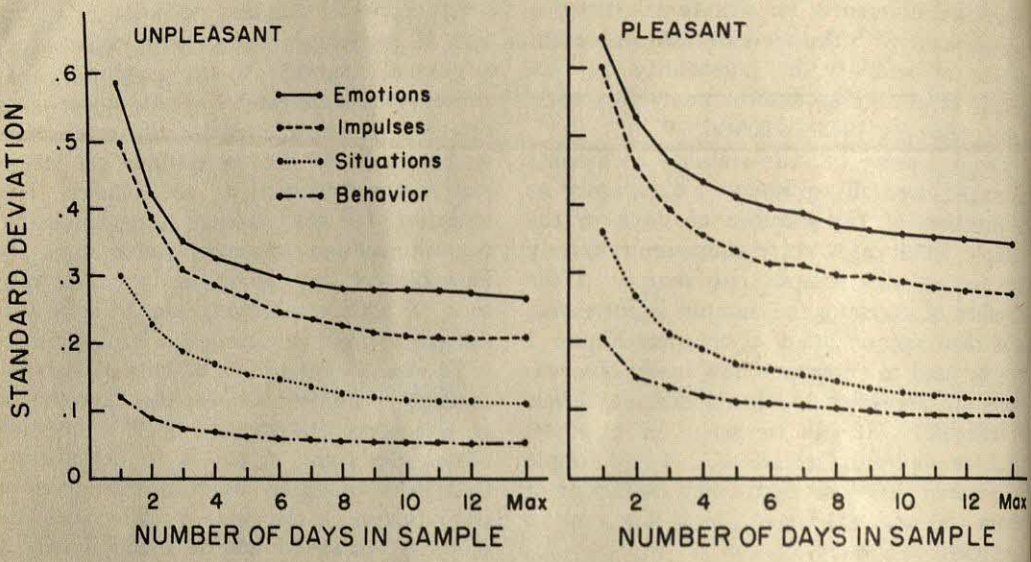


Figure 2. Standard deviations as a function of the number of days in the odd–even samples in Study 1. (Values plotted represent the mean standard deviation for the variables in a category.)

Within-subject reliability coefficients: The stability of the organization of variables within an individual. No matter what else personality is, it is widely recognized that it involves an organization of variables within an individual. Yet, despite universal agreement on this point, Carlson (1971) found not a single study that addressed itself to the organization of variables within the individual. The first step in such an inquiry might be to establish whether such organization can be demonstrated to have a reasonable degree of stability. If what has been observed about error of measurement among individuals applies to the organization of variables within individuals, then the stability of such organization should be demonstrable when data are averaged over sufficient observations but not when they consist of single observations.

For each of the 28 subjects, correlations were obtained across variables within a category. A high correlation for a subject indicated that his or her profile was stable over days. Pleasant and unpleasant experiences were treated separately, and the data across variables within a category were averaged for different numbers of days. It can be seen in Figure 3 that when profiles were derived from single observations, the average intrasubject stability coefficient was less than .25 over all categories. As the number of days from which individual profiles were derived was increased, the stability coefficients also increased. The average correlation for profiles based on the maximum number of odd days versus the maximum number of even days varied between .60 and .79, depending on the category. Relatively high stability coefficients were even obtained for profiles of situations. Table 3 presents the ranges and means of the correlations of the profiles within individuals for the different categories, with the data based on the maximum number of days in the odd–even samples. It is apparent that there are marked individual differences in the stability of profiles, with some subjects exhibiting almost perfect stability, and others exhibiting a lack of stability. It may be concluded that within-subject reliability coefficients provide evidence for a relatively high degree of stability of the organization of variables within most individuals when the data are derived from sufficient observations but provide no such evidence when the data are derived from single observations. It may further be concluded that there are marked individual differences in the degree to which individuals exhibit stability in their personality profiles.

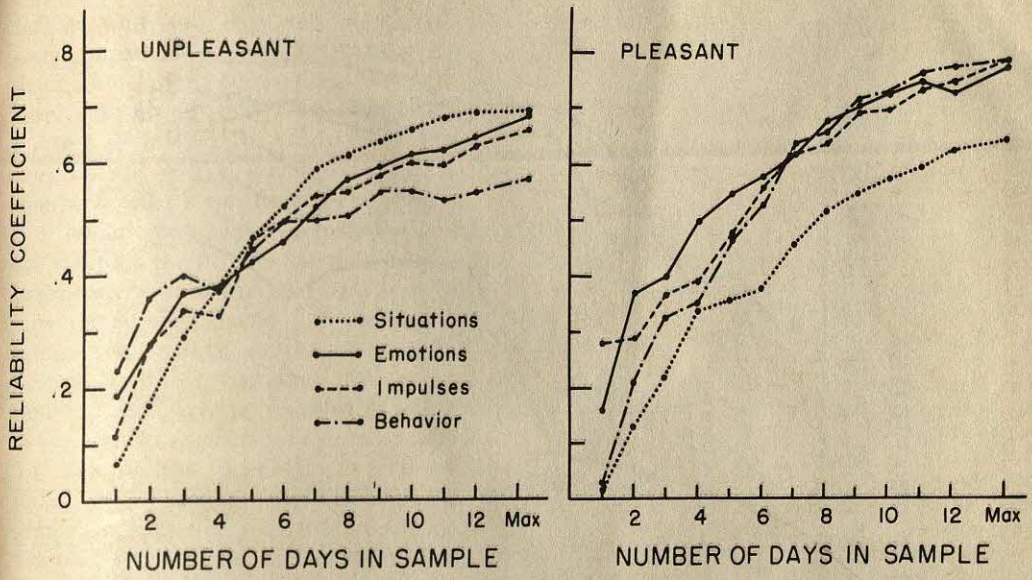


Figure 3. Within-subject reliability coefficients as a function of the number of days in the odd–even samples in Study 1. (Values plotted represent the mean of the correlations for the variables in a category. [From "Traits are Alive and Well" by S. Epstein. In D. Magnusson & N. S. Endler (Eds.), *Personality at the crossroads: Current issues in interactional psychology.* Copyright 1977 by Erlbaum, Hillsdale, N.J. Reprinted by permission.])

Study 2: Stability of Behavior Observed by Others

Method

The data for this study were obtained from an investigation by Barry Leon (Note 4) of social and impulsive behavior in everyday life. Thirty-two pairs of women students at Smith College served as subjects and observers in the study. For a couple to be selected, the observer had to have ample opportunity to observe the subject during and outside of class. The observer kept records of the subject's behavior on eight items related to impulsivity and sociability. Ratings were made on 3-point scales for frequency of occurrence. The following is an example of one of the items: "She actively sought out the company of others: (a) never, (b) once or twice, or (c) three or more times."

Table 3
Means and Ranges of Within-Subject Odd–Even Reliability Coefficients for Maximum Number of Days Sampled in Study 1

Category	*n*	Least reliable subject within category: Subject no.	Least reliable subject within category: *r*	Most reliable subject within category: Subject no.	Most reliable subject within category: *r*	Mean reliability for all subjects
Unpleasant experience						
Emotions	8	28	.24	16	.96	.68
Situations	10	10	−.05	28	.95	.69
Impulses	12	14	.18	20	.95	.66
Behavior	12	2	−.17	25	.97	.60
Pleasant experience						
Emotions	6	12	.13	3	.95	.78
Situations	9	2	−.04	7	.97	.65
Impulses	7	2	−.15	23	.99	.79
Behavior	7	10	.24	28	1.00	.79

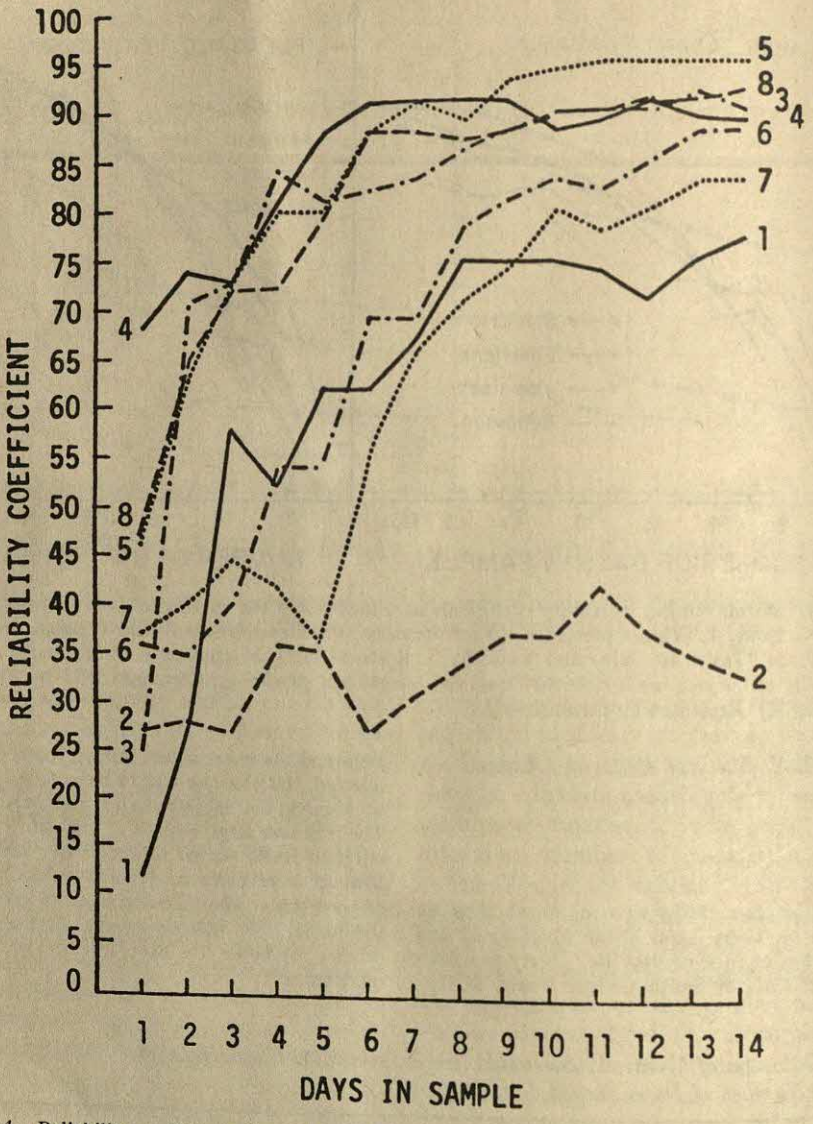


Figure 4. Reliability coefficients for each of the variables in Study 2 as a function of the number of days in the odd–even samples.

Results

As in the previous studies, odd–even correlations were computed for the data averaged over different numbers of days. It can be seen in Figure 4 that for six out of eight variables the pattern is the same, with relatively low reliabilities for a 1-day sample and with increasing reliabilities as the number of observations is increased, until a relatively high level of reliability is obtained. Variable 2, which referred to the closeness of acquaintances with whom a subject initiated conversations and which was a poorly constructed item, since it did not unambiguously define a quantitative continuum, never achieved high reliability, whereas Variable 4, which referred simply to the number of times the subject initiated contacts with others, started with a high reliability coefficient and maintained it throughout.

To determine whether the establishment of a rating set during the course of the study influenced the results, correlations between

the 1st and 2nd day were compared with correlations between the next-to-last and last day. The latter correlations were slightly lower, indicating that the establishment of a rating set during the study could not have contributed to the increase in correlations that occurred with increasing observations.

A further possibility for contamination of the stability coefficients lies in the inferential processes of the observers. That is, a judge's view on stability could have influenced her ratings independent of the behavior of her subject. To the extent that this occurred, it would most clearly be revealed in the ratings of variables that required the most inference. To examine this possibility, variables were sorted into three groups according to the degree of inference required on the part of the observer. It was found that what differences occurred were in the direction of the items which required the least inference producing the highest stability coefficients. Items 3 and 4, which required no inference, produced reliability coefficients of .90 and .89, respectively, which were among the highest obtained (see Figure 4). As demonstrated later, other findings also indicate that the more objective the data, the higher the stability coefficients rise as a function of averaging over observations.

Study 3: Stability of Directly Measured Objective Behavior

Method

This study, in addition to replicating many aspects of Study 1, examined discrete items of objectively measured behavior. It was conducted in two separate classes as an exercise in research.

Nineteen undergraduate seniors in a class in clinical psychology and 15 1st-year graduate students in a class in personality, at the beginning of each class period during the second half of the semester, filled out an adjective checklist on their current emotional state. Following this, they recorded information from daily behavior tallies that they had made since the last meeting. In addition, unknown to the class, the instructor kept records of selected items of behavior.

Current emotional state was rated by the students on 5-point graphic scales anchored at one end with "not at all" and at the other with "very." The scales were identified by clusters of three adjectives, such as happy–cheerful–joyous, the clusters having been determined by factor analysis of adjective checklists used in previous, similar studies (e.g., Epstein, 1976). Ten scales for positive emotions and 10 separate, opposite scales for negative emotions were randomly arranged. In addition to the usual emotions, scales included feelings of self-esteem, arousal, and ability to assimilate information, as these were of special interest for a theory relating emotions to the self-concept (see Epstein, 1973, 1976, in press). After rating their current emotional states, the students transcribed from their daily tally sheets the frequency with which the following occurred since the last class meeting: social telephone calls made and received, social letters written and received, entertainment events attended, and number of headaches and stomach aches. Next, with the instructor keeping time for the class, subjects took three 30-sec samples of their pulse rate and recorded the mean and range. The above information was recorded on computer-scored standard answer sheets. Without the knowledge of the students, the instructor kept records of the following: borrowing of a Number 2 pencil routinely made available to students who had forgotten to bring their own, number of minutes late to class, number of errors and omissions in providing information on the answer sheet, and number of erasures in filling in the spaces for recording answers. In addition, the instructor kept records throughout the semester of number of absences, number of papers turned in late, and number of papers not turned in at all. These will be referred to as fixed variables, as they do not vary over days.

Students in both classes were combined into a single group for the analysis of all variables except minutes late and pencils borrowed, which were recorded only for the undergraduate class. The data finally selected consisted of the records of 15 undergraduates and 11 graduates who had filled out forms properly on at least six Tuesdays and six Thursdays. For all split-half correlations, a balance was maintained between Tuesdays and Thursdays to control for the unequal number of days between meetings.

Results

To evaluate the effects of bias in the self-ratings, it is important to consider that the data can be divided according to level of objectivity in two ways. First, they can be classified as subjective, in the sense that they are descriptions of inner states, and objective, in the sense that they can be observed by an external observer. Second, the latter category can be subdivided according to the degree to which there is the possibility for bias to enter. For example, when a subject records the number of telephone calls or letters he or she has received, the figure may be distorted according to the impression the subject wishes to create. The data can be arranged in three categories of decreasing opportunity for sub-

Table 4
Intercorrelations of Objective Events With Each Other and With Self-Rated Emotions for a 12-Day Sample, and Reliability Coefficients of Objective Events for 1-Day and 12-Day Samples for Study 3

Objective event	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Emotion									
																Fri	A-O	A-I	Unh	Fru	Unw	Trd	Tns	Pls	Cnf
1. Heart rate mean		.74												.34				.34	.31						.33
2. Heart rate range	.74									.57	.32			.45				.37	.36	.39	.42				.46
3. Calls made				.77	.33	.32																			
4. Calls received			.77		.42	.46						.35													
5. Letters written			.33	.42		.79					.52	.43				.33									
6. Letters received			.32	.46	.79						.42	.45													
7. Headaches								.77						.46	.48	.43		.51	.44	.35	.40		.30	.36	.32
8. Stomach-aches							.77									.41	.31	.51	.50	.33	.49			.35	.30
9. Entertainment														.31											
10. Errors		.57												.45											
11. Erasures		.32			.52	.42							.64										.39	.38	.34
12. Papers missing				.35	.43	.45								.47		.41		.43	.40	.41	.42		.32	.62	.51
13. Absences											.64														
14. Lateness	.34	.45					.46		.31	.45		.47			.40	.49	.55	.65	.57	.45	.76	.52	.36	.49	.73
15. Pencils forgotten							.48							.40		.32		.35				.45			
Reliability (1-day)	.66	−.10	.42	.33	.43	.39	.69	.68	−.11	.46	.10	NA	NA	.53	.68	.58	.40	.39	.24	.18	.30	.31	.60	.34	.39
Reliability (12-day)	1.0	.88	.91	.94	.79	.90	.90	.91	.70	.88	.60	NA	NA	.94	.93	.86	.84	.71	.76	.70	.65	.73	.91	.85	.69
M	67	3.3	1.8	1.7	1.7	1.7	1.6	1.4	1.3	1.5	.20	1.3	5.0	3.7	4.8	1.7	1.7	1.6	1.7	1.9	1.6	2.0	2.0	1.6	1.7
σ	26	2.1	.76	.52	.64	.62	.56	.56	.32	.52	.67	1.2	.08	.53	.39	.77	.58	.64	.68	.68	.56	.64	.79	.57	.65

Note. Only correlations of at least .30 are reported. For all variables ($N = 26$) other than pencils forgotten and lateness ($N = 15$), correlations of .39 and .50 are significant respectively at the .05 and .01 levels. For pencils forgotten and lateness, correlations of .51 and .64 are significant at the .05 and .01 levels. Fri = Frightened; A-O = anger-out; A-I = Anger-in; Unh = Unhappy; Fru = Frustrated; Unw = Unworthy; Trd = Tired; Tns = Tense; Pls = Powerless; Cnf = Confused. NA = not applicable; these were fixed variables that consisted of one score for the entire semester.

jective bias as follows: (a) self-recorded events outside the classroom, such as number of telephone calls reported; (b) self-recorded physiological reactions in class, such as the count of pulse rate, with the instructor doing the timing; and (c) examiner-recorded events, such as minutes late to class and number of erasures on the answer sheets. It will be helpful to keep this scheme in mind when examining the influence of bias in self-report on the stability coefficients.

In Table 4, stability coefficients are presented for a 1-day and a 12-day sample for all variables in the study. Also included are the intercorrelations for a 12-day sample of the objective events with each other and with the negative emotions. The results for positive emotions are not included because they add little to the findings for the negative emotions, producing by and large equivalent opposite results. Given a total sample of 12 days, split-half reliability coefficients could, of course, only be computed for subsamples of 6 days. The reliability coefficients for the 12-day sample are estimates obtained by the Spearman-Brown formula for determining the reliability of a total test from the correlation between its halves. To assess the accuracy of such estimates for the current data, 6-day reliability coefficients were estimated from 3-day samples. In all cases, the estimates were almost identical to the values actually observed for a 6-day sample. The 12-day reliability estimates are presented in Table 4, since the intercorrelations of the different variables with each other are all based on 12-day samples, and it is helpful in evaluating the validity coefficients to examine the reliability coefficients for the same number of observations.

All items in Table 4 exhibit a marked increase in reliability from a 1-day to a 12-day sample. For the 12-day sample, 21 of the 23 items have a reliability coefficient of at least .70, and 9 of these are at least .90. Of these 9, only one involves the report of an inner state. The increase in reliability that occurs when data are averaged over events and arranged in order of increasing objectivity results in the following: For the 10 unpleasant inner states, the respective mean coefficients for a 1-day and a 12-day sample are .37 and .77, respectively. For self-recorded behavior that is externally observable, the corresponding figures are .40 and .96. For self-recorded physiological reactions, the corresponding figures are .27 and .94. For examiner-recorded variables, the corresponding figures are .44 and .84. It is noteworthy that all three categories that include externally observable events yield higher reliability coefficients than the category that refers to inner states. That the most objective category, examiner-recorded behavior, attained a somewhat lower level of reliability (.84) than the other two categories of externally observable behavior is attributable to one variable, number of erasures, which had a relatively low frequency of occurrence. The other two variables in this category, lateness and pencils forgotten, obtained reliability coefficients of .94 and .93, respectively. It may be concluded that the increase in reliability as a function of averaging over an increasing number of observations cannot be accounted for by subjective bias associated with self-report, since the same phenomenon was as well demonstrated for objective data. This same conclusion is supported by inspection of Figure 5, which presents the data grouped in a somewhat different fashion. It is apparent in Figure 5 that self-ratings of emotions tend to differ in the direction of producing lower reliability coefficients.

Having established high levels of reliability, let us now examine the data with respect to validity. In Table 4, it is evident that the number of significant correlations of objective events with other objective events and with inner states is well beyond chance. Further, the correlations form coherent patterns. Calls made, calls received, letters written, and letters received are all variables that involve communication with others and are intercorrelated. Erasures, papers missing, and lateness to class, all suggestive of carelessness and lack of organization, are related to the feeling states of tension, powerlessness, and confusion. Heart rate mean, heart rate range, headaches, and stomachaches, suggesting heightened physiological arousal and psychosomatic reactivity, are correlated with unhappiness, confusion, and anger-in but, interestingly, not with anger-out. Heart rate

range, but not heart rate mean, despite the former having a lower reliability coefficient, is significantly correlated with errors (.57) and with erasures (.32), suggesting a variable of physiological lability that is associated with behavioral lability.

It may be concluded that high levels of stability can be demonstrated for objective as well as subjective data when the data are averaged over a sufficient number of events and that once reliability is established, evidence of validity may not be far behind.[2]

Study 4: Relationship of Personality Inventories to Behavior

There are four ways of interpreting the evidence that personality inventories have in the past produced, at best, low correlations when evaluated against a criterion of objective behavior. One is that the two are measuring nonoverlapping aspects of behavior; another is that the inventories are inadequate; a third is that the objective criteria are inadequate; and a fourth is that both are inadequate. Let us consider the possibility that the criteria are inadequate, either because they are unreliable or because they are so narrow and limited in representativeness as to share little variance in common with the broader attributes sampled by personality inventories (see Davidson & Jaccard, 1975; Eagly, 1978; Fishbein & Ajzen, 1974; Jacard, 1974; McGowan & Gormly, 1976; Weigel & Newman, 1976). To the extent that either of these limitations exists, an increase in the sample of objective behavior comprising the criteria should produce an increase in the correlations between personality inventories and objective behavior. Study 4 was undertaken to examine this possibility.

Method

Forty-five undergraduates kept records (on forms similar to those used in the other studies) of their feelings and behavior on 14 consecutive days, not including weekends. Each day, at the same time, a subject set aside 10 min to rate his or her current feelings and to obtain three 30-sec samples of pulse rate. Feelings were rated on bipolar scales with 10 subdivisions that were anchored at one end by a cluster of adjectives, such as happy–cheerful–joyous, and at the other end by an opposite cluster, such as unhappy–sad–depressed. A tally was kept of the following, recorded at the end of each day: number of social phone calls made and received, number of social letters written and received, number of social contacts initiated with groups of three or more people, number of headaches, and number of stomachaches. On the morning of each day, hours of sleep during the previous night were recorded, and a rating was made of soundness of sleep.

Before they began recording their daily behavior, subjects took a battery of personality inventories, including a specific and a general form of a specially constructed inventory made to resemble as closely as possible the forms used for daily recording. The specific form contained precise descriptions of situations and responses. The general form was similar in wording to most standard personality inventories in that it was more vague or general in description of events and response options. An example of an item in the general form is, "How often do you make social phone calls? Consider social calls to include all calls other than business calls." Responses for the general form were selected from a 5-point scale anchored at one end with "almost never" and at the other with "very often." The corresponding item in the specific form was, "How many social phone calls do you make, on the average, over a 5-day period, not including weekends? Consider social calls to include all calls other than business calls." The response options were as follows: less than 1, 1–3, 4–6, 7–9, 10 or more. The general form was always administered before the specific form. Both forms included all the items of objective behavior included in the daily forms, with the exception of heart rate

---

[2] The results are quite different when data from a 1-day sample are examined. A 1-day sample yields a total of eight correlations that are significant at the .05 level when the same variables as in Table 4 are related to each other, as compared to 57 significant relationships when the variables are averaged over 12 days (see Table 4). Moreover, of the eight significant correlations for the 1-day sample, five were produced with the correlation of a fixed variable, papers missing, which is based on the entire semester. The significant correlations for which both variables were based on a 1-day sample were as follows: calls received and calls made (.49), stomachaches and headaches (.58), and unhappy and stomachaches (.39).

To further illustrate the difference between the results for a 1-day and a 12-day sample, consider the interesting, unanticipated findings for heart rate range for a 12-day sample and the absence of any significant relationships for heart rate range when a 1-day sample was examined. In Table 4 it can be seen that for a 12-day sample, heart rate range produced significant correlations with the objective variables of heart rate mean and errors and with self-rated feelings of frustration, unworthiness, and confusion, thereby suggesting an interesting psychosomatic relationship between autonomic lability and behavioral and cognitive lability, a relationship of which there is not a trace in the 1-day sample.

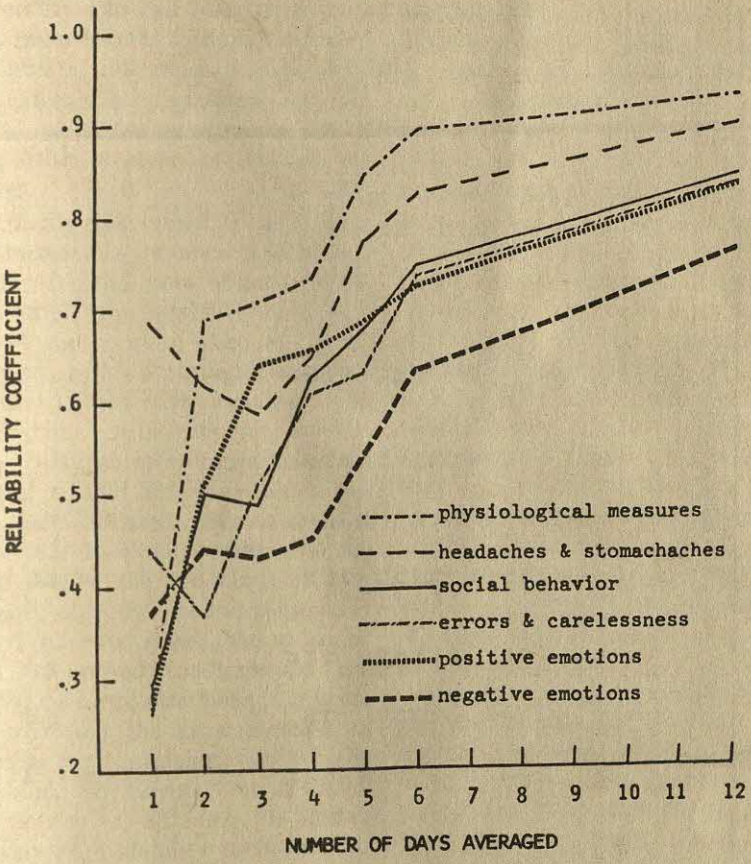


Figure 5. Reliability coefficients as a function of the number of days in the odd–even samples in Study 3. (Values plotted represent the mean of the correlations for the variables in a category. The values for the 12-day samples were estimated from the 6-day samples by the Spearman-Brown formula.)

mean and heart rate range. Frequency of emotions, in the form of an adjective checklist containing single adjectives, was included only in the general form, as it did not seem reasonable to ask subjects to estimate precisely how many times they experienced a particular emotion per day. Subjects responded with estimates of the frequency they experienced the emotions by checking "never," "rarely," "sometimes," "frequently," or "nearly always." Scores on the adjective checklists were obtained by combining individual adjectives into the same 6-item clusters as in the daily forms.

In addition to the above two specially designed inventories, the following more standard personality inventories were administered: the Guilford-Zimmerman Temperament Survey (1949), the Eysenck Personality Inventory (Eysenck & Eysenck, 1968), the Epstein-Fenz Manifest Anxiety scales and the Epstein Hostility scales (Fenz & Epstein, 1965), and the Epstein-O'Brien Self-Esteem Scale (Epstein, 1976). The latter three were specially designed for previous research and have been demonstrated to have some degree of construct validity (cf. Epstein, 1962, 1976; Epstein & Fenz, 1970; Fenz & Epstein, 1965).

Results

The findings on reliability replicate the results of the previous studies that demonstrated a marked rise in stability coefficients as a result of averaging over an increasing number of observations. In Table 5, the stability coefficients for ratings of inner states for a 1-day sample range from .22 for external versus internal direction of attention to .59 for feeling attractive versus unattractive. The mean reliability coefficient for the 15 inner states is .45, which is somewhat higher than in the previous study. The increase in reliability in the present study for a 1-day sample can be attributed to the use of bipolar scales that bring six, rather than three, adjectives to bear on each dimension and that by contrasting the opposites, elucidate the construct. For the entire 14-day sample, as

Table 5

Reliability Coefficients for 1-Day, 7-Day, and 14-Day Samples of Daily Feelings and Events and Correlations of Trait Inventories With 14-Day Samples of Daily Feelings and Objective Events for Study 4

Variable recorded daily	Reliability coefficient			Correlation with scales in trait inventories													
				Trait form of matching scale		Epstein-Fenz anxiety scale			Epstein hostility scale			Epstein-O'Brien self-esteem scale	Guilford-Zimmerman scales				Eysenck scales	
	1-day sample	7-day sample	14-day sample	Specific form	General form	Ins	MT	AA	DH	UH	OH		Res	Asc	Soc	EmS	Ext	Nrt
Emotion																		
Happy–sad	43	66	80		44	−31	−35		−31						36		29	
Kind–angry	55	78	88		48	−43		−29	−40						36	41		−40
Secure–threatened	41	67	80		42	−39	−39	−39	−43	−35	−42							−41
Aroused–tired	32	66	80		45													
Spontaneous–inhibited	41	79	88		40	−36			−31		−30	30		36	41		45	−33
Calm–tense	34	83	91		58	−42	−33	−36	−37	−37	−41	32				49		−50
External attention–introspective	22	55	71		30	−47	−35		−38		−29		−33					−35
Worthy–unworthy	56	79	88		20	−42	−41	−35	−46	−29	−57	47				29		−31
Integrated–disorganized	52	84	91		56	−49	−44	−40	−32		−47	51		35	31	44		−40
Powerful–helpless	59	85	92		44	−54	−47	−45	−46		−48	37	−31	34	35	36	37	−43
Peaceful–agitated	53	82	90		54	−32		−29		−31						34		−36
Attractive–unattractive	59	89	94		55				−33		−48	36			39			
Optimistic–pessimistic	41	85	92		66	−59	−58	−46	−59	−31	−59	55		30	32	44	33	−53
Alert–unreactive	34	64	78		59	−44	−50	−29	−46		−35	40				33		
Outgoing–seclusive	46	70	82		42	−39		−34							51		47	
M correlation	45	75	86		47													
Objective event																		
Heart rate mean	70	88	94															
Heart rate range	49	87	93						40	29								37
Calls made	26	81	90	71	45													
Calls received	45	91	95	72	61													
Letters written	63	59	74	44	38													
Letters received	20	59	74	36	28													
Social contacts	63	94	97	73	46													
Hours of sleep	09	72	84	61	47								−37		40		52	
Soundness of sleep	34	83	91	67	58	−38												
Hours of study	47	77	87	79	80											30		
Headaches	39	75	86	44	51		57	49	49	33				31				
Stomachaches	62	83	91	66	76	31	41	55	49	38						−30		41
M correlation	44	79	88	61	53													33

Note. For bipolar dimensions, positive scores are in the direction of the first term. For standard inventories, only correlations of at least .29 (significant at .05 level) are reported. Correlations of .38 are significant at the .01 level. Decimals are omitted. Ins = Insecurity, MT = Muscle tension, AA = Autonomic arousal, DH = Disturbing hostile feelings, UH = Undercontrolled hostility, OH = Overcontrolled hostility, Res = Restraint, Asc = Ascendance, Soc = Sociability, EmS = Emotional stability, Ext = Extraversion, Nrt = Neuroticism.

estimated from the split-half 7-day coefficients by the Spearman-Brown formula, the reliability coefficients for inner states range from .71 for external versus internal direction of attention to .94 for feeling attractive versus unattractive, with a mean of .86 for the 15 inner states.

A similar increase in reliability as a function of the increase in number of occasions sampled occurs for the more objective data. For a 1-day sample, the reliability coefficients range from .09 for hours of sleep to .70 for heart rate mean, with an average correlation of .44 for the 12 objective variables. The estimated reliability coefficients for a 14-day sample from 7-day samples range from .74 for letters received and letters written to .97 for social contacts initiated, with an average correlation of .88. It is noteworthy that for the objective events, the two kinds of data that were most objective in the sense that they involved nothing more than a count taken under supervised conditions in the classroom, that is, heart rate mean and heart rate range, exhibited the same increase as the other measures, attaining two of the highest reliability coefficients, .94 and .93, respectively. Moreover, although the former had a relatively high stability coefficient for a 1-day sample, the latter did not.

Let us now turn to a consideration of validity by examining the relationships of the personality inventories to the daily recorded events, beginning with the inventories designed to match the daily records on an item-by-item basis. It can be seen in Table 5 that the general form of the inventory has a mean validity coefficient of .53 for objective events, as compared to a validity coefficient of .61 for the specific form. There was apparently a moderate gain in validity as the wording in the self-report inventory was made to more precisely match the criterion. For the specific form, 4 of the 10 validity coefficients are greater than .70, indicating a substantial degree of validity for 40% of the items. These items, with their validity coefficients, are as follows: calls made (.71), calls received (.72), social contacts initiated (.73), and hours of study (.79). Although hours of study requires a degree of inference, as it is not clear exactly what constitutes study, number of letters received and written simply requires a tally of unambiguous events. Thus, the results are not easily explained away by the assumption that daily tallies of events and self-report estimates over an extended period can be viewed as common method variance, in that they both require reports by subjects. Moreover, if all that is involved is common method variance, the correlations of self-report estimates with daily recordings should be as great for a 1-day sample of the latter as for a 14-day sample, which inspection of the data indicates is not the case, and could have been predicted from the consideration that reliability provides a limiting condition for validity. It is noteworthy, in this respect, that in 9 of the 10 variables of objective events under consideration, the validity coefficients for the specific form of the inventory against the criterion of objective events over a 14-day sample are higher than the matching reliability coefficients for a 1-day sample of objective events.[3]

Validity can further be examined by noting the relationships of the standard personality inventories with the 14-day samples of the daily records. It should not be expected, of course, that the relationships will be very high, given the different ranges of generalization encompassed by the recording of specific kinds of behavior and the broad personality dimensions measured by the personality inventories. In Table 5 it can be seen that there are far more significant correlations than would be expected by chance and that the relationships are, for the most part, coherent and relatively high for what is usually obtained for validity coefficients against objective criteria. There is a large number of relationships that break the .30 barrier, and no correlation is in an opposite direction from expectancy. All three of the anxiety scales in the Epstein-Fenz inventory (Fenz & Epstein, 1965) correlate significantly

[3] For a 1-day sample, the mean validity coefficient for self-ratings of emotions was .19 as contrasted with .47 for a 14-day sample. For the general form of the inventory of objective events, the mean validity coefficients for 1-day and 14-day samples were, respectively, .43 and .53. For the specific form of the same inventory, the corresponding figures were .40 and .61.

in the expected direction with daily ratings of threat and tension. Although it is true that the anxiety scales also correlate with daily ratings of other negative feelings, such positive correlations are to be expected, as there is a general tendency for negative emotions to cluster together (see factor analysis of similar data reported in Epstein, 1976). One or more of the anxiety scales correlates significantly negative with soundness of sleep, stomachaches, and headaches. Moreover, the scale of muscle tension correlates more highly with headaches, often a symptom of muscle tension, than with stomachaches, often a symptom of autonomic arousal, whereas the reverse is the case for the scale of autonomic arousal. The scales of overcontrolled hostility and undercontrolled hostility are both directly correlated with daily ratings of threat and of tension, which is reasonable considering that hostility is often a reaction to threat. Yet, only the scale of overcontrolled hostility is significantly correlated with daily ratings of inhibition, inward direction of attention, feelings of helplessness, and lack of reactivity, all of which suggest restraint.

In view of the observation in the previous study that heart rate range is associated with behavioral lability, it is noteworthy that heart rate range in this study is significantly correlated with undercontrolled hostility but not with overcontrolled hostility, whereas heart rate mean is associated with neither. It is of further interest that heart rate range, but not heart rate mean, is significantly associated with the scale of disturbance over hostile feelings (.40) and with Eysenck and Eysenck's (1968) scale of neuroticism (.37), which suggests that, despite its lower reliability than heart rate mean, heart rate range is a more interesting personality variable.

In conformity with the theory of the self-concept (cf. Epstein, 1973, 1976) that guided the selection of the items in the daily forms, the Epstein-O'Brien Self-Esteem Scale (Epstein, 1976) correlates with daily feelings of worthiness (.47), integration (.51), optimism (.55), and alertness (.40). Among the Guilford-Zimmerman scales (1949), the scale of restraint is significantly negatively (−.37), and the scale of sociability significantly positively (.40) associated with number of social contacts initiated. The scale of emotional stability is positively associated with daily ratings of integration (.44), kindliness (.41), optimism (.44), and soundness of sleep (.30) and is negatively associated with frequency of headaches (−.30). The Eysenck (Eysenck & Eysenck, 1968) scale of extraversion is most strongly associated with daily ratings of spontaneity (.45), with outgoing feelings (.47), and with number of social contacts initiated (.52), all of which are recognized elements of extraversion according to Eysenck.[4] The reader may judge for himself or herself by further examination of Table 5 whether the results provide evidence of coherent relationships between self-report inventories and records of daily behavior and feelings.

In summary, the correlations of the daily records with a specific form of an inventory that closely matched the items in the daily form were in the vicinity of .60, the correlations with a more general form of the same inventory were in the vicinity of .50, and the correlations with relevant scales in more standard personality inventories were in the vicinity of .40.[5] It should be noted that all of these correlations are highly significant, that all are above the presumed .30 personality barrier, that none was corrected for attenuation due to unreliability, and that most would be about .10 higher if they were thus corrected. It may be concluded that

---

[4] Not surprisingly, as in the previous study, there were far fewer significant correlations for a 1-day sample, and among these the correlations were lower and less coherent. For example, the Guilford-Zimmerman scale of Sociability (1949) correlated significantly with nine variables, including the variable of number of social contacts initiated, when the data were based on a 14-day sample, but with only one variable when the data were based on a 1-day sample. That variable was self-rated attractiveness, which happens to have relatively high reliability for a 1-day sample (see Table 5).

[5] Similar results have been reported by Zuckerman (1977) and his colleagues (Mellstrom, Cicala, & Zuckerman, 1976; Mellstrom, Zuckerman, & Cicala, 1978; Zuckerman, Note 5). In an interesting series of studies involving inventories, self-rating, and objective measures of fear in a variety of settings, it was observed that the more specifically a trait rating corresponded to a state rating or to an objective measure of behavior, the higher the correlation.

evidence for a respectable degree of validity in self-report inventories can be demonstrated when the criterion consists of an adequate sample of behavior. This is not to imply that scores on personality inventories permit a high degree of accuracy in predicting behavior. The problem of accuracy in predicting behavior will be discussed later.

Discussion

Stability of Behavior and Error of Measurement

The classic debate on stability in personality can be resolved by noting that the problem lay in a failure to take into account error of measurement as it relates to temporal reliability. Single items of behavior, no matter how carefully measured, like single items in a test, normally have too high a component of error of measurement to permit demonstration of high degrees of stability. Once this is recognized, the solution to two related problems becomes apparent. First, the contradictory findings between a few studies that have reported stability in personality using data derived from ratings and a much larger number of studies that failed to find evidence of stability using data derived from the direct measurement of behavior can be accounted for by the observation that the former studies examined adequate samples of behavior, whereas the latter examined single events. Second, the failure to relate self-ratings, ratings by others, and personality inventories to objective behavioral criteria can be accounted for by the unreliability of the behavioral criteria, which almost always consisted of single items of behavior, usually measured in a laboratory setting.

After a review of the literature, the following hypothesis and corollary were formulated: (a) Stability can be demonstrated over a wide range of variables, so long as the behavior in question is averaged over a sufficient number of occurrences. This applies equally to data derived from the direct measurement of objective behavior, from self-reports, and from ratings by others. (b) Significant relationships can be demonstrated between ratings by others and self ratings, including standard personality inventories on the one hand and objective behavior on the other, so long as the objective behavior is sampled over an appropriate level of generality and averaged over a sufficient number of occurrences. Four studies conducted to test the hypothesis and corollary provided unequivocal support for their validity. The studies demonstrated that when single events are examined, there is little evidence for stability, but that when averaging is done over a sufficient sample of events, there is strong evidence for stability, as well as convergence among the different kinds of data. The studies also demonstrated that once high levels of reliability are established, evidence of construct validity is apt to emerge in relationships among the different variables, including ones that do not share common method variance. Thus, error of measurement appears to be the crucial consideration in demonstrating stability in personality and in relating self-ratings and ratings by others to objective data.[6]

It is difficult to believe that with rare exception (e.g., Tryon, 1973), the concept of error of measurement was overlooked throughout the long history of the debate on stability of personality. Perhaps there is a lesson to be learned from this. Can it be that overevaluation of the experimental method, as normally practiced, blinded researchers to the inherent limitations of studying behavior on single occasions? Given the awe in which laboratory experimental procedures have been held, who would have dared to think that they often fail to meet one of the most fundamental scientific tests of all, temporal reliability (replicability)?[7]

---

[6] Other evidence consistent with this thesis is provided by Fishbein and Ajzen (1974), Jacard (1974), Magnusson and Heffler (1969), McGowan and Gormly (1976), Stagner (1977), Tryon (1973), and Weigel and Newman (1976).

[7] To recognize the limitations of the psychological experiment is not to deny its virtues, the primary one of which is to establish causal relationships. Moreover, there are remedies for the limitations. One remedy is to conduct experiments over sessions when this is feasible and appropriate for the phenomenon under consideration, thereby permitting temporal reliability to be assessed and to be increased by averaging over sessions. Another is to encourage replications in different laboratories and to not take

Who would have dared to think that the emperor was wearing no clothing? In both cases the solution is not only emotionally unacceptable but seems too simple to be true. Yet having arrived at the solution, one finds that the future is more hopeful, for delusions can neither warm an emperor nor advance the cause of science. (See also Brunswik, 1947, 1956; Hammond, 1954, 1955; Magnusson & Heffler, 1969; Pervin, 1977; and Tryon, 1973, for other expressions of concern about inadequate sampling of situations.) The broader implications for psychological research of the low temporal reliability of single observations will be discussed in a second article devoted to this topic (Epstein, Note 2).

Issue of Cross-Situational Stability

It has been argued that the real issue with respect to the existence of traits is the demonstration of cross-situational stability. The demonstration of stability per se may, of course, establish nothing more than the existence of narrow habits, which never has been at issue. It is thus important to consider how the findings reported in this article relate to the issue of cross-situational stability.

It will be recalled that a relatively high level of stability in behavior was demonstrated over a wide variety of behaviors, some narrowly conceived, such as making telephone calls, and others broadly conceived, such as feelings of kindliness and acts of nurturance, all of which could be elicited by a variety of situations. Inspection of the data indicated that the same responses were, in fact, elicited by a variety of situations. As the stability that was demonstrated occurred over the normal range of situational variability in everyday life, it would seem fair to state that a meaningful level of cross-situational stability was demonstrated. Expressed otherwise, it was demonstrated that there is enough cross-situational stability in everyday life so that useful statements about individual behavior can be made without having to specify the eliciting situations. This, of course, is the way a trait is usually defined, and the findings demonstrate the utility of such a concept.

Further support for the existence of broad response dispositions was provided by the intercorrelations among the variables. Coherent clusters of correlations occurred not only among variables sharing the same method variance, such as ratings of inner feelings, but also among variables assessed by different methods. Thus, a broadly defined variable in a personality inventory, such as Eysenck's extraversion scale (Eysenck & Eysenck, 1968), produced highly reliable correlations with daily records of feelings of inhibition and of desire for seclusion, as well as with a tally of social contacts initiated, all of which are clearly related to the construct of extraversion as defined by Eysenck. Number of erasures and papers missing, which were recorded by the examiner without the subject's awareness, not only correlated reliably with each other but correlated reliably with daily self-reports of feelings of tension, powerlessness, and confusion.

The conclusion that there are relatively broad, stable response dispositions, or traits, does not conflict with the assumption that situations often exert a strong influence on behavior. People obviously do not manifest response dispositions independent of the setting. That is why it is usually necessary to cancel out situational effects, including subtle background effects that often go unrecognized, by averaging over occasions to demonstrate stability in behavior. By the same token, demonstrations that behavior varies with situations cannot be taken as evidence against traits. The fact that people read in a library and swim in a swimming pool does not establish that there is no generality, or "cross-situational stability," in either swimming or in reading behavior. More to the point is that some people are more prone to

---

any finding seriously until it has been replicated on several occasions. Undoubtedly there are some experimental situations that produce temporally reliable results on a single occasion, whereas others require averaging over occasions. The former very likely include highly potent, ego-involving variables as well as certain very simple effects. However, which situations actually produce temporally reliable results on single occasions cannot simply be assumed, but must be established by empirical procedures in which the temporal reliability of the psychological experiment itself is the object of investigation. This matter is further pursued in Epstein (Note 2).

swim than others when there is a reasonable opportunity to do so, and this may include swimming in pools, in lakes, and in oceans. Further, one cannot test such a cross-situational proclivity to swim by observing a person once in the vicinity of a swimming pool and once in the vicinity of a lake, as there may be many reasons for that person to forego swimming on a particular occasion. Behavior is obviously determined by more than response dispositions. Given an adequate sample of occasions, however, response dispositions will out.

Issue of Predictability of Behavior

How stable is stable? It has been demonstrated that there is sufficient stability in behavior so that over a sufficient sample of events, threads of consistency become apparent. As noted above, in any one instance behavior is determined largely by the situation. This, of course, allows one to predict situational effects averaged over subjects but not necessarily an individual's rank order, as represented by a correlation coefficient, from one situation to the next.

It is important to recognize that no situation for an individual can ever exactly reproduce another, if only because the time lag in between must have somewhat changed the individual, including his or her momentary motivational state. Moreover, always behaving in the same way in the same situation without regard to contextual changes would be behaving in a pathologically rigid manner. On the basis of such considerations alone, it should not have been surprising that the correlations between single items of behavior on different occasions are usually below .30.

To predict individual behavior with reasonable accuracy, correlations in the vicinity of .80 or .90 are required. In the four studies reported in this article, correlations of such magnitude were obtained when the average of a sample of 14 days was correlated with the average of another sample of 14 days, but rarely for smaller samples of behavior. This indicates that one can predict average behavior accurately from a similar sample of average behavior. However, the prediction is only actuarial and is a far cry from predicting with confidence to individual instances of behavior. How useful is such actuarial prediction? It is extremely useful, for it tells us that in the long run, we can depend on people behaving true to character. As I have noted elsewhere (Epstein, 1977), whether we are betting our finances on the outcome of material events or betting our happiness on the outcome of human relationships, it can be extremely rewarding to be right more often than allowed for by chance. From such a perspective, even correlations much lower than .80 or .90 can be useful.[8] With this in mind, let us examine the correlations of self-report statements with daily behavior.

When self-report items that were highly specific in wording were compared with corresponding objective criteria, correlations in the vicinity of .60 were obtained. When the self-report items were worded in a more diffuse manner, the correlations fell to about .50. When standard personality inventories were used that defined much broader styles of functioning than the behavioral criteria to which they were compared, the correlations were, on the average, about .40, but several were above .50. Although none of these allows great accuracy in prediction, all allow much better than chance prediction from information that is readily obtained. Moreover, there is reason to believe that these correlations could be improved by including equally broad categories of items in both the criterion sample and the self-report measure (see studies by the Blocks [in Block, 1977] and by Hartshorne & May, 1928, 1929, described in this paper; see also Fishbein & Ajzen, 1974). It remains to be determined how high such correlations can be brought when a broad sample of matching behavior for the two sets of variables is obtained, and this is combined with averaging the criterion over a sufficient number of occasions to bring it to as high a level of temporal reliability as is customarily

[8] Any correlation that is significantly greater than zero can, of course, be useful for theoretical purposes. In fact, it would be inefficient to have to develop methods and measures that permitted high degrees of prediction to test theories, particularly in early stages of theory development in which it is often desirable to explore many variables, both dependent and independent, rather than pursue a few in depth.

obtained for personality inventories. If both measures had sufficiently high stability coefficients, then at least the unreliability of the criterion would not be a limiting factor for validity, as it customarily has been.

Finally, it should be noted that not everyone is equally predictable. This was demonstrated in the present article by the finding that within-subject correlations varied over a range that suggested almost no stability in a few individuals and extremely high stability in others, with most individuals demonstrating a moderately high degree of stability. Bem and Allen (1974) observed that a sample of subjects who described themselves as highly consistent produced moderately high stability coefficients, whereas a sample of subjects who reported they were not consistent produced low coefficients. The authors concluded that stability in personality can be demonstrated only for specially selected samples of subjects, which is aptly reflected in their title, "On Predicting Some of the People Some of the Time." According to our findings, it is not necessary to select particular classes of people to demonstrate stability unless one has failed to obtain a sufficient sample of behavior to begin with, in which case one needs individuals with unusually high stability to compensate for a high degree of error of measurement. As demonstrated in the present article, given an adequate sample of behavior to begin with, it should be possible "to predict most of the people much of the time."

Reference Notes

1. Magnusson, D. *The person and the situation in the traditional measurement model* (Tech. Rep. No. 426). Stockholm: University of Stockholm, Psychological Laboratories, December 1974.
2. Epstein, S. *The stability of behavior: II. Implications for research.* Manuscript submitted for publication, 1979.
3. Mischel, W. On the future of personality measurement. Unpublished paper in *The future of personality measurement.* Symposium presented at the meeting of the American Psychological Association, Chicago, August 1975.
4. Leon, B. *The stability of social and impulsive behavior as a function of number of observations.* Unpublished manuscript, 1977.
5. Zuckerman, M. *General and specific traits as predictors of self-reported states and behavior in real situations.* Paper presented at the Symposium on Interactional Psychology, Stockholm, June 1975.

References

Alker, H. A. Is personality situationally specific or intrapsychically consistent? *Journal of Personality*, 1972, *40*, 1–16.

Allport, G. W. *Personality: A psychological interpretation.* New York: Holt, 1937.

Allport, G. W. *Pattern and growth in personality.* New York: Holt, Rinehart & Winston, 1961.

Allport, G. W. Traits revisited. *American Psychologist*, 1966, *21*, 1–10.

Argyle, M., & Little, B. R. Do personality traits apply to social behavior? *Journal of Theory of Social Behavior*, 1972, *2*, 1–35.

Averill, J. R. The disposition of psychological dispositions. *Journal of Experimental Research in Personality*, 1973, *6*, 275–282.

Bem, D. J. Constructing cross-situational consistencies in behavior: Some thoughts on Alker's critique of Mischel. *Journal of Personality*, 1972, *40*, 17–26.

Bem, D. J., & Allen, A. On predicting some of the people some of the time: The search for cross-situational consistencies in behavior. *Psychological Review*, 1974, *81*, 506–520.

Block, J. *Lives through time.* Berkeley, Calif.: Bancroft Books, 1971.

Block, J. Advancing the psychology of personality: Paradigmatic shift or improving the quality of research. In D. Magnusson & N. S. Endler (Eds.), *Personality at the crossroads: Current issues in interactional psychology.* Hillsdale, N.J.: Erlbaum, 1977.

Bowers, K. S. Situationism in psychology: An analysis and a critique. *Psychological Review*, 1973, *80*, 307–336.

Brunswik, E. *Systematic and representative design of psychological experiments.* Berkeley: University of California Press, 1947.

Brunswik, E. *Perception and the representative design of psychological experiments.* Berkeley: University of California Press, 1956.

Burton, R. V. Generality of honesty reconsidered. *Psychological Review*, 1963, *70*, 481–499.

Campus, N. Transituational consistency as a dimension of personality. *Journal of Personality and Social Psychology*, 1974, *29*, 593–600.

Carlson, R. Where is the person in personality research? *Psychological Bulletin*, 1971, *75*, 203–219.

Cattell, R. B. *The scientific analysis of personality.* London: Penguin, 1965.

Cronbach, L. J. The two disciplines of scientific psychology. *American Psychologist*, 1957, *12*, 671–684.

Cronbach, L. J. Beyond the two disciplines of scientific psychology. *American Psychologist*, 1975, *30*, 116–127.

Davidson, S. R., & Jaccard, J. J. Population psychology: A new look at an old problem. *Journal of Personality and Social Psychology*, 1975, *31*, 1073–1083.

Eagly, A. H. Sex differences in influenceability. *Psychological Review*, 1978, *85*, 86–116.

Ekehammar, B. Interactionism in personality from a historical perspective. *Psychological Bulletin*, 1974, *81*, 1026–1048.

Ekehammar, B., & Magnusson, D. A method to study stressful situations. *Journal of Personality and Social Psychology*, 1973, *27*, 176–179.

Endler, N. S. Estimating variance components from mean squares for random and mixed effects analysis of variance models. *Perceptual and Motor Skills*, 1966, *22*, 559–570.

Endler, N. S., & Hunt, J. McV. S–R inventories of hostility and comparisons of the proportions of variance from persons, responses, and situations for hostility and anxiousness. *Journal of Personality and Social Psychology*, 1968, *9*, 309–315.

Endler, N. S., & Hunt, J. McV. Generalizability of contributions from sources of variance in the S–R inventories of anxiousness. *Journal of Personality*, 1969, *37*, 1–24.

Endler, N. S., Hunt, J. McV., & Rosenstein, A. J. An S–R inventory of anxiousness. *Psychological Monographs*, 1962, *76*(17, Whole No. 536).

Endler, N. S., & Magnusson, D. Toward an interactional psychology of personality. *Psychological Bulletin*, 1976, *83*, 956–974.

Epstein, S. The measurement of drive and conflict in humans: Theory and experiment. In M. R. Jones (Ed.), *Nebraska Symposium on Motivation* (Vol. 10). Lincoln: University of Nebraska Press, 1962.

Epstein, S. The self-concept revisited, or a theory of a theory. *American Psychologist*, 1973, *28*, 404–416.

Epstein, S. Anxiety, arousal, and the self-concept. In I. G. Sarason & C. D. Spielberger (Eds.), *Stress and anxiety* (Vol. 3). Washington, D.C.: Hemisphere, 1976.

Epstein, S. Traits are alive and well. In D. Magnusson & N. S. Endler (Eds.), *Personality at the crossroads: Current issues in interactional psychology*. Hillsdale, N.J.: Erlbaum, 1977.

Epstein, S. The self-concept: A review and the proposal of an integrated theory of personality. In E. Staub (Ed.), *Personality: Basic issues and current research*. Englewood Cliffs, N.J.: Prentice-Hall, in press.

Epstein, S., & Fenz, W. D. Habituation to a loud sound as a function of manifest anxiety. *Journal of Abnormal Psychology*, 1970, *75*, 189–194.

Eysenck, H. J. *Personality structure and measurement*. San Diego, Calif.: Knapp, 1969.

Eysenck, H. J. *The structure of human personality* (3rd ed.). London: Methuen, 1970.

Eysenck, H. J., & Eysenck, S. B. G. *The Manual of the Eysenck Personality Inventory*. San Diego, Calif.: Educational and Industrial Testing Service, 1968.

Fenz, W. D., & Epstein, S. Manifest anxiety: Unifactorial or multifactorial composition? *Perceptual and Motor Skills*, 1965, *20*, 773–780.

Fishbein, M., & Ajzen, I. Attitudes toward objects as predictors of single and multiple behavioral criteria. *Psychological Review*, 1974, *81*, 59–74.

Fiske, D. W. The inherent variability of behavior. In D. Fiske & S. Maddi (Eds.), *Functions of varied experience*. Homewood, Ill.: Dorsey Press, 1961.

Golding, S. L. Flies in the ointment: Methodological problems in the analysis of the percentage of variance due to persons and situations. *Psychological Bulletin*, 1975, *82*, 278–288.

Guilford, J. P., & Zimmerman, W. S. *The Manual of the Guilford-Zimmerman Temperament Survey*. Beverly Hills, Calif.: Sheridan Supply Co., 1949.

Hammond, K. R. Representative vs. systematic design in clinical psychology. *Psychological Bulletin*, 1954, *51*, 150–159.

Hammond, K. R. Probabilistic functioning and clinical method. *Psychological Review*, 1955, *62*, 255–262.

Hartshorne, H., & May, M. A. *Studies in the nature of character: Vol. 1. Studies in deceit*. New York: Macmillan, 1928.

Hartshorne, H., & May, M. A. *Studies in the nature of character: Vol. 2. Studies in service and self control*. New York: Macmillan, 1929.

Hartshorne, H., May, M. A., & Shuttleworth, F. K. *Studies in the nature of character. Vol. 3. Studies in the organization of character*. New York: Macmillan, 1930.

Jaccard, J. J. Predicting social behavior from personality traits. *Journal of Research in Personality*, 1974, *7*, 358–367.

Jones, E. E., & Nisbett, R. E. *The actor and observer: Divergent perceptions of the causes of behavior*. New York: General Learning Press, 1971.

Lehman, H. C., & Witty, P. A. Faculty psychology and personality traits. *American Journal of Psychology*, 1934, *44*, 486–500.

Magnusson, D. An analysis of situational dimensions. *Perceptual and Motor Skills*, 1971, *32*, 851–867.

Magnusson, D., & Endler, S. Interactional psychology: Present status and future prospects. In D. Magnusson & N. S. Endler (Eds.), *Personality at the crossroads: Current issues in interactional psychology*. Hillsdale, N.J.: Erlbaum, 1977.

Magnusson, D., & Heffler, B. The generality of behavioral data: III. Generalization potential as a function of the number of observation instances. *Multivariate Behavioral Research*, 1969, *4*, 29–42.

McGowan, J., & Gormly, J. Validation of personality traits: A multicriteria approach. *Journal of Personality and Social Psychology*, 1976, *34*, 791–795.

Mellstrom, M., Cicala, G. A., & Zuckerman, M. General versus specific trait anxiety measures in the prediction of fear of snakes, heights, and darkness. *Journal of Consulting and Clinical Psychology*, 1976, *44*, 83–91.

Mellstrom, M., Zuckerman, M., & Cicala, G. A. General versus specific traits in the assessment of anxiety. *Journal of Consulting and Clinical Psychology*, 1978, *46*, 423–431.

Mischel, W. *Personality and assessment*. New York: Wiley, 1968.

Mischel, W. Continuity and change in personality. *American Psychologist*, 1969, *24*, 1012–1018.

Mischel, W. Toward a cognitive social learning reconceptualization of personality. *Psychological Review*, 1973, *80*, 252–283.

Moos, R. H. Situational analysis of a therapeutic community milieu. *Journal of Abnormal Psychology*, 1968, *73*, 49–61.

Moos, R. H. Sources of variance in response to questionnaires and in behavior. *Journal of Abnormal Psychology*, 1969, *74*, 405–412.

Moos, R. H. Differential effects of psychiatric ward settings on patient change. *Journal of Nervous and Mental Disease*, 1970, *5*, 316–321.

Murray, H. A. *Explorations in personality*. New York: Oxford University Press, 1938.

Olweus, D. Personality and aggression. In J. K. Cole & D. D. Jensen (Eds.), *Nebraska Symposium on Motivation* (Vol. 21). Lincoln: University of Nebraska Press, 1973.

Olweus, D. Personality factors and aggression: With special reference to violence within the peer group. In J. de Wit, & W. W. Hartup (Eds.), *Determinants and origins of aggressive behavior*. The Hague, The Netherlands: Mouton, 1974.

Olweus, D. A critical analysis of the "modern" interactionist position. In D. Magnusson & N. S. Endler (Eds.), *Personality at the crossroads: Current issues in interactional psychology*. Hillsdale, N.J.: Erlbaum, 1977. (a)

Olweus, D. Aggression and peer acceptance in preadolescent boys. Two short-term longitudinal studies of ratings. *Child Development*, 1977, *48*, 1301–1313. (b)

Patrick, A., Zuckerman, M., & Masterson, F. An extension of the trait–state distinction from affects to motive measures. *Psychological Reports*, 1974, *34*, 1251–1258.

Pervin, L. A. Performance and satisfaction as a function of individual–environment fit. *Psychological Bulletin*, 1968, *69*, 56–86.

Pervin, L. A. The representative design of person–situation research. In D. Magnusson & N. S. Endler (Eds.), *Personality at the crossroads: Current issues in interactional psychology*. Hillsdale, N.J.: Erlbaum, 1977.

Peterson, D. R. *The clinical study of social behavior*. New York: Appleton-Century-Crofts, 1968.

Raush, H. L. Interaction sequences. *Journal of Personality and Social Psychology*, 1965, *2*, 487–499.

Raush, H. L., Dittman, A. T., & Taylor, T. J. Person, setting and change in social interaction. *Human Relations*, 1959, *12*, 361–378.

Raush, H. L., Farbman, I., & Llewellyn, L. G. Person, setting and change in social interaction: II. A normal–control study. *Human Relations*, 1960, *13*, 305–332.

Sarason, I. G. The growth of interactional psychology. In D. Magnusson & N. S. Endler (Eds.), *Personality at the crossroads: Current issues in interactional psychology*. Hillsdale, N.J.: Erlbaum, 1977.

Sarason, I. G., Smith, R. E., & Diener, E. Personality research: Components of variance attributable to the person and the situation. *Journal of Personality and Social Psychology*, 1975, *32*, 199–204.

Schalling, D. The trait–situation interaction and the physiological correlates behavior: The example of smoking. In D. Magnusson & N. S. Endler (Eds.), *Personality at the crossroads: Current issues in interactional psychology*. Hillsdale, N.J.: Erlbaum, 1977.

Shweder, R. A. How relevant is an individual difference theory of personality? *Journal of Personality*, 1975, *43*, 455–484.

Stagner, R. On the reality and relevance of traits. *The Journal of General Psychology*, 1977, *96*, 185–207.

Thorndike, E. L. *Principles of teaching*. New York: Seiler, 1906.

Tryon, R. C. Basic unpredictability of individual responses to discrete stimulus presentations. *Multivariate Behavioral Research*, 1973, *8*, 275–295.

Vale, J. R., & Vale, G. R. Individual differences and general laws in psychology: A reconciliation. *American Psychologist*, 1969, *24*, 1093–1108.

Vernon, P. W. *Personality assessment. A critical survey*. New York: Wiley, 1964.

Wachtel, P. L. Psychodynamics, behavior therapy and the implacable experimenter: An inquiry into the consistency of personality. *Journal of Abnormal Psychology*, 1973, *82*, 324–334.

Weigel, R. H., & Newman, L. S. Increasing attitude-behavior correspondence by broadening the scope of the behavioral measure. *Journal of Personality and Social Psychology*, 1976, *33*, 793–802.

Zuckerman, M. General and situation-specific traits and states: New approaches to assessment of anxiety and other constructs. In M. Zuckerman & C. D. Spielberger (Eds.), *Emotions and anxiety. New concepts, methods and applications*. Hillsdale, N.J.: Erlbaum, 1976.

Zuckerman, M., & Mellstrom, M., Jr. The contributions of persons, situations, modes of responses, and their interactions in self-reported responses to hypothetical and real anxiety-inducing situations. In D. Magnuson & N. S. Endler (Eds.), *Personality at the crossroads: Current issues in interactional psychology*. Hillsdale, N.J.: Erlbaum, 1977.

Received June 23, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 7, 1127-1145

Psychological Differentiation: Current Status

Herman A. Witkin, Donald R. Goodenough, and Philip K. Oltman
Educational Testing Service, Princeton, New Jersey

The status of the differentiation hypothesis, derived from differentiation theory, is examined in light of the evidence that has accumulated since the hypothesis was proposed a decade and a half ago. Assuming that greater or less differentiation is an organismic attribute, the hypothesis postulates that behaviors reflecting extent of differentiation are likely to be interrelated, resulting in self-consistency in individual functioning across domains. The newer evidence has, on the whole, confirmed the linkages among the behaviors examined in the earlier research that was done to test the differentiation hypothesis. The differentiation hypothesis has also proven useful in generating predictions about linkages to behaviors in new domains (cognitive restructuring, interpersonal competencies, and cerebral lateralization); these predictions have been tested and generally confirmed. The differentiation construct appears able to account for phenomena that cannot be accommodated by other lower order constructs, such as field dependence–independence. On all these grounds the differentiation construct continues to serve a useful function. Whereas the differentiation hypothesis has dealt only with the interrelatedness among components of a cluster of behaviors subsumed under differentiation, the newer evidence carries suggestions for a hierarchical ordering of these components and for the nature of causal connections among them.

The differentiation construct was introduced into the stream of research on field dependence–independence in 1962 to accommodate new findings and to guide further research on the broad patterns of psychological functioning associated with individual differences in manner of establishing the upright in space (Witkin, Dyk, Faterson, Goodenough, & Karp, 1962/1974).

Briefly stated, differentiation is a major formal property of an organismic system. A less differentiated system is in a relatively homogeneous state; a more differentiated system is in a relatively heterogeneous state. A system that is more differentiated shows greater self–nonself segregation, signifying definite boundaries between an inner core of attributes, feelings, and needs identified as the self, and the outer world, particularly other people. In a less differentiated system, in contrast, there is greater connectedness between self and others. A system that is more differentiated is also characterized by greater segregation of psychological functions; that is, functions are more separate from each other and activities within each are more specialized.[1]

To make the differentiation construct a useful guide for research, it was necessary to operationalize it by specifying the particular ways in which greater or less differentiation

Herman A. Witkin died on July 8, 1979, while this article was in press.

The preparation of this review was aided by Grant MH-21989 from NIMH.

Requests for reprints should be sent to Donald R. Goodenough, Center for Assessment and Research in Human Development, Educational Testing Service, Princeton, New Jersey 08541.

[1] Although more differentiated systems are likely to be more complexly organized (i.e., relationships among system components and between the system and its environment are more elaborate), there is no inherent relation between differentiation and effectiveness of integration (i.e., the harmonious working together of system components with each other, and of the system as a whole with its environment). Supporting the concept that level of differentiation is unrelated to effectiveness of integration are the results of many studies that have shown that more differentiated and less differentiated people are not different in sheer presence or absence of psychopathology, in other words, in being well adjusted or poorly adjusted (Witkin, 1965).

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3707-1127$00.75

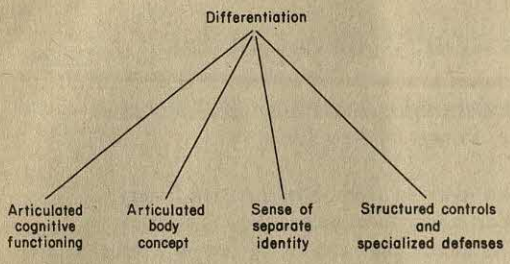


Figure 1. 1962 model of differentiation.

may show itself in different psychological domains, and by formulating testable hypotheses. Assuming that the development of differentiation is an organismwide process, the main hypothesis we proposed was that greater or less differentiation is likely to be characteristic of an individual's activities in diverse domains. According to this differentiation hypothesis, functional manifestations or indicators of greater or less differentiation should be related to some extent, resulting in self-consistency in functioning across domains. We held this expectation, though recognizing that special training and experience during growth may contribute to unequal progress among areas, and that dedifferentiation during aging or with psychological impairment may affect functioning irregularly from one area to another.

In the theoretical conception that guided our research, shown schematically in Figure 1, differentiation was considered a high-order individual-differences construct, with four lower order constructs radiating from it. The first of these was articulated cognitive functioning. To experience parts of organized fields as discrete and to organize unstructured fields, which articulation implies, is to show differentiation in the cognitive domain. The second was an articulated body concept. This refers to an impression of the body as having definite limits or boundaries, and the parts within as discrete yet interrelated and joined into a definite structure. The third was sense of separate identity—that is, identification of attributes, needs, and values that are recognized as one's own and as distinct from those of others. The availability of an internal core of characteristics and standards was expected to enable the person to function with relatively little need for guidance and support from others, and to maintain an internally derived perspective in the face of contradictory perspectives of others. The fourth was availability of structured controls over impulse expression and the use of specialized defenses for dealing with potentially disturbing experiences. Both signify more differentiated functioning, in contrast to the diffuse expression of impulse and the use of nonspecific defenses. By 1962 there was a good deal of evidence that measures of these lower order constructs were interrelated, as hypothesized.

Since then, the differentiation hypothesis has been the stimulus for a large body of further research.[2] A segment of that research, concerned with the field-dependence–independence cognitive-style component of differentiation, has recently been reviewed elsewhere (Witkin & Goodenough, in press; Witkin & Goodenough, Note 5). Our main task here is the broader one of examining the differentiation hypothesis in light of the newer evidence. We consider that evidence under the headings of the four lower order constructs of the 1962 model of differentiation.

Articulated Versus Global Field Approach

The articulated verus global concept emerged from a line of research on individual differences in perceptual and intellectual functioning that began with studies of perception of the upright in the rod-and-frame test (RFT), body-adjustment test (BAT), and rotating-room test (RRT). In each of these tests, subjects differed in the extent to which they used the external visual field or the body itself as the main referent for locating the upright; and they were self-consistent across tests in their tendency to rely primarily on the one referent or the other. These contrasting tendencies were designated *field dependent and field independent.*

Later, manner of locating the upright was shown to be related to relative ease of separating part of an organized field from the field as a whole, as in the embedded-figures test (EFT), in which the task is to find a simple

[2] For recent comprehensive bibliographies see Cox and Witkin (Note 1); Witkin, Cox, and Friedman (Note 2); Witkin, Cox, Friedman, Hrishikesan, and Siegel (Note 3); Witkin et al., (Note 4).

figure within a complex design (Witkin et al., 1954/1972). Since the requirement of separating the item from its embedding context in this task appeared similar to the task of keeping a rod or the body apart from the surrounding field in the RFT, BAT, and RRT, we redefined the field-dependence–independence concept as a disembedding ability in perception. Disembedding ability in perception was subsequently related to disembedding ability in intellectual activities, and disembedding ability in both domains was related to structuring competence in both domains. To analyze and structure fields is to show articulated cognitive functioning as a characteristic approach to the field; to follow the field as given is to use a global approach. This greatly enlarged individual-differences dimension was conceived as an articulated versus global field approach dimension, and was designated a cognitive style.

Since 1962, a great deal of research has been carried out on the articulated–global field approach conception. (See Witkin & Goodenough, in press, and Witkin & Goodenough, Note 5, for recent reviews of that research.) This research has confirmed the picture of self-consistency in cognitive functioning established earlier, and extended it by linking a wide array of spatial–visual cognitive dimensions to field dependence–independence in perception of the upright. A common feature of situations in which these dimensions are represented is that the person has to restructure percepts or symbolic representations to meet the requirements of the task. Field-independent people are better able to accomplish such restructuring, in contrast to field-dependent people who are more likely to follow the prevailing organization of a stimulus array as given. Examples of restructuring acts on which people who are field-independent in perception of the upright do better are locating a simple figure in a complex organized gestalt (as in the EFT; e.g., Goodenough & Karp, 1961), achieving alternative perspectives in spatial–visualization tasks (e.g., Gardner, Jackson, & Messick, 1960; Gough & Olton, 1972), showing conservation in Piagetian situations (e.g., Pascual-Leone, 1969), and breaking a set in Einstellung problems (e.g., Busse, 1968; Dinius, 1975). Most factor-analytic studies have suggested that the restructuring competence related to field independence in perception of the upright may be limited to performance on spatial–visualization types of tasks, but the evidence on the issue is not conclusive.

Recent research on the articulated–global field approach concept also suggests that manner of mixing visual and vestibular information, rather than cognitive restructuring (Witkin & Goodenough, in press; Witkin & Goodenough, Note 5), may underlie individual differences in performance on tests of perception of the upright. In a later section of this article, we consider alternatives to the disembedding hypothesis that may account for the relation that exists between manner of establishing the upright and performance in cognitive restructuring tasks such as the EFT.

Articulated Body Concept

People's conception of their bodies is the product of bringing into a coherent entity a wide array of experiences while growing up. These experiences have their primary sources in children's apprehension of their bodies through vision; in the tactile and kinesthetic feedback arising from their manual exploration of their bodies and from their motoric and bodily functions; and in the handling of their bodies by other people. The bodies of others provide still another source of experience. The typical progression in development of the body concept is from a relatively global view of the body, to awareness of its constituent parts and their interrelatedness, as well as awareness of the outer limits of the body. The outcome of this progression is an articulated body concept —that is, one in which components of the body are experienced as discrete and joined into a bounded whole. According to the differentiation hypothesis, individual differences in position on the articulated–global body-concept dimension should be related to differences in position on the field-dependence–independence cognitive-style dimension.[3] This expectation

[3] The differentiation hypothesis supposes that measures of the postulated indicators of extent of differentiation would tend to be related to each other. In the studies described in our earlier report (Witkin et al., 1962/1974), intercorrelations among measures of these indicators were examined and were generally found to be significant. Reflecting the early and continuing role of

was confirmed in our early studies, in which we used an articulation of body concept scale, applied to human figure drawings, to assess that aspect of the body concept (Faterson & Witkin, 1970; Witkin et al., 1962/1974). This relationship has been further confirmed with a high degree of consistency in many subsequent studies (e.g., Adevai, Silverman, & McGough, 1968; Pizzamiglio & Carli, 1973; Schuck, Cross, & Mills, 1970; Winestine, 1969).

Some studies have examined articulation of body concept by means other than figure drawings. One set of studies used the Finger Apposition Test (Jacobson, 1944), which requires subjects to reproduce with their own hands the finger positions of pairs of hands shown in a series of photographs. The ability to manipulate the hands vicariously in imagination, to achieve a particular relationship between them, was conceived to require an articulated conception of one's own body. A similar kind of reasoning lay behind the use of the Laterality Orientation Test (Culver, 1969) and the Hands Test (Thurstone, 1938) in another set of studies. In these tests the subject is required to identify the sidedness of drawings of right or left body parts. Most studies found the expected relationship between scores on these tests and measures of field dependence–independence (Adams, 1974; Adevai et al., 1968; Bachelis, 1965; Culver, Cohen, Silverman, & Shmavonian, 1964; Epstein, 1961; Gaughran, 1964).

Still other studies considered only the body-boundary component of an articulated body concept. Some of these used the "barrier index" (Fisher & Cleveland, 1958), an indirect indicator of body-concept boundaries based on frequency of bounded percepts in a subject's Rorschach productions. Although a few studies found a relation between boundary scores and field independence (Jacobson, 1965; analyses by Fisher of data from Young, 1959, cited in Witkin et al., 1962/1974), most have not (e.g., Bachelis, 1965; Carlson, Tucker, Harrow, & Quinlan, 1971; Siegel, 1977). Other studies came closer to the body itself in their assessment of body-boundary characteristics, by examining discreteness of experiences on the surface of the body through such indicators as the two-point threshold, traced letter identification, and tactile localization (Adams, 1974; Adevai et al., 1968; Bachelis, 1965; Cohen, Silverman, & Shmavonian, 1962; Paquette, 1974; Silverman, Cohen, Shmavonian, & Greenberg, 1961). In most instances, the anticipated relationship was found between field independence and the measures of discreteness of experience on the surface of the body.

If the various measures derived from tasks involving the body itself do indeed allow inferences about articulation of body concept, we have, in these studies, a suggestion that the relationship between field dependence–independence and that body-concept dimension is not method specific, in the sense of being limited to the figure-drawing, articulation-of-body-concept assessment technique. Additional studies using more experimental approaches to examination of the body concept are clearly needed, however.

Sense of Separate Identity

The extensive evidence derived from studies carried out since 1962 under the aegis of the sense of separate identity concept, recently reviewed by Witkin and Goodenough (1977), has essentially confirmed and extended past findings.

In one approach, sense of separate identity was assessed through its expression in self-object differentiation. Winestine (1969) used strength of the twinning reaction, as an indication of standing on that dimension, for each member of the pairs of male twins she studied. The twinning reaction was conceived to be stronger the greater the degree of interidentification of the boy with his twin, as assessed from an interview with the boy. Boys with a stronger twinning reaction were significantly

---

perception of the upright and perceptual disembedding in our research, as well as the availability of effective tests of these functions, it has predominantly been relations between performance on the RFT and/or EFT and measures of each of the other indicators that have been examined in the subsequent research on differentiation, reviewed in this and the following sections of this article. Although there is some additional evidence on relations among measures of the other indicators in the more recent literature, that evidence is limited and is not considered here.

more field dependent. Olesker (1978), following the separation–individualization model of Mahler (Mahler, 1966; Mahler, Pine, & Bergman, 1975), assessed degree of self–object differentiation through extended observations of young children in a nursery school. Again, greater self–object differentiation was significantly related to greater field independence. Baraga (1978), also working within the Mahler framework, found that less attachment at age 1 was associated with greater field independence at age 3, although the expected relations between field dependence–independence and particular maternal characteristics and child behaviors were not found. Finally, Paul (1975) observed that children who showed difficulty in separating from parents after 1 week of nursery school experience were more field dependent than children who did not.

Another body of evidence has demonstrated that people who show greater autonomy of the external visual field in perception of the upright also tend to function more autonomously of others in social-interpersonal situations, particularly under conditions of ambiguity (e.g., Shulman, 1976; Solar, Davenport, & Bruehl, 1969; Weinberg, 1970). This finding broadens the picture of self-consistency in functioning across domains in ways consistent with the differentiation hypothesis.

The evidence linking autonomy in perception of the upright to autonomy of external referents in social behavior has affected our usage of the cognitive-style and field-dependence–independence concepts (Witkin & Goodenough, in press; Witkin & Goodenough, Note 5). We now apply the designation *cognitive style* to the contrasting tendencies to rely primarily on external referents or on the self in psychological functioning.[4] We have also transferred the designation *field dependence–independence* to this process dimension from its early use as a label for disembedding ability in perception.

Another important product of the more recent research on sense of separate identity is its demonstration that field-dependent people tend to have an interpersonal orientation and field-independent people an impersonal orientation (Witkin & Goodenough, 1977). Compared to field-independent people, field-dependent ones favor social over solitary situations (e.g., Coates, Lord, & Jakabovics, 1975; Crandall & Sinkeldam, 1964; Nadeau, 1969); they prefer to be physically close to others in an interaction situation (Greene, 1976; Justice, 1970; Wineman, 1974); they are selectively attentive to social sources of information (e.g., Fitz, 1971; Fitzgibbons, Goldberger, & Eagle, 1965; Konstadt & Forman, 1965); they are open in expressing their feelings and thoughts (DeMers, 1971; Greene, 1976; Sousa-Poza & Rohrberg, 1976), an approach likely to elicit similar behavior from others toward them. These attributes and behaviors seem likely to provide the person with information about what others may be feeling and thinking, as well as experience in dealing with people. They may thereby facilitate getting along with others. Probably contributing to that same end are such characteristics ascribed to field-dependent people as considerate and attentive to others (e.g., Elliott, 1961). Indeed, evidence is beginning to emerge that field-dependent people may be more effective in social interactions than more field-independent people (e.g., Oltman, Goodenough, Witkin, Freedman, & Friedman, 1975). The social attributes and behaviors useful in interpersonal relations, found among field-dependent people, and their apparently greater facility in getting along with others add up to a picture of interpersonal competencies.

Research on the social skills associated with field dependence–independence is limited, reflecting the relatively recent emergence of interest in the social domain in the history of work on psychological differentiation. More precise specification of the interpersonal competencies characteristic of relatively field-dependent people and of the social skills com-

---

[4] Other dimensions that on the surface may appear similar to field dependence–independence in fact are not. An example is the locus-of-control construct. This dimension is conceptually different from the field-dependence–independence dimension, and many studies have shown, almost invariably, that measures of the two dimensions are not related (e.g., Roodin, Broughton, & Vaught, 1974; Shapson, 1973; Tobacyk, Broughton, & Vaught, 1975). Field dependence–independence is conceived as a process variable, representing degree of autonomy of external referents in processing information from field and self; locus of control is a belief or attitudinal variable, representing greater or less fatalism as an outlook on life.

mon among more field-independent people,[5] as well as a clearer conceptualization of these competencies and skills within the differentiation framework, are tasks that lie ahead.

The involvement of cognitive restructuring skills and interpersonal competencies in field dependence–independence makes that cognitive-style dimension bipolar, in the sense of having no clear high and low ends: People at the field-independent pole have more developed cognitive restructuring skills, and people at the field-dependent pole are more likely to give evidence of interpersonal competencies. To the extent that at both poles we find characteristics adaptive in particular situations, the field-dependence–independence cognitive-style dimension is not value biased.

Segregation of Psychological Functions

The main trademark of segregation of psychological functions, a major manifestation of differentiation, is specificity of activities and experiences. Signifying the development of specificity are the formation of structured controls for routing impulses and expending energy and the formation of specialized defenses for dealing with potentially disturbing experiences. The differentiation hypothesis leads us to expect that these indicators of greater differentiation would be related to field independence.

With regard to controls, early in life, impulses are likely to find expression in diffuse systemwide reactions. As the child grows older, structured systems of control are developed that make possible the specific channeling of impulse. This accomplishment diminishes the likelihood of a "spilling over" of the content of one domain into another, thereby avoiding the too easy mixing of the ideational, affective, motivational, and motoric.

With regard to defenses, some may be considered relatively unspecialized, in the sense that they involve dealing with experience in a global fashion. Examples are denial and repression, which are characterized by a turning away from perception of the content of an immediate experience, or the memory of a past experience, often in its entirety. In the case of denial, the existence of a disturbing external reality as a whole may not be acknowledged. Repression, when used as a generalized defensive strategy, has been conceived in psychoanalytic theory to be the blocking from awareness of not only objectionable instinctual demands but the impulses, thoughts, and feelings associated with them as well. In contrast to these nonspecific ways of functioning, defenses such as isolation, intellectualization, and projection are more specialized. These defenses are characteristically directed at particular components of experience, which may be obliterated or made more pallid, or they may act to separate components of experience that are actually connected, without the components necessarily being lost. In the case of isolation, the connections between ideas and feelings (or wishes or impulses) that are actually related are lost from awareness, although the affect and ideation are consciously available to the person. Intellectualization is a defense in which affect is neglected or excluded, whereas the ideational components of the experience remain in force. Even though its affective content is washed out, the experience continues to be available to the person in its intellectual aspects as something that can be named and defined. Finally, in projection, specific characteristics and purposes are attributed to the person who is its object; it is thus not indiscriminate in its action.

To the extent that structured controls signify specialization of functions and contribute to specificity of behavior, we may expect from the differentiation hypothesis that field-independent people will provide more evidence of the availability and use of struc-

---

[5] Very little is yet known about the social skills to be found among field-independent people, beyond the fact that such people tend to be limited in interpersonal competencies. Some of these people, who have been described as aloof, distant, and solitary, are likely to stay away from interpersonal involvement as much as possible. There are indications that when field-independent people do invest in the social sphere, their social skills may in fact be expressions of their restructuring abilities, taking the form of ordering and organizing social situations, perspectivism, and accuracy in the more cognitive aspects of person perception resulting from use of an analytical approach. These skills would appear to contrast with those more common among field-dependent people that particularly involve facility in getting along with others. The evidence on these points is hardly conclusive, but it suggests directions for the research that is needed.

tured controls, compared to field-dependent people. The characteristics of specialization and specificity of defenses, such as intellectualization, isolation, and projection, would similarly lead us to expect these defenses to be more commonly used by field-independent people, and the use of more global defenses, such as repression and denial, to be more common among field-dependent people.

Structured Controls

In one set of studies, degree of structure of controls was inferred from the content of a battery of projective tests or from particular projective test responses reflecting directness of affective discharge, with controls judged to be more structured when the expression of affect was modulated or mediated (Freedman & Marks, 1965; Gardner, Holzman, Klein, Linton, & Spence, 1959; Guskin, 1955; Witkin et al., 1962/1974; Witkin et al., 1954/1972). Most of these clinical studies found the expected tendency for field-independent subjects to show more structured controls, although the evidence is not entirely consistent. In line with these findings is the observation (Witkin, Lewis, & Weil, 1968) that field-dependent patients showed more diffuse anxiety in their verbal productions during psychotherapy when these productions were assessed according to the Gottschalk-Gleser method (Gottschalk & Gleser, 1969), although these two kinds of patients were not different in overall amount of anxiety expressed.

Other studies have assessed structure of controls from observations of behavior. Crutchfield and Starkweather (Note 6) found that judges described field-dependent adults they had observed over a period of time as people who "undercontrol impulses" and "act with insufficient thinking and deliberateness." Performance on the mazes task has been observed for the evidence it provides of impulsivity versus control in carrying out a specific activity. In most studies, field-dependent subjects, both children and adults, were more likely to provide evidence of impulsive behavior than were field-independent subjects (e.g., Best, 1975; Gorman, 1968; Podell & Phillips, 1959; Swan, 1974). A related set of studies was concerned with motoric inhibition. The method used required the subject to carry out a well-practiced act—such as drawing a line, walking a line, or writing one's own name—very slowly. Many studies with children have shown capacity for motor inhibition to be greater among field-independent than field-dependent children (e.g., Eddy, 1974; Maccoby, Dowley, Hagen, & Degerman, 1965; Mayer, 1967; Mumbauer & Miller, 1970). Other studies with children did not find a relation between field dependence–independence and assessments of impulse control from general behavior manifestations (Reppucci, 1970; Seitz, 1972) or from behavior manifestations together with psychological test data (Gardner & Moriarty, 1968). On the other hand, with a high degree of consistency, hyperactive children, in whom impulsivity is a prominent characteristic, have proved to be field dependent (e.g., Campbell, Douglas, & Morgenstern, 1971; Cohen, Weiss, & Minde, 1972; Douglas, 1972; Halverson & Waldrop, 1976).

Measures of impulsivity from self-report inventories have generally proved not to be related to field dependence–independence. This outcome may be a function of particular characteristics of these inventories.[6]

---

[6] Self-report inventory measures of defenses generally do not relate to field dependence–independence either. The possibility that the contrast between the results from self-report inventories and the results obtained with projective test responses and behavioral observations may be a function of the methodologies involved is supported by several observations made in other areas of research on psychological differentiation. As one illustration, mothers' reports of their actual behavior with their children tend to distinguish significantly between the child-rearing experiences of relatively field-dependent and field-independent children, but mothers' expressed attitudes toward child rearing generally do not (Witkin & Goodenough, in press; Goodenough & Witkin, Note 7). As another example, studies of the position subjects assume in relation to those with whom they are interacting have shown that field-dependent people prefer to be closer to others than do field-independent people; on the other hand, subjects' responses to hypothetical situations involving the use of interpersonal space have not shown such a difference (Witkin & Goodenough, 1977). As still another example, whereas field-dependent people are more responsive to group standards when the standard is derived from actual interaction with a group, no difference has typically been found between field-dependent and field-independent people when the standard was given in the form of a statement attributed to either an authoritative source or a bogus group average (Witkin

By most signs, field-independent people thus give evidence of greater regulation of affective discharge and of motor activities, suggesting the presence of more structured controls.

Specialized Defenses

The nature of the defenses predominantly used by field-dependent and field-independent people has been examined in many studies employing a variety of approaches.

In one approach, clinical assessments of defenses were made. The clinical techniques used have included the Rorschach (e.g., Bertini, 1962; Levine & Spivak, 1964; Morrison & Centers, 1969; Schimek, 1968); an interview (Witkin et al., 1954, 1972) or therapy interactions (Safer, 1975); a battery of psychological tests and behavioral observations (Gardner & Moriarty, 1968); Blacky Pictures and the Sentence Completion Test (Lutzky & Schmeidler, 1963; Silverstone, 1966); and the Defense Mechanism Inventory, developed by Gleser and Ihilevich (1969) (e.g., Bogo, Winget, & Gleser, 1970; Donovan, Hague, & O'Leary, 1975; Erickson, Smyth, Donovan, & O'Leary, 1976; Williamson, 1977). In another kind of study, repressors and isolators were identified through performance on a metacontrast situation (Almgren, 1971; Nilsson & Almgren, 1970). The weight of the evidence from these studies together is that field-independent people are prone to use isolation, intellectualization, and projection as characteristic defenses, whereas field-dependent people are more likely to use repression and denial.

Evidence relevant to the issue of defenses has also come from studies of the effects of stress on recall and perception. These studies have shown that field-dependent people, compared to field-independent ones, are more likely to forget previously learned stressful material than neutral material (e.g., Duvall, 1970; Rosen, 1963; Schimek & Wachtel, 1969; Uhlmann & Saltz, 1965). Field-dependent people have also been shown to be more affected by stressful material in perception (e.g., Almgren, 1971; Duvall, 1970; Lefcourt, Gronnerud, & McDonald, 1973; Minard & Mooney, 1969).

Still another kind of evidence comes from the literature on dream forgetting, a phenomenon that has been attributed to repression. This evidence, recently reviewed by Goodenough (1976), suggests that people who characteristically fail to recall their dreams, particularly stressful ones, are likely to be field dependent.

The expectation from the differentiation hypothesis that field-independent people are more prone to use specialized defenses, such as isolation, intellectualization, and projection,[7] and field-dependent people to use less specialized defenses, such as repression and denial, has received substantial confirmation from the available evidence.[8]

Further Evidence on "Keeping Things Separate"

Greater discreteness of functions and experiences, signifying segregation and specialization of activities, has been observed among field-

& Goodenough, 1977). It appears that the most consistent relations are found between field dependence–independence and a variety of personality and social variables when assessment of these variables is made under conditions in which subjects interact with others and in which actual behavior is used as the basis of assessment. Relations are less likely to be found when hypothetical situations or attitudes are the media of assessment. To the extent that self-report inventories include items of these kinds, their predictive value in relation to field dependence–independence may be diminished.

[7] The use of more specialized defenses has the general consequence of separating the self from emotions by reducing the potency of affect, giving ascendancy to the more cognitive aspects of experience. There is an evident parallel between this emphasis on the cognitive side of psychic life among field-independent people and their impersonal orientation, with its emphasis on ideas and abstract principle. Thus, field-independent people have been characterized as highly cathecting intellectual activities (Crutchfield, Woodworth, & Albrecht, 1958.)

[8] The use of more specialized or less specialized defenses carries no implication as to effectiveness of adjustment. Thus, the use of intellectualization may be a valuable aid in the development of the individual's intellectual potential, by giving intellectual processes freedom from intrusion of feelings, impulses, and wishes that may divert these processes and impair their effectiveness. Extreme reliance on intellectualization, however, may hamper the development of the individual's emotional life. Similarly, repression may help the individual to function unhampered by threatening feelings and memories that if allowed entry into awareness, could be disruptive; extreme reliance on repression may constrict development of salient parts of the personality.

independent people in psychological domains other than controls and defenses.

One illustration comes from a study by Brilhart (1970) that examined the influence of affective features of a communication on subjects' evaluations of its cognitive content. Two messages were prepared on a given theme; one was "good" in its cognitive structure and content, the other "poor." In addition, each kind of message was delivered by a good or poor speaker. Among field-dependent subjects, but not among field-independent ones, estimates of the logic of the message were influenced by the kind of speaker who delivered it. Thus, field-independent subjects, more than field-dependent ones, kept cognitive features of the message separate from the personal context in which it was presented.

A quite different kind of evidence on the maintenance of discreteness may be found in the domain of stimulus generalization. A recent review (Goodenough, 1976) indicates that generalization along a stimulus continuum is less among field-independent than among field-dependent people. Thus, field-independent people more quickly attach the critical response to a particular stimulus in the array, inhibiting expression of that response to other stimuli.

The evidence reviewed here is, on the whole, consistent with expectations from the differentiation hypothesis, that greater field independence is likely to be associated with greater specialization of psychological functions.

Neurophysiological Differentiation

On the premise that differentiation is a broad organismic construct, we may expect greater or less differentiation to manifest itself in the neurophysiological domain as well as in the psychological domain. It is only recently that the rapidly expanding knowledge about asymmetrical specialization of the hemispheres of the brain has made it possible to formulate explicit hypotheses, within differentiation theory, regarding linkages between psychological and neurophysiological differentiation (Oltman, Note 8).

A great deal of evidence now indicates that among normal right-handed people, the left and right hemispheres have characteristically different modes of processing information (e.g., Dimond & Beaumont, 1974; Milner, 1975). The left hemisphere, served most directly by the right visual field and right ear, seems particularly suited to the processing of verbal-conceptual and executive motor functions. The mode of functioning of the right hemisphere, served most directly by the left visual field and left ear, has been characterized in ways that imply holistic processing of configurations or gestalten (e.g., Bogen, 1975). These observations lend themselves to interpretation within the differentiation framework. Specialization of the functions of the two hemispheres may be taken as an indicator of neurophysiological differentiation, much as specialization has served as an indicator of psychological differentiation. The differentiation hypothesis leads us to predict that greater specialization in the psychological domain will be linked to greater specialization in the neurophysiological domain. In particular, we may expect that field-independent people, compared to field-dependent ones, will show evidence of greater lateral specialization of the hemispheres; this will show itself in greater specialization of the left hemisphere for verbal and motor control processing, and greater specialization of the right hemisphere for configurational-gestalt processing. The emphasis, it should be noted, is on degree of lateralization of different types of processing in the respective sides, not on the dominance of one hemisphere over the other as a generalized hemisphericity tendency.

Evidence of the linkages between psychological and neurophysiological differentiation, as predicted by differentiation theory, is now available from a number of studies using a variety of approaches.

It is a well-established observation that in right-handed individuals the perception of faces is processed primarily by the right hemisphere (e.g., Hilliard, 1973; Rizzolatti, Umiltà, & Berlucchi, 1971). Using a composite face-perception technique (Gilbert & Bakan, 1973), Oltman, Ehrlichman, and Cox (1977) found a significant tendency for relatively field-independent subjects to show a stronger left visual field bias than did field-dependent subjects. Thus, when confronted with a task

that is more suited to the processing mode of the right hemisphere, the field-independent observers gave particular weight to cues from the left visual field, which project most directly to the right hemisphere. Field-dependent subjects showed little or no lateralization on this task.

Complementing the Oltman et al. (1977) study of right-hemisphere functions are studies of left-hemisphere functions. One such study, by Sousa-Poza and Rohrberg (Note 9), took as its point of departure the observation by Kimura (1973) that right-handed speakers typically make most of their speech-related gestures with their right hands. Sousa-Poza and Rohrberg found that this relationship was moderated by field dependence–independence, however. Although relatively field-independent speakers gestured primarily with their right hands, conforming to the modal pattern described by Kimura, field-dependent speakers showed more bilateral speech-related gestures.

The Kimura (1973) finding apparently reflects left-hemisphere control, in general, of both speech and speech-related right-hand movements. On this basis, the relatively greater number of bilateral speech-related movements among field-dependent people may reflect a more bilateral representation of language. Results consistent with the idea of more definite left-hemisphere language representation among field-independent people have come from studies using the dichotic listening technique. Those studies show that persons with stronger right-ear superiority in the processing of verbal material tend to be relatively field independent (Dawson, 1977; Pizzamiglio, 1974; Pizzamiglio & Cecchini, 1971; Waber, 1977b).

Assessments of extent of lateralization of left-hemisphere and right-hemisphere processing, as a function of field dependence–independence, were recently combined in a single study (Zoccolotti & Oltman, 1978). Rizzolatti et al. (1971) had earlier found that overall, discrimination reaction times were faster to letters when they were presented tachistoscopically to the right visual field (which projects directly to the left "verbal hemisphere"), and were faster to faces shown in the left visual field (which projects directly to the right "configurational hemisphere"). Using the same stimuli employed by Rizzolatti et al., Zoccolot and Oltman found this pattern of opposi visual-field superiorities, as a function of t type of processing required, among their fiel independent, but not among their field-d pendent, subjects. These two kinds of subje were not different in overall performance either task. A recent study by Rapaczyns and Ehrlichman (Note 10), using a face recog tion task, found reversed, rather than small hemifield differences in their field-depende subjects, but the complexity of the task th used and the less accurate overall performanc of their field-dependent subjects make int pretation of these results difficult.[9]

In addition to leading to differences in b havior of the kinds considered, differentiati of function of the two hemispheres may also expected to show itself more directly in diff ences in electrical activity of the hemispher The more similar the activities in the two hem spheres (or the less different their function the more similar are the electroencephalogra (EEG) recordings from the left and rig hemispheres likely to be. Oltman, Semple, a Goldstein (in press) recently found th fluctuations over time, in integrated EE amplitudes, showed significantly higher c relations between left and right hemispher among field-dependent subjects.

Still another kind of evidence, of a mo indirect nature, comes from studies of su groups that differ in degree of lateralization well as in performance on tests of field pendence–independence or on tests of spati visualization ability, one of the cognitive structuring dimensions in which field inc pendence expresses itself. For example, seve studies have shown that women tend to be l strongly lateralized than men (e.g., Kimu

---

[9] In the face-recognition data of Rapaczynski a Ehrlichman (Note 10), the field-independent subje showed the usual left-visual-field superiority, but fie dependent subjects showed a significant right-fi superiority. The task used required a verbal report o fixation digit in addition to face recognition, and fie dependent subjects showed lower overall performa on the face task. The combination of a verbal report a face discrimination in the same series of trials com cates inferences about extent of lateralization in processing of the face stimuli, particularly since level of performance also differed between the grou

1969; McGlone & Davidson, 1973; Ray, Morell, Frediani, & Tucker, 1976), and there is considerable evidence that people who are not fully right-handed are less lateralized than those who are clearly right-handed (e.g., Gilbert & Bakan, 1973; Silverman, Adevai, & McGough, 1966; Zangwill, 1960). As to the cognitive style of these less lateralized groups, women are relatively more field dependent than men (e.g., Witkin et al., 1962/1974), and they do less well on spatial tests (e.g., Maccoby & Jacklin, 1974). Similarly, people who are less than fully right-handed are relatively more field dependent than clear right-handers (e.g., Dawson, 1977; Pizzamiglio, 1974; Silverman et al., 1966), and they tend to be lower in spatial ability (Levy, 1969; Miller, 1971), although the evidence on this point is not entirely consistent (e.g., Briggs, Nebes, & Kinsbourne, 1976). The indications that women and non-right-handers are less lateralized, and relatively field dependent as well, are consistent with the proposal from the differentiation hypothesis that degree of lateralization is related to standing on the field-dependence–independence dimension.

In all, the evidence that greater neurophysiological differentiation tends to be associated with a more field-independent cognitive style lends support to the differentiation hypothesis.

Theoretical Implications

We assume that level of differentiation is a broad organismic dimension. On that basis, we have hypothesized that the tendency to function in more differentiated or less differentiated fashion is likely to be evident, to some degree, throughout an individual's psychological and neurophysiological activities. This expectation of self-consistency has received considerable support in the evidence that has been reviewed. The manifestations of greater or less differentiation found to be interrelated are quite diverse, and together cover considerable psychological and neurophysiological territory.

Beyond confirming the linkages among functions and behaviors identified in the earlier work, the newer research has shown that the differentiation hypothesis meets the requirement of any useful conception, that it be able to generate testable predictions in new domains. Now implicated in the differentiation network, through that research, are such additional domains as cerebral lateralization, cognitive restructuring, and interpersonal behavior. The differentiation construct also seems able to account for findings and relationships that cannot be accounted for by other constructs in our conceptual framework. Thus, the field-dependence–independence construct, now defined as extent of autonomous functioning, is not able to accommodate such phenomena as nature of controls and defenses or extent of cerebral lateralization, whereas the broader differentiation construct can do so.

The differentiation hypothesis is essentially a proposal to account for an association, among components of a network of functions and behaviors, responsible for a particular kind of individual self-consistency. It does not concern itself with the issue of hierarchical ordering of the components; nor does it concern itself with the issue of causal influences of the components on each other, either in the person's current functioning or during development. The issues of hierarchical ordering and of causal interconnections have recently come to the fore, stimulated in part by the addition of new areas to the differentiation network.

A hierarchical ordering primarily signifies an arrangement of constructs from more general to more specific; it may carry the implication as well that higher order constructs influence lower order ones. In our 1962 model (Figure 1), differentiation was the most general construct, and the series of postulated indicators of extent of differentiation constituted an array of narrower, lower order constructs, all of equal status. From the more recent research, proposals have emerged for amalgamating some of the lower order constructs, as well as proposals for a larger number of delineated constructs varying in degree of specificity. A plausible structure incorporating these proposals is shown in its main outlines in Figure 2 and is discussed later along with some elaborations and possible alternatives. As in our earlier model, differentiation continues to be located at the apex. At the level immediately below the apex are the three major indicators of differentiation: self–nonself segregation, segre-

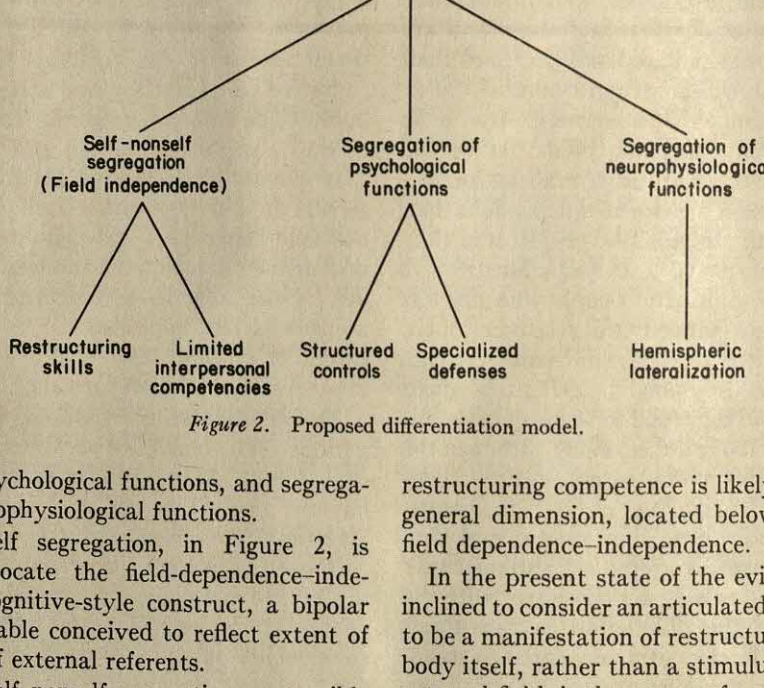


Figure 2. Proposed differentiation model.

gation of psychological functions, and segregation of neurophysiological functions.

Self–nonself segregation, in Figure 2, is where we locate the field-dependence–independence cognitive-style construct, a bipolar process variable conceived to reflect extent of autonomy of external referents.

Limited self–nonself segregation, responsible for less autonomous functioning or a field-dependent cognitive style, signifies continued connectedness with others. Such connectedness, it seems reasonable to postulate, fosters the development of an interpersonal orientation, "fellow feeling," a fund of knowledge about others, and facility in getting along with others—in other words, interpersonal competencies. Interpersonal competencies are, in this view, a derivative of greater connectedness with others, and hence a construct below the level of self–nonself segregation or field dependence–independence.

It also seems possible that cognitive-restructuring abilities may be a product of field independence. Whether a person tends to rely primarily on external referents or to be self-reliant might be expected to influence the development of cognitive-restructuring skills. A person who functions less autonomously may adhere to a percept or symbolic representation as given when dealing with restructuring tasks. A more autonomously functioning person may go beyond the information given when that is required by situational demands or inner needs. If this view of the relationship between autonomy and restructuring is correct, then restructuring competence is likely to be a ve general dimension, located below the level field dependence–independence.

In the present state of the evidence, we a inclined to consider an articulated body conce to be a manifestation of restructuring when t body itself, rather than a stimulus array in t external field, is the source of experience, as the examples of restructuring consider earlier. Articulation emphasizes the mo cognitive aspects of the body concept, and pa ticularly the person's view of the spatial repr sentation of the body. Though formation the body concept is a process that takes pla over the course of development, it may be se to entail the imposition of structure on stimulus array much as do some of the pro lem-solving tasks we considered, which call f structuring at the moment they are carried ou In suggesting that an articulated body co cept be subsumed under cognitive restructu ing, we are changing the location of that co struct from the higher order status it had our earlier model in Figure 1.

Consistent with the concept that extent autonomy may exert a causal influence on t development of interpersonal competencies an restructuring skills is the evidence from a larg body of cross-cultural research (e.g., Witki 1978; Witkin & Berry, 1975; Witkin et a 1974). It appears that societies that encourag autonomy from societal and parental authorit in young children, in comparison to societie that stress conformity, tend to produce peopl who show the associated characteristics

field independence in perception of the upright, competence in cognitive restructuring, and self-reliance. The contrasting set of characteristics is more common in societies that encourage conformity. Extent of socialization for autonomy is thus implicated as central to the development of this cluster of characteristics.

We have proposed that the greater self–nonself polarity and more autonomous functioning of field-independent people may foster the development of cognitive restructuring skills among them, but it is not as likely to foster the development of interpersonal competencies. We have also proposed that the greater recourse of field-dependent people to external referents stimulates the development of interpersonal competencies, but may be responsible for these people's lesser cognitive restructuring skills. Relatively field-dependent and field-independent people may thus be seen as making their main developmental investment in different domains, with the result that their psychological development proceeds along different pathways. This conception views the development of differentiation as multilinear; genuine development takes place along both the pathways of limited differentiation and greater differentiation. This conception also implies that cognitive styles, which are process variables, influence the development of patterns of abilities—cognitive restructuring skills and interpersonal competencies—and so may be regarded as expressing themselves in these abilities.

The second major indicator of differentiation, segregation of psychological functions, is manifested in other ways, in structured controls and specialized defenses. These manifestations thus have the status of more specific constructs below the level of segregation of psychological functions.

The third major indicator of differentiation, segregation of neurophysiological functions, is located at the same level in our model as the psychological indicators, segregation of psychological functions and self–nonself segregation. Below the level of neurophysiological segregation we place the construct lateralization of cerebral functions, in which neurophysiological segregation manifests itself.

Within the differentiation framework, segregation of neurophysiological functions may provide causal pathways to the development of the psychological manifestations of differentiation, in addition to the socialization and experiential pathways that have been suggested at various points. (To avoid undue complexity, these pathways are not shown in Figure 2.) Psychological functions are rooted in a substratum of neurophysiological activity; therefore, specialization of functions at a neurophysiological level may be an important determinant of specialization of psychological functions. The possible influence of neurophysiological specialization on the development of cognitive restructuring and of specialized defenses may be examined as illustrative.

With regard to cognitive restructuring, one might suppose that the processing modes that are characteristic of the left and right hemispheres are maximally effective if they are localized in separate hemispheres. Levy (1969, 1974) has proposed that when language representation is bilateral, verbal functions will predominate at the expense of configurational-gestalt processes. It may be, as Teuber (1975) has suggested, that speech is established earlier and is more resilient, thereby attaining a favored position wherever it is located. That is, extent of hemispheric specialization may have little effect on verbal competence, but it is likely to facilitate spatial-configurational processes ordinarily mediated by the right hemisphere, either alone or in complex interaction with the left hemisphere's propositional mode. Although the scope of the cognitive-restructuring dimension has not yet been conclusively established, the weight of the evidence currently available suggests that the cognitive correlates of field independence in perception of the upright may be limited to spatial-visual restructuring abilities. If this suggestion is confirmed by future research, Levy's hypothesis may help to explain such a possible limitation, and individual differences in spatial-restructuring abilities may be understood as a consequence, in part, of degree of cerebral lateralization.[10]

---

[10] Waber (1977a), working within the framework of Levy's (1969) hypothesis, has recently proposed a model that implicates hormonal factors. There is evidence, although it is not entirely consistent, that individuals of both sexes who reach puberty relatively

Turning to the possible influence of cerebral specialization on specialization of defenses, we have proposed that greater differentiation manifests itself in clear separation between affect and ideas or percepts. This separation may perhaps be related to the newly emerging evidence that the right hemisphere, in addition to mediating spatial-configurational processes, may also be implicated in the expression and control of affect (Carmon & Nachshon, 1973; Gainotti, 1972; Galin, 1974; Heilman, Scholes, & Watson, 1975; Schwartz, Davidson, & Maer, 1975). If there is a special role for the right hemisphere in affective experience and expression, we might then ask whether greater or less cerebral lateralization may enter into determining how affect and ideation are interrelated. It is perhaps possible that one of the bases for the relative lack of separation of feelings and ideas among less differentiated people is a neural organization in which parts of the brain ordinarily involved in their processing are not as specialized for these purposes. Such a possibility seems consistent with the differentiation framework and is capable of generating testable predictions.

The possible involvement of cerebral specialization in the psychological components of the differentiation cluster suggests that causal influences may cross the three main pathways radiating from differentiation, in the model sketched in Figure 2, making for multidetermination of these components. In still broader perspective, it may well be that the socialization practices that encourage self–nonself segregation and individual autonomy (see Witkin & Goodenough, in press, and Goodenough & Witkin, Note 7, for recent reviews) have much in common with those that encourage the development of such expressions of segregation of psychological functions as structured controls and specialized defenses (Dyk, 1969; Dyk & Witkin, 1965; Witkin et al., 1962/1974). Moreover, socialization practices may well influence the development of neurophysiological differentiation, which in turn may have its own consequences for a variety of specific psychological functions implicated in the differentiation framework. That socialization experiences may indeed affect the development of lateralization is suggested by the finding of Kimura (1967) and Geffner and Hochberg (1971) that their lower socioeconomic status (SES) children developed the usual right-ear advantage in dichotic listening later than did middle and upper SES children. Although we do not yet know the precise nature of the operative variables involved, or their overlap with the socialization variables implicated in the development of self–nonself segregation and segregation of psychological functions, it seems that experience may be one important source of individual differences in neurophysiological differentiation.

Complementing these various routes of causal influence are possibilities for more directly acting causal effects. Thus, training has been shown to enhance specific restructuring skills such as disembedding (Witkin & Goodenough, in press; Goodenough, & Witkin, Note 7), and it may affect restructuring as a general competence as well (Dolecki, 1976). The development of interpersonal competencies is, of course, also subject to direct influence by appropriate social experiences.

Much progress has been made over the years in elaborating and making more specific the propositions of differentiation theory and enlarging its empirical underpinning. With this expansion have come new questions that now require answers, an inevitable accompaniment of the evolution of any theory. The varied directions for further work that have been opened by recent research on differentiation theory provide an encouraging sign of the theory's continuing usefulness.

early tend to be less lateralized and less competent in cognitive restructuring, although not different in verbal functioning, as compared to late maturers (e.g., Broverman, Broverman, Vogel, Palmer, & Klaiber, 1964; Petersen, 1976; Waber, 1977b). Waber suggests that the relationship between pubertal age and restructuring ability is mediated by hormonal influences on the development of hemispheric specialization.

Reference Notes

1. Cox, P. W., & Witkin, H. A. *Field-dependence-independence and psychological differentiation: Bibliography with index* (Suppl. 3, ETS RB 78-8). Princeton, N.J.: Educational Testing Service, 1978.
2. Witkin, H. A., Cox, P. W., & Friedman, F. *Field-dependence–independence and psychological differentiation: Bibliography with index* (Suppl. 2, ETS RB 76-28). Princeton, N.J.: Educational Testing Service, 1976.

3. Witkin, H. A., Cox, P. W., Friedman, F., Hrishikesan, A. G., & Siegel, K. N. *Field-dependence–independence and psychological differentiation: Bibliography with index* (Suppl. 1, ETS RB 74-42). Princeton, N.J.: Educational Testing Service, 1974.
4. Witkin, H. A., et al. *Field-dependence–independence and psychological differentiation: A bibliography through 1972 with index* (ETS RB 73-62). Princeton, N.J.: Educational Testing Service, 1973.
5. Witkin, H. A., & Goodenough, D. R. *Field dependence revisited* (ETS RB 77-16). Princeton, N.J.: Educational Testing Service, 1977).
6. Crutchfield, R. S., & Starkweather, J. A. *Differences among officer personnel in perception of the vertical under distorting influence of a tilted frame* (Project No. 505-041-0001). Berkeley: University of California, Institute of Personality Assessment and Research, December 1953.
7. Goodenough, D. R., & Witkin, H. A. *Origins of the field-dependent and field-independent cognitive styles* (ETS RB 77-9). Princeton, N.J.: Educational Testing Service, 1977.
8. Oltman, P. K. Field independence and extent of lateralization. In B. D. Cohen (Chair), *Psychophysiological studies of field dependence–independence*. Symposium presented at the meeting of the American Psychological Association, Washington, D.C., September 1976.
9. Sousa-Poza, J. F., & Rohrberg, R. *Field dependence and lateralization of hand movements during speech.* Unpublished manuscript, Laboratoire de Communication, Département de Psychiatrie, Centre Hospitalier Universitaire, Sherbrooke, Québec, Canada, 1974.
10. Rapaczynski, W., & Ehrlichman, H. Opposite visual hemisfield superiorities in face recognition as a function of cognitive style. *Neuropsychologia*, in press.

References

Adams, R. C. Perceptual correlates of the rod-and-frame test: A critical response. *Perceptual and Motor Skills*, 1974, *38*, 1044–1046.

Adevai, G., Silverman, A. J., & McGough, W. E. Perceptual correlates of the rod-and-frame test. *Perceptual and Motor Skills*, 1968, *26*, 1055–1064.

Almgren, P. E. Relations between perceptual defenses, defined by the metacontrast technique, and adaptive patterns in two serial behavior tests. *Psychological Research Bulletin*, 1971, *11*, 1–22.

Bachelis, L. A. Body-field perceptual differentiation as a variable in creative thinking (Doctoral dissertation, Yeshiva University, 1965). *Dissertation Abstracts*, 1965, *26*, 3475. (University Microfilms No. 65-11, 977)

Baraga, E. S. The relationship of social and maturational variables to development of field-dependence in preschoolers: A longitudinal study (Doctoral dissertation, University of North Dakota, 1977). *Dissertation Abstracts International*, 1978, *39*, 351B–352B. (University Microfilms No. 78-10313)

Bertini, M. Il tratto difensivo dell'isolamento nella sua determinazione dinamica e strutturale. *Contributi dell'Instituto di Psicologia*, 1962, *24*, 307–326.

Best, P. K. Psychological differentiation in the deaf (Doctoral dissertation, Wayne State University, 1974). *Dissertation Abstracts International*, 1975, *35*, 4239A. (University Microfilms No. 74-29, 786)

Bogen, J. E. Some educational aspects of hemispheric specialization. *UCLA Educator*, 1975, *17*, 24–32.

Bogo, N., Winget, C., & Gleser, C. G. Ego defenses and perceptual styles. *Perceptual and Motor Skills*, 1970, *30*, 599–605.

Briggs, G. G., Nebes, R. D., & Kinsbourne, M. Intellectual differences in relation to personal and family handedness. *Quarterly Journal of Experimental Psychology*, 1976, *28*, 591–601.

Brilhart, B. L. Relationships of speaker-message perception to perceptual field-independence. *Journal of Communication*, 1970, *20*, 153–166.

Broverman, D. M., Broverman, I. K., Vogel, W., Palmer, R. D., & Klaiber, E. L. The automatization cognitive style and physical development. *Child Development*, 1964, *35*, 1343–1359.

Busse, T. V. Establishment of the flexible thinking factor in fifth-grade boys. *Journal of Psychology*, 1968, *69*, 93–100.

Campbell, S. B., Douglas, V. I., & Morgenstern, G. Cognitive styles in hyperactive children and the effect of methylphenicate. *Journal of Child Psychology and Psychiatry*, 1971, *12*, 55–67.

Carlson, K., Tucker, G., Harrow, M., & Quinlan, D. Body image and mental illness: A study of the human figure drawings of psychiatric patients. *Psychiatry and Art*, 1971, *3*, 162–169.

Carmon, A., & Nachshon, I. Ear asymmetry in perception of emotional nonverbal stimuli. *Acta Psychologica*, 1973, *37*, 351–357.

Coates, S., Lord, M., & Jakabovics, E. Field dependence–independence, social-non-social play and sex differences in preschool children. *Perceptual and Motor Skills*, 1975, *40*, 195–202.

Cohen, N. J., Weiss, G., & Minde, K. Cognitive styles in adolescents previously diagnosed as hyperactive. *Journal of Child Psychology and Psychiatry*, 1972, *13*, 203–209.

Cohen, S. I., Silverman, A. J., & Shmavonian, B. M. Psychophysiological studies in altered sensory environments. *Journal of Psychosomatic Research*, 1962, *6*, 259–281.

Crandall, V. J., & Sinkeldam, C. Children's dependent and achievement behaviors in social situations and their perceptual field dependence. *Journal of Personality*, 1964, *32*, 1–22.

Crutchfield, R. S., Woodworth, D. G., & Albrecht, R. E. *Perceptual performance and the effective person* (WADC-TN-58-60). Lackland Air Force Base, Tex.: Personnel Laboratory, Wright Air Development Center, Air Research and Development Command, April 1958. (ASTIA No. AD-151 039.)

Culver, C. M. Test of right–left discrimination. *Perceptual and Motor Skills*, 1969, *29*, 863–867.

Culver, C. M., Cohen, S. I., Silverman, A. J., & Shmavonian, B. M. Cognitive structuring, field dependence–independence, and the psychophysiological

response to perceptual isolation. In J. Wortis (Ed.), *Recent advances in biological psychiatry* (Vol. 4). New York: Plenum Press, 1964.

Dawson, J. L. M. Alaskan Eskimo hand, eye, auditory dominance and cognitive style. *Psychologia*, 1977, *20*, 121–135.

DeMers, N. A. *Predicting parental and marital empathic development*. Unpublished master's thesis, University of Waterloo, Ontario, Canada, 1971.

Dimond, S. J., & Beaumont, J. G. (Eds.). *Hemisphere function in the human brain*. New York: Halstead Press, 1974.

Dinius, S. H. The role of personality characteristics in the choice of accounting as a major and in success in accounting courses (Doctoral dissertation, Auburn University, 1974). *Dissertation Abstracts International*, 1975, *35*, 3608B. (University Microfilms No. 75-556)

Dolecki, P. G. The effects of a perceptual training program on conservation task performance and field independence in first grade children (Doctoral dissertation, Rutgers University, 1976). *Dissertation Abstracts International*, 1976, *37*, 878A. (University Microfilms No. 76-17, 311)

Donovan, D. M., Hague, W. H., & O'Leary, M. R. Perceptual differentiation and defense mechanisms in alcoholics. *Journal of Clinical Psychology*, 1975, *31*, 356–359.

Douglas, V. I. Stop, look and listen: The problem of sustained attention and impulse control in hyperactive and normal children. *Canadian Journal of Behavioural Science*, 1972, *4*, 259–282.

Duvall, N. S. Field articulation and the repression-sensitization dimension in perception and memory (Doctoral dissertation, University of North Carolina at Chapel Hill, 1969). *Dissertation Abstracts International*, 1970, *30*, 3864B. (University Microfilms No. 70-3228)

Dyk, R. B. An exploratory study of mother–child interaction in infancy as related to the development of differentiation. *Journal of the American Academy of Child Psychiatry*, 1969, *8*, 657–691.

Dyk, R. B., & Witkin, H. A. Family experiences related to the development of differentiation in children. *Child Development*, 1965, *30*, 21–55.

Eddy, S. The relationship of field articulation to delay capacity in children (Doctoral dissertation, Boston University, 1974). *Dissertation Abstracts International*, 1974, *35*, 2988B. (University Microfilms No. 74-26, 441)

Elliott, R. Interrelationships among measures of field dependence, ability, and personality traits. *Journal of Abnormal and Social Psychology*, 1961, *63*, 27–36.

Epstein, L. The relationship of certain aspects of the body-image to the perception of the upright. (Doctoral dissertation, New York University, 1958). *Dissertation Abstracts*, 1961, *22*, 639. (University Microfilms No. 61-2626)

Erickson, R. C., Smyth, L., Donovan, D. M., & O'Leary, M. R. Psychopathology and defensive style of alcoholics as a function of congruence–incongruence between psychological differentiation and locus of control. *Psychological Reports*, 1976, *39*, 51–54.

Faterson, H. F., & Witkin, H. A. Longitudinal study of development of the body concept. *Developmental Psychology*, 1970, *2*, 429–438.

Fisher, S., & Cleveland, S. E. *Body image and personality*. Princeton, N. J.: Van Nostrand, 1958.

Fitz, R. J. The differential effects of praise and censure on serial learning as dependent on locus of control and field dependency (Doctoral dissertation, Catholic University of America, 1970). *Dissertation Abstracts International*, 1971, *31*, 4310B. (University Microfilms No. 71-1457)

Fitzgibbons, D., Goldberger, L., & Eagle, M. Field dependence and memory for incidental material. *Perceptual and Motor Skills*, 1965, *21*, 743–749.

Freedman, S. J., & Marks, P. A. Visual imagery produced by rhythmic photic stimulation: Personality correlates and phenomenology. *British Journal of Psychology*, 1965, *56*, 95–112.

Gainotti, G. Emotional behavior and hemispheric side of the lesion. *Cortex*, 1972, *8*, 41–55.

Galin, D. Implications for psychiatry of left and right cerebral specialization. *Archives of General Psychiatry*, 1974, *31*, 572–583.

Gardner, R. W., Holzman, P. S., Klein, G. S., Linton, H. P., & Spence, D. P. Cognitive control: A study of individual consistencies in cognitive behavior. *Psychological Issues*, 1959, *1*(4), 1–86.

Gardner, R. W., Jackson, D. N., & Messick, S. J. Personality organization in cognitive controls and intellectual abilities. *Psychological Issues*, 1960, *2*(4), (Monograph 8).

Gardner, R. W., & Moriarty, A. *Personality development at preadolescence*. Seattle: University of Washington Press, 1968.

Gaughran, E. P. Physiognomic perception: A study of a cognitive style in relation to qualities of body experience (Doctoral dissertation, New York University, 1963). *Dissertation Abstracts*, 1964, *24*, 4288. (University Microfilms No. 63-7215)

Geffner, D. S., & Hochberg, I. Ear laterality performance of children from low and middle socioeconomic levels on a verbal dichotic listening task. *Cortex*, 1971, *2*, 193–203.

Gilbert, C., & Bakan, P. Visual asymmetry in perception of faces. *Neuropsychologia*, 1973, *11*, 355–362.

Gleser, G. C., & Ihilevich, D. An objective instrument for measuring defense mechanisms. *Journal of Consulting and Clinical Psychology*, 1969, *33*, 51–60.

Goodenough, D. R. The role of individual differences in field dependence as a factor in learning and memory. *Psychological Bulletin*, 1976, *83*, 675–694.

Goodenough, D. R., & Karp, S. A. Field dependence and intellectual functioning. *Journal of Abnormal and Social Psychology*, 1961, *63*, 241–246.

Gorman, B. S. Field dependence and visual maze learning. *Perceptual and Motor Skills*, 1968, *27*, 142.

Gottschalk, L. A., & Gleser, G. C. *The measurement of psychological states through the content analysis of verbal behavior*. Berkeley: University of California Press, 1969.

Gough, H. G., & Olton, R. M. Field independence as related to nonverbal measures of perceptual performance and cognitive ability. *Journal of Consulting and Clinical Psychology*, 1972, *38*, 338–342.

Greene, L. R. Effects of field dependence on affective reactions and compliance in dyadic interactions. *Journal of Personality and Social Psychology*, 1976, *34*, 569–577.

Guskin, S. L. *Individual differences in dependence on visual or kinesthetic cues*. Unpublished master's thesis, University of North Carolina at Chapel Hill, 1955.

Halverson, C. F., & Waldrop, M. F. Relations between preschool activity and aspects of intellectual and social behavior at age 7½. *Developmental Psychology*, 1976, *12*, 107–112.

Heilman, K. M., Scholes, R., & Watson, R. T. Auditory affective agnosia: Disturbed comprehension of affective speech. *Journal of Neurology, Neurosurgery, and Psychiatry*, 1975, *38*, 69–72.

Hilliard, R. D. Hemispheric laterality effects on a facial recognition task in normal subjects. *Cortex*, 1973, *9*, 246–258.

Jacobson, G. R. Effects of sensory reduction on measures of body image and perceptual orientation. *Virginia Journal of Science*, 1965, *16*, 411.

Jacobson, J. R. A method of psychobiologic evaluation. *American Journal of Psychiatry*, 1944, *101*, 343–348.

Justice, M. T. Field dependency, intimacy of topic and interperson distance (Doctoral dissertation, University of Florida, 1969). *Dissertation Abstracts International*, 1970, *31*, 395B–396B. (University Microfilms No. 70-12, 243)

Kimura, D. Functional asymmetry of the brain in dichotic listening. *Cortex*, 1967, *3*, 163–178.

Kimura, D. Spatial localization in left and right visual fields. *Canadian Journal of Psychology*, 1969, *23*, 445–458.

Kimura, D. Manual activity during speaking. I: Right-handers. *Neuropsychologia*, 1973, *11*, 45–50.

Konstadt, N., & Forman, E. Field dependence and external directedness. *Journal of Personality and Social Psychology*, 1965, *1*, 490–493.

Lefcourt, H. M., Gronnerud, P., & McDonald P. Cognitive activity and hypothesis formation during a double entendre word association test as a function of locus of control and field dependence. *Canadian Journal of Behavioural Science*, 1973, *5*, 161–173.

Levine, M., & Spivak, G. *The Rorschach index of repressive style*. Springfield, Ill.: Charles C Thomas, 1964.

Levy, J. Possible basis for the evolution of lateral specialization of the human brain. *Nature*, 1969, *224*, 614–615.

Levy, J. Psychobiological implications of bilateral asymmetry. In S. J. Dimond & J. G. Beaumont (Eds.), *Hemisphere function in the human brain*. New York: Halstead Press, 1974.

Lutzky, H. C., & Schmeidler, G. R. Perceptual leveling in institutional and other children. *Journal of Genetic Psychology*, 1963, *103*, 45–51.

Maccoby, E. E., Dowley, E. M., Hagen, J. W., & Degerman, R. Activity level and intellectual functioning in normal preschool children. *Child Development*, 1965, *36*, 761–770.

Maccoby, E. E., & Jacklin, C. N. *The psychology of sex differences*. Stanford, Calif.: Stanford University Press, 1974.

Mahler, M. Notes on the development of basic moods: The depressive affect. In R. M. Lowenstein, L. M. Newman, M. Schur, & A. J. Solnit (Eds.), *Psychoanalysis—A general psychology*. New York: International Universities Press, 1966.

Mahler, M., Pine, F., & Bergman, A. *The psychological birth of the human infant, symbiosis and individuation*. New York: Basic Books, 1975.

Mayer, A. D. Preferences in level of conceptualization as related to motor inhibition and field dependence (Doctoral dissertation, Columbia University, 1967). *Dissertation Abstracts*, 1967, *28*, 2628B. (University Microfilms No. 67-15, 501)

McGlone, J., & Davidson, W. The relation between cerebral speech laterality and spatial ability with special reference to sex and hand preference. *Neuropsychologia*, 1973, *11*, 105–113.

Miller, E. Handedness and the pattern of human ability. *British Journal of Psychology*, 1971, *62*, 111–112.

Milner, B. (Ed.). *Hemispheric specialization and interaction*. Cambridge, Mass.: MIT Press, 1975.

Minard, J. G., & Mooney, W. Psychological differentiation and perceptual defense: Studies of the separation of perception from emotion. *Journal of Abnormal Psychology*, 1969, *74*, 131–139.

Morrison, D., & Centers, L. Investigation of the convergent validity of different measures of cognitive style. *Journal of Projective Techniques and Personality Assessment*, 1969, *33*, 168–172.

Mumbauer, C. C., & Miller, J. O. Socioeconomic background and cognitive functioning in preschool children. *Child Development*, 1970, *41*, 471–480.

Nadeau, G. H. Cognitive style in preschool children: A factor analytic study (Doctoral dissertation, University of Minnesota, 1968). *Dissertation Abstracts*, 1969, *29*, 2623B. (University Microfilms No. 68-17, 702)

Nilsson, A., & Almgren, P. A. Para-natal emotional adjustment: A prospective investigation of 165 women. Part II: The influence of background factors, psychiatric history, parental relations, and personality characteristics. *Acta Psychiatrica Scandinavica*, 1970, Suppl. 220, 64–141.

Olesker, W. Cognition and the separation-individuation process: A study of three boys at nursery school. *Psychoanalysis and Contemporary Thought*, 1978, *1*, 237–268.

Oltman, P. K., Ehrlichman, H., & Cox, P. W. Field independence and laterality in the perception of faces. *Perceptual and Motor Skills*, 1977, *45*, 255–260.

Oltman, P. K., Goodenough, D. R., Witkin, H. A., Freedman, N., & Friedman, F. Psychological differentiation as a factor in conflict resolution. *Journal of Personality and Social Psychology*, 1975, *32*, 730–736.

Oltman, P. K., Semple, C., & Goldstein, L. Cognitive style and interhemispheric differentiation in the EEG. *Neuropsychologia*, in press.

Paquette, E. C. Awareness and accuracy of perception of physiological change: Personality and perceptual correlates (Doctoral dissertation, University of Houston, 1973). *Dissertation Abstracts International*, 1974, *34*, 5688B–5689B. (University Microfilms No. 74-11, 828)

Pascual-Leone, J. *Cognitive development and cognitive style: A general psychological integration.* Unpublished doctoral dissertation, University of Geneva, Geneva, Switzerland, 1969.

Paul, E. *A study of the relationship between separation and field-dependency in a group of three-year-old nursery-school children.* Unpublished master's thesis, Bank Street College of Education, New York, 1975.

Petersen, A. C. Physical androgyny and cognitive functioning in adolescence. *Developmental Psychology*, 1976, *12*, 524–533.

Pizzamiglio, L. Handedness, ear-preference, and field-dependence. *Perceptual and Motor Skills*, 1974, *38*, 700–702.

Pizzamiglio, L., & Carli, R. Caratteristiche psicometriche di alcuni test di dipendenza-indipendenza dal campo. *Archivio di Psicologia, Neurologia e Psichiatria* 1973, *34*, 276–286.

Pizzamiglio, L., & Cecchini, M. Development of the hemispheric dominance in children from 5 to 10 years-of-age and their relations with the development of cognitive processes. *Brain Research*, 1971, *31*, 361–378.

Podell, J. E., & Phillips, L. A developmental analysis of cognition as observed in dimensions of Rorschach and objective test performance. *Journal of Personality*, 1959, *27*, 439–463.

Ray, W. J., Morell, M., Frediani, A. W., & Tucker, D. Sex differences and lateral specialization of hemispheric functioning. *Neuropsychologia*, 1976, *14*, 391–394.

Reppucci, N. D. Individual differences in the consideration of information among two-year-old children. *Developmental Psychology*, 1970, *2*, 240–246.

Rizzolatti, G., Umiltà, C., & Berlucchi, G. Opposite superiorities of the right and left cerebral hemispheres in discriminative reaction time to physiognomical and alphabetical material. *Brain*, 1971, *94*, 431–442.

Roodin, P. A., Broughton, A., & Vaught, G. M. Effects of birth order, sex, and family size on field dependence and locus of control. *Perceptual and Motor Skills*, 1974, *39*, 671–676.

Rosen, H. *Recall as a function of differentiation and hostility.* Unpublished doctoral dissertation, Wayne State University, Detroit, Mich., 1963.

Safer, J. Effects of sex and psychological differentiation on responses to a stressful group situation (Doctoral dissertation, New School for Social Research, 1975). *Dissertation Abstracts International*, 1975, *36*, 3068B–3069B. (University Microfilms No. 75-25, 458)

Schimek, J. G. Cognitive style and defenses: A longitudinal study of intellectualization and field independence. *Journal of Abnormal Psychology*, 1968, *73*, 575–580.

Schimek, J. G., & Wachtel, P. L. Exploration of effects of distraction, competing tasks and cognitive style in attention deployment. *Perceptual and Motor Skills*, 1969, *28*, 567–580.

Schuck, J. R., Cross, H. A., & Mills, D. H. A signal detection analysis of the rod-and-frame test. *Perception & Psychophysics*, 1970, *7*, 276–280.

Schwartz, G. E., Davidson, R. J., & Maer, F. Right hemisphere lateralization for emotion in the human brain: Interactions with cognition. *Science*, 1975, *190*, 286–288.

Seitz, E. K. The relationship between cognitive abilities and impulse control in Project Headstart children (Doctoral dissertation, New York University, 1971). *Dissertation Abstracts International*, 1972, *32*, 6061B. (University Microfilms No. 72-13, 402)

Shapson, S. M. Hypothesis testing and cognitive style in children (Doctoral dissertation, York University, Toronto, Canada, 1973). *Dissertation Abstracts International*, 1976, *36*, 5317B–5318B.

Shulman, E. Conformity in a modified Asch-type situation (Doctoral dissertation, City University of New York, 1975). *Dissertation Abstracts International*, 1976, *36*, 6455B. (University Microfilms No. 76-11, 971)

Siegel, H. B. Body boundary and field-dependence. *Perceptual and Motor Skills*, 1977, *45*, 1097–1098.

Silverman, A. J., Adevai, G., & McGough, W. E. Some relationships between handedness and perception. *Journal of Psychosomatic Research*, 1966, *10*, 151–158.

Silverman, A. J., Cohen, S. I., Shmavonian, B. M., & Greenberg, G. Psychophysiological investigations in sensory deprivation: The body-field dimension. *Psychosomatic Medicine*, 1961, *23*, 48–62.

Silverstone, S. Structural personality differences in essential hypertension and peptic ulcer (Doctoral dissertation, Yeshiva University, 1966). *Dissertation Abstracts*, 1966, *27*, 604B–605B. (University Microfilms No. 66-7667)

Solar, D., Davenport, G., & Bruehl, D. Social compliance as a function of field dependence. *Perceptual and Motor Skills*, 1969, *29*, 299–306.

Sousa-Poza, J. F., & Rohrberg, R. Communicational and interactional aspects of self-disclosure in psychotherapy: Differences related to cognitive style. *Psychiatry*, 1976, *39*, 81–91.

Swan, G. A. Machiavellianism, impulsivity, field dependence-independence, and performance on the Prisoners' Dilemma game (Doctoral dissertation, Wayne State University, 1973). *Dissertation Abstracts International*, 1974, *34*, 5695B. (University Microfilms No. 74-11, 167)

Teuber, H. L. Why two brains? In B. Milner (Ed.), *Hemispheric specialization and interaction.* Cambridge, Mass.: MIT Press, 1975.

Thurstone, L. L. Primary mental abilities. *Psychometric Monographs*, 1938, No. *1*, 1–121.

Tobacyk, J. J., Broughton, A., & Vaught, G. M. Effects of congruence–incongruence between locus of control and field dependence on personality functioning. *Journal of Consulting and Clinical Psychology*, 1975, *43*, 81–85.

Uhlmann, F. W., & Saltz, E. Retention of anxiety material as a function of cognitive differentiation. *Journal of Personality and Social Psychology*, 1965, *1*, 55–62.

Waber, D. P. Biological substrates of field dependence: Implications of the sex difference. *Psychological Bulletin*, 1977, *84*, 1076–1087. (a)

Waber, D. P. Sex differences in mental abilities, hemispheric lateralization, and rate of physical growth at

adolescence. *Developmental Psychology*, 1977, *13*, 29–38. (b)

Weinberg, H. J. Changing perceptions on the RFT by conditioning subjects to relieve dissonance and/or escape from the anxiety in a new manner (Doctoral dissertation, University of Nebraska, 1970). *Dissertation Abstracts International*, 1970, *31*, 1637A. (University Microfilms No. 70-17, 768)

Williamson, J. H. Field dependence-independence, internal-external control, and chief mode of defense in male high school seniors (Doctoral dissertation, St. John's University, New York, 1976). *Dissertation Abstracts International*, 1977, *37*, 6383A. (University Microfilms No. 77-1605)

Wineman, J. H. Projected utilization of interpersonal distance as a function of psychological differentiation (Doctoral dissertation, University of Utah, 1973). *Dissertation Abstracts International*, 1974, *34*, 5698B–5699B. (University Microfilms No. 74-11, 431)

Winestine, M. C. Twinship and psychological differentiation. *Journal of the American Academy of Child Psychiatry*, 1969, *8*, 436–455.

Witkin, H. A. Psychological differentiation and forms of pathology. *Journal of Abnormal Psychology*, 1965, *70*, 317–336.

Witkin, H. A. *Cognitive styles in personal and cultural adaptation. The 1977 Heinz Werner Lectures*. Worcester, Mass.: Clark University Press, 1978.

Witkin, H. A., & Berry, J. W. Psychological differentiation in cross-cultural perspective. *Journal of Cross-Cultural Psychology*, 1975, *6*, 4–87.

Witkin, H. A., Dyk, R. B., Faterson, H. F., Goodenough, D. R., & Karp, S. A. *Psychological differentiation*. Potomac, Md.: Erlbaum, 1974. (Originally published, Wiley, 1962.)

Witkin, H. A., & Goodenough, D. R. Field dependence and interpersonal behavior. *Psychological Bulletin*, 1977, *84*, 661–689.

Witkin, H. A., & Goodenough, D. R. The essence and origins of cognitive styles: Field dependence and field independence. *Psychological Issues* (Monograph). New York: International Universities Press, in press.

Witkin, H. A., et al. *Personality through perception*. Westport, Conn.: Greenwood Press, 1972. (Originally published, Harper, 1954.)

Witkin, H. A., Lewis, H. B., & Weil, E. Affective reactions and patient-therapist interactions among more differentiated and less differentiated patients early in therapy. *Journal of Nervous and Mental Disease*, 1968, *146*, 193–208.

Witkin, H. A., et al. Social conformity and psychological differentiation. *International Journal of Psychology*, 1974, *9*, 11–29.

Zangwill, O. L. *Cerebral dominance and its relation to psychological function*. Edinburgh, Scotland: Oliver & Boyd, 1960.

Zoccolotti, P., & Oltman, P. K. Field dependence and lateralization of verbal and configurational processing. *Cortex*, 1978, *14*, 155–163.

Received April 20. 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 7, 1146–1152

Self As Actor, Active Observer, and Passive Observer: Implications for Causal Attributions

John D. Cunningham
School of Behavioural Sciences
Macquarie University, Sydney, Australia

Philip A. Starr
University of
Southern California

David E. Kanouse
Rand Corporation
Santa Monica, California

Miller and Norman have distinguished between active observers (those on the receiving end of an actor's behavior) and passive observers (onlookers of an event involving an actor and an active observer). Following the concept of hedonic relevance, it was hypothesized that active observers will attribute the actor's behavior to personal dispositions of the actor even more strongly than passive observers will. In a series of hypothetical emotional events, subjects were depicted either as actors ("You like Ted"), active observers ("Ted likes you"), or passive observers ("Ted likes Paul"). They then rated the degree to which the actor, active observer, or some "other reason" had caused the given event. Although the actor–observer effect was obtained overall, an interaction between subject role and positivity of verb indicated that it occurred much more strongly in negative-verb events than in positive-verb events. That is, subjects—as either actors or active observers—tended to deny their responsibility for negative events but did not claim praise for positive events. Implications for the effects of egotism on attribution are discussed.

The intriguingly simple hypothesis of Jones and Nisbett (1971) that "there is a pervasive tendency for actors to attribute their actions to situational requirements, whereas observers tend to attribute the same actions to personal dispositions" (p. 2) has sparked a profusion of research in recent years. In any interpersonal situation that involves both actors and observers, it is necessary to distinguish between two types of observers, active and passive. Miller and Norman (1975) stated that

> the term *active observer* refers to any participant in a social interaction situation who, in addition to observing the behavior of the other participants, influences the behavior of the other participants and is himself behaviorally influenced by the other participants. Active observers, therefore, function as both actors and observers. The term *passive observer* . . . refers to an individual who neither influences nor is influenced by the actor he is observing. (p. 503)

Clearly, this distinction is critical, since the active observer's attributions are more likely to be swayed than the passive observer's by the hedonic relevance and personalism of the actor's behavior (Chaikin & Cooper, 1973; Hill & Jones, 1973; Jones & Davis, 1965).

Among all of the interpersonal studies to be found in the actor–observer literature, few have investigated this distinction explicitly (Miller & Norman, 1975; Wolfson & Salancik, 1977; Weber, 1975; Cunningham & Antill, Note 1). Miller and Norman found, as predicted, that active observers attributed more behavioral responsibility to the actor and less responsibility to the interaction setting—a Prisoner's Dilemma game—than passive observers did. Supportive of Miller and Norman's results, divergent attributions were produced

Thanks are due John H. Harvey and two anonymous reviewers for their perceptive comments on a previous draft of this article.

Requests for reprints should be sent to John Cunningham, School of Behavioural Sciences, Macquarie University, North Ryde, New South Wales 2113, Australia.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3707-1146$00.75

by actors and active observers in two other bargaining studies (Snyder, Stephan, & Rosenfield, 1976; Stephan, Rosenfield, & Stephan, 1976). Yet Wolfson and Salancik's (1977) "involved (i.e., active) observers," who watched an actor operate a model race car set while anticipating competing against either the actor or another subject, made more situational (external) and fewer dispositional (internal) attributions for the actor's failing performance than uninvolved (passive) observers. To the extent that both involved and noncompetitive observers' attributions paralleled those of actors, these observers seemed to have adopted an empathic set (Regan & Totten, 1975; Stephan, 1975) rather than the motivationally defensive posture of the bargaining game players. The emergence of either empathic or "egotistic" (Snyder et al., 1976) attribution thus seems to be a complex function of the patterning of interaction (trial-by-trial bargaining vs. observer role anticipating actor role), outcome (success–failure vs. no explicit result), and sex of subjects. Unfortunately, the unstructured-conversation paradigm of Cunningham and Antill (Note 1) and Weber (1975)—the only apparent active–passive observer studies that did not use a competitive or achievement setting—yielded no differences between the attributions of active and passive observers.

The present study was conducted in the hope that a simplified paradigm would bring clarity to the issue of similarity or divergence in attributions made by actors, active observers, and passive observers. Jones and Nisbett (1971) have argued that both informational bias and information-processing bias contribute to differences in attributions made by actors and observers. The studies by Miller and Norman (1975) and by Wolfson and Salancik (1977) differed, among other ways, in the amount of information available to the active observer; in the latter study, the actor performed in full view of the active observer, but in the former, he was never seen. Regan and Totten (1975) showed that through induction of an empathic set, information processing alone can yield actor–observer differences. In a questionnaire format, the present study sought to minimize informational complexity but permit information-processing differences by locating the actor as the subject of a sentence ("You like Paul") and the active observer as the object of a sentence ("Ted likes you").

Using success and failure at hypothetical achievement events, Ruble (1973) manipulated actor–observer perspective by phrasing sentence subjects in the third person for the passive-observer role (e.g., "John Doe didn't work well with others on a project"), as opposed to the second person for the actor role (e.g., "You worked well with others on a project"). Incorporating these actor and passive-observer roles, this procedure was adapted in the present study to include the active-observer role as well by phrasing certain sentence objects in the second person (e.g., "Ted likes you"). Emotional verbs were used particularly to determine whether the predominance of causal attributions to sentence objects for subjective verbs in passive-observer events (Abelson & Kanouse, 1966; McArthur, 1972) would shift to sentence subjects in active-observer events, as the results of Miller and Norman (1975) would predict.

A second goal of the study examined the effect of extremity of event on ascriptions of causality. More extreme or severe consequences lead observers to assign greater responsibility to actors (Cunningham & Kelley, 1975; Walster, 1966), whereas actors attribute less responsibility and freedom to themselves (Harvey, Harris, & Barnes, 1975). It is predicted that actors and active observers will place the locus of causality for extreme emotional reactions more in the other person than in themselves, whereas passive observers will attribute greater causality to sentence subjects as emotions become more extreme.

Method

Subjects

Subjects were 24 male undergraduates at the University of California, Los Angeles, who participated in partial fulfillment of an introductory psychology course requirement.

Materials

A series of one-sentence interpersonal events depicting subjective emotional states was constructed (e.g.,

"Ted likes Paul somewhat"). Sentence subjects and objects were either common male first names or the pronoun *you*. Actor–observer perspective was manipulated by four combinations of subject–object pairings, for example, "Ted . . . Paul," "Paul . . . Ted," "You . . . Paul" (self as actor), and "Paul . . . you" (self as active observer). Four positive–negative verb pairs were used (like–dislike, admire–resent, love–hate, and trust–distrust). All verbs were phrased in the present tense. The extremity or intensity of the event was varied by the inclusion of a moderate ("somewhat") or extreme ("deeply") modifier. The complete crossing of all possible combinations of these four variables—actor–observer role, verb, positivity, and extremity—yielded a total of 64 events.

Each event was followed by three one-sentence causes or explanations for the given event. The first cause referred to qualities of the sentence subject, and the second referred to qualities of the sentence object. The third attribution ("some other reason") was left optional for experimental subjects to rate. Subjects were told to "rate each of these explanations on how likely it is that they caused the initial situation," and to make their rating of each cause independent of their ratings of the other causes. A typical item follows:

Ted likes Paul somewhat.
How likely is this because:

A. Ted is the kind of person that likes people.

1 2 3 4 5 6 7 8 9

B. Paul is the kind of person that people like.

1 2 3 4 5 6 7 8 9

C. Some other reason

1 2 3 4 5 6 7 8 9

The 9-point Likert scales were anchored at each end by "not likely" or "definitely likely" and at the midpoint by "moderately likely."

All subjects made ratings for all 64 events. However, all events containing the pronoun *you* were grouped in one booklet of 32 events, and all events with first names only were grouped in another booklet of 32 events. Within each booklet, the sequence of events was randomized, but each subject received the same sequence of events. The order in which subjects received one booklet or the other first was the only between-subjects variable.

Results

As preliminary analyses indicated that no significant differences appeared between pairs of equivalent passive-observer statements (e.g., "Ted likes Paul" and "Paul likes Ted"), these data were pooled in subsequent analyses so that the subject-role variable had three levels: actor, active observer, and passive observer.[1] All ratings were subjected to multivariate analysis of variance (Dixon, 197 Harris, 1975) with order of presentation booklet as the only between-subjects variab and with subject role, positivity, extremity, a verbs as repeated measures. The depende variables were the ratings of sentence acto active observers, and "other reasons" causes of the given event.

Order of Presentation Effects

The multivariate effects involving order presentation of booklets are sporadic a difficult to interpret. Also, no significant mai effect or interaction involving order emerge when the three dependent variables we analyzed separately. But when all three d pendent variables were taken together—i the multivariate analysis—subjects who r ceived the booklet containing passive-observe events first rated causality for negative-ver events lower than positive events and lowe than subjects who received that bookle second, Order × Positivity interaction, $F($ 196) = 32.7, $p < .001$. Those who receive the booklet containing passive-observer event first gave lower ratings of causality for event for which self "loves" or "hates" someon else than all of the other combinations, Orde × Verb × Subject Role multivariate inte action, $F(27, 573.06) = 3.19$, $p < .001$.[2] Thos who received the self-relevant booklet fir rated moderate negative events involvi self as actor or active observer highes whereas those receiving the other bookl first rated extreme negative events involvi self as actor or active observer highest, Orde × Positivity × Extremity × Subject Ro multivariate interaction, $F(9, 477.16) = 3.4$ $p < .001$. However, none of these order effect changed the conclusions to be drawn fro the other results, nor did they correspond

[1] Pooling of the equivalent passive-observer sta ments resulted in twice as many observations in passive-observer condition as in either the actor active-observer conditions. Unequal cell sizes w accommodated by the BMD computer program us in the analyses (Dixon, 1973).

[2] Fractional degrees of freedom appear because the algorithm employed by the program to deal wi unequal cell sizes (Dixon, 1973).

any apparent way to effects obtained when the order of actor and observer roles was manipulated (Sherrod & Farber, 1975, in failure situations).

Actor–Observer Hypothesis and Hedonic Relevance

A significant multivariate main effect for subject role lends support to the actor–observer hypothesis, as attributions to the self either as actor or as active observer were lowest of all means, $F(9, 477.16) = 67.9$, $p < .001$. When this analysis was performed on ratings of actor and active observer only on events that involved the self (i.e., actor role and active-observer role), the subject-role main effect emerged even more strongly, $F(6, 392) = 100.6$, $p < .001$. This confirms Miller and Norman's (1975) distinction between active and passive observers to the extent that subjects attributed less causality to themselves as active observers than when others were portrayed as active observers.

However, the Jones–Nisbett (1971) effect was ancillary to an interaction between subject role and positivity that reflected the hedonic relevance of the event. That is, subjects were much more reluctant to attribute causality to the self for negative verbs (e.g., hate and distrust) than for positive verbs, multivariate $F(9, 477.16) = 26.7$, $p < .001$. Indeed, an orthogonal comparison pitting negative verb-event ratings of self as actor and of self as active observer against the other 16 means in this interaction was highly significant, $F(1, 23) = 64.0$, $p < .001$. Thus, contrary to Miller and Norman's findings, the active observer did not attribute greater causality to the actor than a passive observer did. Rather, the active observer attributed—in negative events predominantly—less causality to self, as active observer, than a passive observer did.

Following the significant multivariate interaction between subject role and positivity, univariate interactions appeared for causal attributions to active and active observer separately, $F(3, 42) = 5.30$ and 8.96, ps $< .005$ and .001, respectively. As can be seen in Table 1, subjects tended to deny responsibility for negative events when they were cast either as actors or as active observers. Orthogonal comparisons probing these interactions showed that negative verb ratings of self as actor were much lower than all other actor ratings, $F(1, 23) = 29.7$, $p < .001$, and that negative verb ratings of self as active observer were much lower than all other active-observer ratings, $F(1, 23) = 46.8$, $p < .001$. On the other hand, positive verb ratings of self as actor were only somewhat lower than all other actor ratings, $F(1, 23) = 4.98$, $p < .04$, and positive verb ratings of self as active observer did not differ from all other active-observer ratings ($F < 1$).

Table 1
Mean Causal Attributions to Actor and Active Observer As a Function of Subject Role and Verb Positivity

Causal attribution to	Subject role: Actor	Active observer	Passive observer
Actor			
Positive verb	4.67	5.33	5.29
Negative verb	3.53	5.14	5.17
Active observer			
Positive verb	5.48	5.81	6.12
Negative verb	4.97	3.81	5.33

Note. A sample sentence in the actor condition is "You like Ted"; in the active observer condition is "Ted likes you"; and in the passive observer condition is "Ted likes Paul."

Verb Effects and Extremity Effects

The interaction between actor–observer perspective and positivity was particularly pronounced on events involving the verb *hate* and nearly as pronounced on the verb *dislike*, multivariate interaction, $F(27, 573.06) = 2.78$, $p < .001$. As a consequence, negative verbs, such as dislike and hate, received the lowest causality ratings of all verbs, and positive verbs, especially trust, the highest, multivariate interaction, $F(9, 477.16) = 8.95$, $p < .001$. Positive verbs had higher ratings than negative verbs on actor and active-observer ratings, univariate $F(1, 14) = 23.2$ and 34.5, respectively, both ps $< .001$, but lower ratings on other reasons, $F(1, 14) = 10.3$, $p < .01$. As predicted, more causality was attributed to actors for extremely negative events than for

moderately positive events, as seen in the significant multivariate interaction between extremity and positivity, $F(3, 196) = 4.07$, $p < .01$. Overall, extreme events were rated higher on causality ratings than moderate events, $F(3, 196) = 10.8$, $p < .001$; this was particularly true of other reasons ratings.

Discussion

The main conclusion to be drawn from this study is that the actor–observer hypothesis of Jones and Nisbett (1971) may hold true particularly for blameworthy or socially undesirable events, that is, events involving hatred, distrust, dislike, or resentment. Whenever portrayed as a participant in such an event, as either an actor or an active observer, the subject deflected blame from the self, directing it instead to the other person or to some unspecified other reason. This did not occur to the same extent when the subject was depicted as either an actor or active observer in a socially desirable event, that is, love, trust, liking, or admiration.

Snyder et al. (1976) defined egotism as "the tendency to make attributions that put oneself in the best possible light" (p. 435). Accordingly, subjects in the present experiment displayed limited egotism. Although they strongly denied their role in undesirable events, they did not claim excessive credit for desirable events. Thus, persons may be more likely to avoid blame for blameworthy acts than they are to demand praise for praiseworthy acts. This finding is explicable in terms of the results of a study by de Charms, Carpenter, and Kuperman (1965), who found that more freedom is attributed to a person who acts in order to please a liked other than to one who acts to avoid the threat of punishment. If persons are held responsible for an act to the extent that they are perceived as having the freedom to do it or not (Harvey, 1976), then, presumably, they will be held more responsible for acts of self-aggrandizement (claiming praise) than for acts of self-protection (blame avoidance). Others may regard excessive self-aggrandizement as boastful and, therefore, undesirable. Hence, a balance between demands for a favorable public image and for a robust level of self-esteem may be struck if individuals avoid blame but do not always claim praise for their actions.

The muting of self-aggrandizement in this study seems to contradict the findings of Snyder et al. (1976), who found that egotism was as potent a factor in a bargaining game after winning as after losing. That is, their subjects were as likely to flatter themselves on success as they were to deny personal responsibility for failure. However, the competitive nature of the bargaining game may have led the subjects to adopt an assertive "heads I win, tails you lose" attitude toward the situation. Similarly, when competing only against themselves, actors are more likely to take credit for their successes than their failures (Harvey, Arkin, Gleason, & Johnston, 1974; Johnson, Feigenbaum, & Weiby, 1964). The context of the present study, which involved neither competition nor achievement, might have encouraged a more even-handed set among subjects. Unfortunately, attributional data from the bargaining study of Miller and Norman (1975) do not address this particular question.

By supporting the original proposition of Jones and Nisbett (1971) that actors attribute less causality to themselves than passive observers do, the present findings contradict the results of Miller and Norman (1975). Their bargaining subjects ascribed more behavioral responsibility to themselves and perceived more disposition in their behavior than did observers. However, in the physical absence of the bargaining opponent, the opponent's behavior probably did not "engulf the field" (Heider, 1958), a presumed prerequisite for the actor–observer effect. Because the bargaining session consisted of a number of trials involving identical choices in the same situation, the actor's own behavior would seem to be the most vivid feature of an otherwise informationally impoverished setting. Such vividness would direct attention to the self, thus fostering attributions of responsibility to the self (Duval & Wicklund, 1972). Moreover, when actor–active observer role was treated as a within-subjects factor, Miller and Norman found no significant differences on ratings of responsibility or personality, in contrast to ratings of causality in the present study.

It may be argued that the artificiality of the present questionnaire format lacks impact and precludes generalization to behavioral situations (Ruble, 1973). However, actor–observer differences have been found in previous questionnaire studies (e.g., Nisbett, Caputo, Legant, & Maracek, 1973). To the extent that this technique has limited realism, it may simply underestimate the strength of actual behavioral effects. This technique also has several advantages. As each subject attributes causality for many events in a questionnaire, a wide range of events can be sampled. By contrast, the findings of Miller and Norman (1975), Snyder et al. (1976), and Stephan et al. (1976) may not generalize beyond the bargaining context that these authors used. Such situations undoubtedly elicit rivalry and face-saving strategies that may apply only to a circumscribed subset of natural settings. Within a limited period of time, it would be difficult indeed to stage, in factorial combinations, behavioral equivalents of each of many events in both passive-observer and self-relevant modes. As well, events need not be informationally or behaviorally rich before people seek to ascribe causality for them. Often, attributions are based on hearsay or on minimal information, as in the present study. Further research would benefit from a systematic sampling of a variety of interpersonal events (as by Carson's, 1969, circumplex) to identify the situations in which active observers display varying degrees of empathic or egotistic attribution. The obvious limitations of a questionnaire study would be transcended by creating behavioral situations that permit more direct extrapolation to face-to-face encounters.

Reference Note

1. Cunningham, J. D., & Antill, J. K. *Actors, active observers, and passive observers in a getting-acquainted situation.* Unpublished data, Macquarie University, New South Wales, Australia, 1976.

References

Abelson, R. P., & Kanouse, D. E. Subjective acceptance of verbal generalizations. In S. Feldman (Ed.), *Cognitive consistency*. New York: Academic Press, 1966.

Carson, R. C. *Interaction concepts of personality*. Chicago: Aldine, 1969.

Chaikin, A. L., & Cooper, J. Evaluation as a function of correspondence and hedonic relevance. *Journal of Experimental Social Psychology*, 1973, *9*, 257–264.

Cunningham, J. D., & Kelley, H. H. Magnitude of effect and causal attributions. *Journal of Personality*, 1975, *43*, 74–93.

de Charms, R., Carpenter, V., & Kuperman, A. The "origin-pawn" variable in person perception. *Sociometry*, 1965, *28*, 241–258.

Dixon, W. J. (Ed.) *BMD: Biomedical computing programs* (3rd ed.) Berkeley: University of California Press, 1973.

Duval, S., & Wicklund, R. A. *A theory of objective self-awareness*. New York: Academic Press, 1972.

Harris, R. J. *A primer in multivariate statistics*. New York: Academic Press, 1975.

Harvey, J. H. Attribution of freedom. In J. H. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research* (Vol. 1). Hillsdale, N.J.: Erlbaum, 1976.

Harvey, J. H., Arkin, R. M., Gleason, J. M., & Johnston, S. Effect of expected and observed outcome of an action on the differential causal attributions of actor and observer. *Journal of Personality*, 1974, *42*, 62–77.

Harvey, J. H., Harris, B., & Barnes, R. D. Actor-observer differences in the perceptions of responsibility and freedom. *Journal of Personality and Social Psychology*, 1975, *32*, 22–28.

Heider, F. *The psychology of interpersonal relations*. New York: Wiley, 1958.

Hill, C. T., & Jones, J. M. Effects of prejudice and perspective on attribution of intentionality. *Proceedings of the 81st Annual Convention of the American Psychological Association*, 1973, *8*, 347–348. (Summary)

Johnson, T. J., Feigenbaum, R., & Weiby, M. Some determinants and consequences of the teacher's perception of causation. *Journal of Educational Psychology*, 1964, *55*, 237–246.

Jones, E. E., & Davis, K. E. From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2). New York: Academic Press, 1965.

Jones, E. E., & Nisbett, R. E. *The actor and the observer: Divergent perceptions of the causes of behavior*. Morristown, N.J.: General Learning Press, 1971.

McArthur, L. A. The how and what of why: Some determinants and consequences of causal attributions. *Journal of Personality and Social Psychology*, 1972, *22*, 171–193.

Miller, D. T., & Norman, S. A. Actor-observer differences in perceptions of effective control. *Journal of Personality and Social Psychology*, 1975, *31*, 503–515.

Nisbett, R., Caputo, C., Legant, P., & Maracek, J. Behavior as seen by the actor and as seen by the observer. *Journal of Personality and Social Psychology*, 1973, *27*, 154–164.

Regan, D. T., & Totten, J. Empathy and attribution: Turning observers into actors. *Journal of Personality and Social Psychology*, 1975, *32*, 850–856.

Ruble, T. L. Effects of actor and observer roles on attributions of causality in situations of success and failure. *Journal of Social Psychology*, 1973, *90*, 41–44.

Sherrod, D. R., & Farber, J. The effect of previous actor/observer role experience on attribution of responsibility for failure. *Journal of Personality*, 1975, *43*, 231–247.

Snyder, M., Stephan, W. G., & Rosenfield, D. Egotism and attribution. *Journal of Personality and Social Psychology*, 1976, *33*, 435–441.

Stephan, W. G. Actor vs. observer: Attributions to behavior with positive or negative outcomes and empathy for the other role. *Journal of Experimental Social Psychology*, 1975, *11*, 205–214.

Stephan, W. G., Rosenfield, D., & Stephan, C. Egotism in males and females. *Journal of Personality and Social Psychology*, 1976, *34*, 1161–1167.

Walster, E. Assignment of responsibility for an accident. *Journal of Personality and Social Psychology*, 1966, *3*, 73–79.

Weber, G. S. Interpersonal causal attributions of active and passive observers (Doctoral dissertation, University of Arizona, 1974). *Dissertation Abstracts International*, 1975, *35*, 4263B. (University Microfilms No. 75-4132)

Wolfson, M. R., & Salancik, G. R. Observer orientation and actor-observer differences in attributions for failure. *Journal of Experimental Social Psychology*, 1977, *13*, 441–451.

Received May 31, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 7, 1153–1159

Self-Esteem and Education: Sex and Cohort Comparisons Among High School Seniors

Patrick M. O'Malley and Jerald G. Bachman
University of Michigan (Ann Arbor)

The present article reports on the self-esteem of 3,183 male and female seniors in a nationwide sample of the high school class of 1977 and draws comparisons with 1,715 males from the class of 1969. It thus provides a partial replication and extension of an earlier study, which showed that educational accomplishments undergo a reduction in centrality—become less important—for self-esteem during the late teens and early twenties. One major finding is that the self-esteem of seniors was correlated with educationally relevant measures almost equally for males in the classes of 1969 and 1977. This finding appears to rule out a secular trend interpretation of the earlier study's results, thereby providing further support for a developmental interpretation. A second finding is that self-esteem correlated with educationally relevant measures about equally for males and females in the class of 1977, suggesting that the impact of educational factors is basically similar for the two sexes. A third finding is that males and females were very similar in levels of self-esteem.

A recent analysis by Bachman and O'Malley (1977), based on a sample of about 1,600 young men from the high school class of 1969, showed that self-esteem is positively correlated with educational success. The study also showed that educational accomplishments seem to have greater importance or centrality for self-esteem during the high school years than during the 5 years beyond high school. Although the data for that study were nationally representative and covered an 8-year longitudinal span, there remained at least two important limitations on the generalizability of the findings. First, and most obviously, the study was limited to young men, thus leaving it unclear whether the findings are applicable to young women. Second, the study spanned the interval from 1966 to 1974, a period in which there were important social changes, including changes in public views about the value of education. Therefore, the data left some uncertainty about whether the changing link between educational success and self-esteem reflected a genuine developmental process in late adolescence or a secular trend that affected the society as a whole.

The present article, reporting new data from a nationwide cross section of male and female seniors from the high school class of 1977, provides a partial replication and extension of the study of the class of 1969. The new data permit us to overcome both of the limitations noted above. Unlike the previous study, the findings reported here are not longitudinal; therefore, we will not be concerned with relating self-esteem to later (post–high school) educational and occupational attainments. Instead, we will see how self-esteem among seniors relates to recent educational success (high school grades), self-ratings of academic ability, college plans, and parental education. We will examine these relationships in parallel fashion for three distinct groups: senior males in the class of 1969, senior males in the class of 1977, and senior females in the class of 1977.

Analyses for this article were supported by National Institute of Education Grant NIE-G-78-0036; the data collections were supported by Office of Education Grant OE-5-85-054 and National Institute on Drug Abuse Grant DA01411. The authors thank F. Andrews, M. Gold, R. Herzog, R. Kulka, and J. Miller for their helpful comments.

Requests for reprints should be sent to Patrick M. O'Malley, Survey Research Center, Box 1248, University of Michigan, Ann Arbor, Michigan 48106.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3707-1153$00.75

One of the basis conclusions from the previous study was that

> those things having to do with a self-concept of educational success—things such as academic skills, past classroom performance, future aspirations, and the like—undergo some reduction in salience or "centrality" for the overall self-esteem of young men as they move through the final years of high school and go on to other experiences. (Bachman & O'Malley, 1977, p. 377)

The study recognized the possibility of different interpretations.

> The first interpretation, and the one which we have emphasized, is that this shift in centrality is a fairly typical part of the developmental sequence followed by young people in this society. During the late high school years and the period which follows, the young person in the process of becoming an adult increasingly anticipates and experiences situations in which self-evaluation depends on factors quite different from success in school; and this means that academic things become less dominant in shaping self-esteem. An alternative interpretation of our findings is that they reflect a particular secular trend or cultural change during the late sixties and early seventies—a general decline in the importance or value that society places upon education and educational success. Trust in government declined dramatically during this period, and some would argue that faith in education as the pathway to success has also suffered a setback. (Bachman & O'Malley, 1977, pp. 378–379)

The present analysis, by comparing data for males in the classes of 1969 and 1977, provides a means of testing whether one of the alternative interpretations outlined above is more valid than the other. First, it must be understood that the earlier longitudinal study revealed a gradual and continuing change from 1966 (start of 10th grade) through 1974 (5 years postgraduation); across a sequence of five data collections, correlations between self-esteem and educational factors grew progressively smaller (Bachman & O'Malley, 1977; Bachman, O'Malley, & Johnston, 1978). Now suppose that this gradual change reflected a substantial secular trend or cultural change affecting the relationship between self-esteem and educational success; then, unless the trend dramatically reversed during the past several years, correlations between educational factors and self-esteem should be substantially lower for seniors in 1977 compared with seniors in 1969. But suppose, on the other hand, that the earlier findings were due solely to developmental changes occurring during late adolescence; then, unless such developmental patterns were somehow different a decade later, the correlations for seniors in 1977 should be similar to those found for seniors in 1969.

A different point of interest in the present analysis is how males and females compare in the extent to which self-esteem is determined by (correlated with) various educationally relevant dimensions. Douvan and Gold (1966), in reviewing the literature on self-esteem more than a decade ago, concluded that the self-esteem of boys and girls depended to some extent on different components. Also, Rosenberg and Simmons (1975) found evidence that indicated that adolescent boys emphasized the importance of competence and achievement more than did girls. If this is true, we may expect that to the extent that boys' needs for competence and achievement depend on educational experiences, correlations of educationally relevant variables with self-esteem should be higher for boys than for girls.

A final point of interest is whether boys have higher levels of self-esteem than girls do. There have been reports that adolescent females' self-images are less favorable than males' (Rosenberg & Simmons, 1975), although others have not found differences (Drummond, McIntire, & Ryan, 1977; Rosenberg, 1965). There are good reasons why females' self-esteem should be lower than that of males, feminists would argue. The females' less advantageous position in society could easily result in lower self-esteem. Simmons and Rosenberg (1975) in their study of adolescent girls found that "at least in 1968, girls appeared to have a more unfavorable self-picture than did boys" (p. 233). As implied in the quote, an interesting question is whether that was still true in 1977. Perhaps the upheaval in women's rights and roles in the period from 1968 to 1977 resulted in an erasure of those differences.

In fact, a comparison has already been made between the findings from the 1968 study (Simmons & Rosenberg, 1975) and findings from a later, similar study conducted from 1974 to 1975 (Bush, Simmons, Hutchinson, & Blyth, 1977–1978) "to see if the new feminist ideology and movement could have had an effect on girls' self-esteem" (p. 472).

Table 1
Self-Esteem Items and Item-Index Correlations

Item	1977 high school seniors: Males n	M	SD	r[a]	Females n	M	SD	r[a]	t test between means
I take a positive attitude toward myself.[b]	1,568	4.13	.93	.54	1,736	3.98	.99	.56	**
I feel I am a person of worth, on an equal plane with others.[b]	1,555	4.27	.85	.49	1,727	4.21	.90	.55	
I am able to do things as well as most other people.[b]	1,560	4.41	.75	.39	1,733	4.30	.81	.50	**
On the whole, I am satisfied with myself.[b]	1,539	4.06	1.05	.48	1,718	4.06	1.03	.57	
I feel I do not have much to be proud of.[c]	1,519	4.03	1.14	.54	1,707	4.05	1.15	.60	
Sometimes I think that I am no good at all.[c]	1,506	3.63	1.32	.51	1,701	3.47	1.39	.50	**
I feel that I can't do anything right.[c]	1,498	4.07	1.11	.41	1,695	4.11	1.08	.53	
I feel that my life is not very useful.[c]	1,501	4.18	1.05	.60	1,687	4.18	1.10	.62	*
Self-esteem index	1,494	4.10	.66	.79[d]	1,689	4.04	.72	.83[d]	

[a] Item-index correlation, corrected for part–whole.
[b] Response of "agree" coded 5 (high self-esteem).
[c] Response of "disagree" coded 5 (high self-esteem).
[d] Coefficient alpha.
* $p < .05$. ** $p < .01$.

Like the 1968 study, this more recent study found that girls' self-esteem in the sixth and seventh grades averaged slightly lower than boys'. The present study extends the time period to 1977 and looks at seniors, rather than sixth and seventh graders.

In sum, this article employs data from two studies to address three theoretically important questions about self-esteem. First, are the patterns of correlation between self-esteem and educational factors essentially the same for males in the high school classes of 1969 and 1977? Second, are these patterns of correlation similar for males and females (in 1977)? Third, are there differences in levels of self-esteem between males and females (in 1977)?

Method

Sample

One source of data is the Youth in Transition sample of young men from the high school class of 1969, described in detail by Bachman et al. (1978). Briefly, this is a nationally representative sample of 2,277 sophomores in 87 schools in the 48 contiguous states. Data collections were made in 1966 when they were sophomores and again in 1968, 1969 (when they were seniors), 1970, and 1974. In the 1969 data collection, the one most relevant here, 1,799 participated (79% of the original target sample).

The second source of data is the Monitoring the Future project being conducted by the University of Michigan's Institute for Social Research. The study is described in detail by Bachman and Johnston (Note 1); only a brief overview is given here. The project consists of a series of annual questionnaire surveys of male and female seniors in high school. In addition, annual follow-up surveys are mailed to a subset of each sample for 6 years following their graduation. The initial base-year data collection is conducted in about 125 schools selected to provide an accurate cross section of high school seniors throughout the United States. The 1977 base-year survey of over 18,000 seniors provided the data to be reported here.[1] The measure of primary interest, self-esteem, was included in only one of the five forms used in the study, so analyses in this report were based on an essentially random 20% of the total. The number of respondents with complete data on the self-esteem index was 3,183.

[1] The study began with the class of 1975; however, 1977 was the first year that included all of the eight self-esteem items presented in this article.

Measures

Self-esteem measures in the two studies were similar but not identical. The Monitoring the Future measure was an eight-item index similar to that used by Rosenberg (1965). Respondents were asked to indicate on a 5-point scale the extent of their agreement or disagreement with the items. The response categories—disagree, mostly disagree, neither, mostly agree, agree—were coded from 1 to 5, with higher values assigned to responses indicating higher self-esteem. The first four items were positively worded, and the second four were negatively worded. Table 1 shows the eight items, with their means, standard deviations, and item-index correlations, separately for males and females. Coefficient alphas were .79 for males and .83 for females.

The measure of self-esteem used in the Youth in Transition study differed slightly from the Monitoring the Future measure. A different 5-point scale was used; respondents were asked to indicate how often each item was true for them, with responses ranging from almost always true to never true. There were 10 items, 7 of which were virtually identical to those used in the 1977 Monitoring the Future study (Items 1–3 and 5–8 in Table 1). (See Bachman & O'Malley, 1977, for the Youth in Transition item wordings, means, standard deviations, and item-index correlations.) Coefficient alpha for males in the class of 1969 was .79, the same as the value for males in the class of 1977.

Because the two studies employed different response scales, and some different items, it is difficult to compare means and standard deviations directly. However, because the two measures are very similar in their basic content as well as index characteristics (coefficient alphas and item-index correlations), their correlations with other measures should be comparable. (The comparability can be heightened by using subscales based on the seven common items. The present article includes correlations based on these subscales. To maintain comparability with other reports using these data, however, we focused primarily on the complete 10-item and 8-item measures.)

Unless otherwise stated, the remaining measures listed below are the same in the two studies.

Father's education and mother's education were measured on a 6-point scale ranging from grade school or less to graduate or professional school. Parental education is a mean of father's and mother's education. Grades were self-reported on a 9-point scale. In the Youth in Transition sample, the grades for junior and senior years were averaged together; in the Monitoring the Future sample, respondents were asked for "your average grades so far in high school."

College plans is a dichotomy indicating whether the respondent planned to enter college or not. Self-concept of school ability is a mean of two items that asked respondents to compare themselves, on a 7-point scale, with others of the same age on school ability and intelligence. For the class of 1969, the measure was obtained as of junior year; for the class of 1977, the measure was made as of senior year.

Statistical Significance

In reporting statistical significance levels, we made no adjustment for the clustered sampling design used for both samples. One reason for not adjusting for design effect was that we did not have good estimates of the design effect for self-esteem and its correlations with other variables. We are confident, however, that these design effects are small because there is little clustering of self-esteem by school. Since the nominal probability levels are nevertheless likely to be slightly inflated, we report statistical significance at both the .01 and the .05 levels.

Results and Discussion

Class of 1977: Male Versus Female Levels of Self-Esteem

Table 1 shows that responses to the self-esteem items for males and females in the class of 1977 indicated a generally high level of self-acceptance. The average response was just above "mostly agree" for the positive items and about "mostly disagree" for the negative items. On six of the eight items, males were higher in self-esteem than females (three of these significantly so); for the total index, males were significantly higher ($p < .05$). Females were more internally consistent, however, as indicated by the higher item-index correlations for all but one item, and the higher coefficient alpha.

The fact that self-esteem was slightly lower for females than for males is consistent with Simmons and Rosenberg (1975) and Bush et al. (1977–1978). Apparently, the recent activities aimed at upgrading women's position in society have not resulted in a total elimination of the differences in self-esteem for high school seniors. But perhaps more important is the fact that the difference in Table 1 is small indeed—less than .09 of a standard deviation. This very small difference is statistically significant given a large number of cases, but studies that use smaller samples would be unlikely to show statistically significant differences. This may explain why Maccoby and Jacklin (1974) reported that statistically significant sex differences were seldom found. They commented that "the similarity of the sexes in self-esteem is remarkably uniform across age levels through college age" (p. 153). But their review was based on a number of studies, most of which involved small samples.

Table 2
Correlations of Self-Esteem With Educational Measures for Seniors in 1969 and in 1977

Measure	Youth in Transition (Class of 1969)	Monitoring the Future (Class of 1977)	
	Males (n = 1,715)[a]	Males (n = 1,494)[a]	Females (n = 1,689)[a]
Father's education	.09 (.10)	.08 (.09)	.11 (.11)
Mother's education	.06 (.07)	.07 (.07)	.08 (.08)
Parental education	.08 (.10)	.08 (.09)	.11 (.11)
Grades	.21 (.21)	.25 (.26)	.24 (.26)
College plans	.18 (.19)	.18 (.20)	.14 (.16)
Self-concept of school ability	.27 (.26)	.28 (.29)	.29 (.30)

Note. All correlations are significantly different from zero at the .01 level, except that between self-esteem and mother's education, for the 1977 senior males, which is significant at the .05 level. The correlations in parentheses are based on the subscale that consists of the 7 items common to both studies (Items 1–3 and 5–8 in Table 1). In no case does a subscale correlation differ from the corresponding full scale correlation by more than .02. The 7-item subscale correlates .95 with the 10-item Youth in Transition measure, .98 with the 8-item Monitoring the Future measure for males, and .99 for females. Subscale coefficient alphas are .76, .76, and .80 for the three samples, respectively.
[a] The number of cases varies slightly for each correlation, with up to 8% missing data.

There is another large-scale study of a nationally representative sample of high school seniors that included a measure of self-esteem similar to those used in the present articles, namely, the "National Longitudinal Study of the High School Class of 1972." Using these data, Conger, Peng, and Dunteman (Note 2) reported that female seniors were statistically significantly lower in self-esteem than males by about .10 of a standard deviation.

A further note of consistency in the finding of slightly lower scores on self-esteem for females than for males comes from the Monitoring the Future data collections from seniors in 1975 and 1976. In those years, an abbreviated four-item self-esteem scale was used. In each year, the females averaged slightly lower than the males.

One other point should be made. Our discussion thus far has taken these self-reports of self-esteem at face value and dealt only with bivariate relationships. It is possible that a more substantial difference between sexes in self-esteem—or even a reversal of the difference reported here—is being masked or suppressed by other related variables that we have not controlled (Rosenberg, 1973). There are several likely candidates for such suppressor variables. Maccoby and Jacklin (1974) noted that boys score higher on lie or defensiveness scales. Thus, boys might defensively overrate themselves on the self-esteem measures. If so, girls may actually have higher self-esteem than boys. A related possible suppressor is social desirability, the tendency to present oneself in as favorable a light as possible. Bush et al. (1977–1978) included such a measure in their study and found that girls were higher than boys in social desirability. Thus, their findings that females were slightly lower in self-esteem could not be due to a social desirability effect. If anything, the differences would be greater if social desirability were controlled. But the correlations among all the variables in their study were low, and controlling actually made little difference in relationships; the zero-order and partial correlations were similar.

Finally, in the Monitoring the Future data, grades could be suppressing a sex effect on self-esteem. We demonstrate that grades and self-esteem are positively correlated; further, females report higher grades than males (data not shown here). Therefore, one would expect females to show higher self-esteem. The fact that they showed lower self-esteem indicates that self-esteem differences would be greater if grades were controlled. But again the differences between zero-order and partial relationships were small. The zero-order correlation of .04 between sex and self-esteem becomes a standardized regression coefficient

of only .08 when grades are controlled, and the substantive significance of a regression coefficient of .08 is dubious indeed.

In sum, the data on sex differences in self-esteem among adolescents showed some consistency across studies and across time. Females scored slightly lower than males, but so slight were the differences that relatively large samples were required to achieve statistical significance. Such differences—on the order of .10 of a standard deviation or less—are of dubious substantive significance. As Maccoby and Jacklin (1974) pointed out, it is the similarity, not the difference, that is remarkable. Although there remains the logical possibility of substantively important effects of sex on self-esteem, the various suggested controls—some of which serve to heighten self-esteem differences between the sexes and some of which serve to lessen differences—do not appear to be sufficiently strong to alter the conclusion that self-esteem scores are distributed similarly among males and females.

Classes of 1969 and 1977: Educational Variables and Self-Esteem

Now we turn to correlations between self-esteem and measures of parental education, grades, college plans, and self-concept of school ability. We compare these correlations, obtained separately for males and females, with those in the Youth in Transition sample of boys in the high school class of 1969.[2] Table 2 shows these correlations. The basic import of Table 2 is that the correlations of self-esteem with the various measures of background, achievement, and aspirations are all similar among the three groups. None of the pairwise comparisons between groups showed a statistically significant difference. In other words, (a) the correlations for male seniors were stable across the 8 years from 1969 to 1977 and (b) the correlations were similar for both males and females in the class of 1977.

This second finding is rather important. As noted earlier, Douvan and Gold (1966) concluded that the self-esteem of boys and girls depended to some extent on different components. The data in Table 2 suggest that the self-esteem of male and female seniors does not depend differentially on the components listed there.

The first correlational finding—that the self-esteem of male seniors correlates with educationally relevant measures about equally for the classes of 1969 and 1977—provides an extension and replication of part of the findings of the previous report. Although it does not deal with the longitudinal aspects of that report, by replicating the cross-sectional aspects it provides some added support for the conclusions reached there.

It appears that self-esteem in high school does depend to some extent on factors of background, achievement, and aspiration. All correlations in Table 2 are positive, and some are at least moderately high (i.e., grades and self-concept of school ability). (All of the correlations in Table 2 are significantly different from zero.)

As discussed above, the earlier longitudinal analysis found that "educational success and its correlates seemed to grow less central to the self-esteem of young men . . . as they moved through high school and beyond," and this was interpreted as "a fairly typical part of the developmental sequence followed by young people in this society" (Bachman & O'Malley, 1977, p. 378). But an alternative explanation was that there was an overall secular trend involving a lessening importance of education for self-esteem—a pattern that would be reflected in reduced correlations between education and self-esteem in later cohorts of students. If the data presented in Table 2 had shown lower correlations for the 1977 seniors (male) than for those from the class of 1969, the secular trend interpretation would have been supported. In fact, however, Table 2 shows a striking degree of similarity in the relationships for the two cohorts; certainly the correlations for the class of 1977 are not at all lower than those for the class of 1969.

Thus, by failing to confirm the secular trend interpretation, the present replication and extension provides additional, albeit indirect,

---

[2] Because the Monitoring the Future sample design eliminated high school dropouts, the present computations for the Youth in Transition sample also eliminate dropouts. The number of cases is therefore slightly different from the earlier report that included dropouts (Bachman & O'Malley, 1977).

support for a basic conclusion from the earlier study: Educational success becomes less central to self-esteem during late high school and the years that follow. Furthermore, given the high degree of similarity between male and female self-esteem data in the class of 1977, it is likely that this process holds true for both sexes.

Reference Notes

1. Bachman, J. G., & Johnston, L. *The Monitoring the Future project: Design and procedures.* Unpublished manuscript, 1978. (Available from the Institute for Social Research, University of Michigan, P.O. Box 1248, Ann Arbor, Mich. 48106).
2. Conger, A. J., Peng, S. S., & Dunteman, G. H. *National longitudinal study of the high school class of 1972: Group profiles on self-esteem, locus of control, and life goals.* Unpublished manuscript, 1977. (Available from Research Triangle Institute, Research Triangle Park, N.C. 27709).

References

Bachman, J. G., & O'Malley, P. M. Self-esteem in young men: A longitudinal analysis of the impact of educational and occupational attainment. *Journal of Personality and Social Psychology*, 1977, *35*, 365–380.

Bachman, J. G., O'Malley, P. M., & Johnston, J. *Youth in Transition VI: Adolescence to adulthood—change and stability in the lives of young men.* Ann Arbor, Mich.: Institute for Social Research, 1978.

Bush, D. E., Simmons, R. G., Hutchinson, B., & Blyth, D. A. Adolescent perception of sex-roles in 1968 and 1975. *Public Opinion Quarterly*, 1977–1978, *41*, 459–474.

Douvan, E., & Gold, M. Modal patterns in American adolescence. In L. W. Hoffman & M. L. Hoffman (Eds.), *Review of child development research.* New York: Russell Sage Foundation, 1966.

Drummond, R. J., McIntire, W. G., & Ryan, C. W. Stability and sex differences on the Coppersmith Self-Esteem Inventory for students in grades two to twelve. *Psychological Reports*, 1977, *40*, 943–946.

Maccoby, E. E., & Jacklin, C. N. *The psychology of sex differences.* Stanford, Calif.: Stanford University Press, 1974.

Rosenberg, M. *Society and the adolescent self-image.* Princeton, N.J.: Princeton University Press, 1965.

Rosenberg, M. The logical status of suppressor variables. *Public Opinion Quarterly*, 1973, *37*, 359–372.

Rosenberg, F., & Simmons, R. G. Sex differences in the self-concept in adolescence. *Sex Roles*, 1975, *1*, 147–159.

Simmons, R. G., & Rosenberg, F. Sex, sex roles, and self-image. *Journal of Youth and Adolescence*, 1975, *4*, 229–258.

Received July 7, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 7, 1160-1171

Deindividuation, Self-Awareness, and Disinhibition

Ed Diener
University of Illinois at Urbana-Champaign

This study was designed to discover whether lack of self-awareness, group unity, lack of conscious planning, and disinhibited behavior occur together in deindividuating settings, as predicted by Diener's theory of deindividuation. Another purpose of the study was to compare the characteristics of deindividuation in groups with the characteristics of lack of self-awareness induced in a nonsocial way. Three conditions were compared: deindividuated, non-self-aware, and self-aware. After the manipulations, participants could choose inhibited versus disinhibited tasks in a supposed "creativity" session, followed by a variety of deindividuation measures. When the dependent variables were factor analyzed, key components of deindividuation loaded together on a single common factor. The results also revealed that the deindividuation group surpassed the other two on the deindividuation factor ($p < .001$) and on most of the individual measures. For some of the variables, the deindividuation and non-self-aware groups differed significantly, suggesting that deindividuation may not be identical in every respect to lack of self-awareness induced in a nonsocial way.

Festinger, Pepitone, and Newcomb (1952) suggested that persons may not be perceived as individuals in some groups. They argued that these "deindividuated" people are freed from restraints and are likely to perform behaviors they would usually inhibit. Zimbardo (1970) further theorized that there are certain input variables that cause deindividuation, certain inferred internal changes such as minimal self-evaluation, and certain "output behaviors" that characterize the state. Spurred by Zimbardo's theory, a number of studies have been conducted that have explored deindividuation in the laboratory (e.g., Baron, 1971; Diener, Westford, Dineen, & Fraser, 1973; Dion, 1971; Maslach, 1974), in field settings (e.g., Diener, Fraser, Beaman, & Kelem, 1976), and cross-culturally (Watson, 1973). Past research, which has been reviewed in detail elsewhere (Diener, 1977; Dipboye, 1977), has focused almost exclusively on the relationship between specific input variables and the resultant behavior. Thus, the majority of the findings can be interpreted in terms of discriminative stimuli for disinhibited behavior without recourse to the concept of deindividuation.

Diener attempted to provide direct support for deindividuation by measuring the internal changes that Zimbardo's (1970) theory suggested should accompany the situational inputs (Diener, 1976; Diener, Dineen, Endresen, Beaman, & Fraser, 1975). He manipulated factors such as anonymity and responsibility and measured aggressive behavior and internal cognitive changes as well. He assessed, for example, memory and self-consciousness. The findings did not appear to offer strong support for the theory. The cognitive measures in most cases did not correlate with the overt disinhibited behavior, and no strong central factor was found when the measures were factor analyzed. However, when the paradigm employed in these studies is considered, it is not surprising that deindividuation did not occur. After subjects were exposed to the manipula-

My thanks are extended to Robert Flax, Thomas Srull, and Carol Diener for their help in conducting the study and preparing the manuscript. I also appreciate the insightful comments of Charles Carver, and the anonymous reviewers.

Requests for reprints should be sent to Ed Diener, Department of Psychology, University of Illinois, Champaign, Illinois 61820.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3707-1160$00.75

tions, they were observed while they "tested a pacifist" by being physically aggressive to him. In such a laboratory situation, subjects were undoubtedly quite self-conscious. In conclusion, past findings on the behavioral consequences of deindividuating conditions have offered some support for the theory. However, no study has found that subjective experiences such as lack of self-awareness or feelings of group unity follow deindividuating conditions and covary together.

Cognitive Factors Underlying Deindividuation

The theory of self-awareness (Wicklund, 1975, 1979) has much in common with the theory of deindividuation. Duval and Wicklund (1972) hypothesized that self-aware persons are more likely than non-self-aware persons to behave in accord with personal and social standards. Much empirical evidence has been accumulated that supports this formulation (e.g., Carver, 1974, 1975; Diener & Srull, 1979, Diener & Wallbom, 1976; Rule, Nesdale, & Dyck, 1975; Scheier, Fenigstein, & Buss, 1974). Given the pivotal position of attention to oneself as a separate entity in both the theories of self-awareness and deindividuation, several authors have suggested that a conceptual integration is in order (Diener, 1977; Duval & Wicklund, 1972), and Ickes, Layden, and Barnes (1978) have found empirical support for such an integration.

Recently Diener (1979) advanced a theory of deindividuation that borrows from the theory of self-awareness, from behavioral theories of self-regulation (Bandura, 1976; Kanfer, 1977), and from earlier models of deindividuation. In Diener's theory, monitoring of one's own behavior and awareness of oneself as an individual are necessary for self-regulation (cf. Carver, in press) and are prevented by certain factors present in some groups. Diener has lessened the emphasis on anonymity as a cause of deindividuation and stressed instead group cohesiveness and uniformity (Ziller, 1964), group activity, and an outward focus of attention. Each of these seems to be a likely cause of deindividuation because it can diminish or prevent self-awareness. According to this theory, the deindividuated person is blocked by environmental factors from becoming self-aware and is thus less likely to regulate his or her behavior in reference to personal and social standards and long-term planning considerations, and more likely to react to immediate cues, motivations, and emotions. Although antisocial behavior does not necessarily result from deindividuation, the person is more influenced by situational factors and less concerned with standards.

Purposes of the Study

The present study was designed to assess whether deindividuation is a useful construct and actually represents a unitary experiential and behavioral phenomenon. A major purpose of the study was to implement seemingly deindividuating conditions and measure both the behavior and the other changes that are defining characteristics of the concept. When in deindividuating circumstances, according to Diener's theory, persons should be less self-aware, plan ahead less, feel close to the group, and be more disinhibited than usual. Therefore, persons in deindividuating situations should have higher scores on the average on each of the central measures of deindividuation than persons who are not in such settings. In addition, these characteristics (low self-awareness, group unity, and disinhibition) should occur together within deindividuated persons and form a single factor when factor analyzed. In the present study, conditions thought to produce deindividuation were maximized, and multiple measures of disinhibited behaviors and postulated intervening cognitive variables were employed. A number of improvements over previous studies (Diener, 1976; Diener et al., 1975) were introduced that seemed likely to increase the likelihood of occurrence of deindividuation.

A second purpose of the present study was to compare the characteristics and effects of group-induced deindividuation with nonsocially induced non-self-awareness. Although Diener's (1979) theory maintains that deindividuating conditions block self-awareness, it seems likely that nonsocial conditions, for example, viewing an exciting movie, can also minimize self-awareness. However, group deindividuating conditions may be more certain to prevent

self-awareness because they not only direct attention away from the self, but also focus attention explicitly on the group. Thus, attention may be less likely to drift back to the self.

A last question addressed by the present study is, What are the characteristics of deindividuation beyond those mentioned earlier? Although group unity and lack of self-awareness are defining characteristics of deindividuation, many additional components have been suggested. For example, Festinger et al. (1952) suggested memory impairment regarding the acts of individual persons. Zimbardo (1970) suggested several additional characteristics such as amnesia, perceptual distortion, and emotionality.

In summary, the major purpose of the present study was to create a situation that appeared to be deindividuating and to assess whether the characteristics, both overt and covert, predicted by the theory result from this situation. A factor analysis of the measures was planned to reveal whether the hypothesized components of deindividuation occur together within individuals. Another purpose of the study was to compare the characteristics of deindividuation and lack of self-awareness resulting from nonsocial causes. Finally, the study was to ascertain the characteristics of deindividuation beyond those that define the state, for example, whether altered state experiences accompany deindividuation.

Method

Overview

The experiment was a 2 × 3 (Sex × Condition) design with the three conditions being deindividuation, non-self-aware, and self-aware. Participants were 59 male and 67 female subjects with 42 persons in each of the three conditions. After a half hour of group activities that were designed to create the three levels of group feeling and self-awareness, subjects were allowed to participate in activities that might "enhance their creativity for a later creativity test." This subsequent session, which was labeled the "individual" session for subjects, was actually the session in which evidence of the behavioral disinhibition was gathered. In the "individual" session participants could choose from disinhibited or inhibited tasks. Subjects were also asked to make creative verbal statements (which could be inhibited or disinhibited) at certain times during the so-called individual session. These two activities following the manipulation session served as the measures of disinhibition. After these activities, participants completed memory measures and an inventory assessing self-reported self-consciousness, group unity, and so on.

Setting

A group of eight same-sex introductory psychology students participated in each session, but unknown to the two real subjects, six of the persons were confederates who manipulated group factors in a preprogrammed way. Although the experimenter was always a male, he was assisted in running the experiment by two persons who were the same sex as the confederates and the subjects at that session. The eight participants (six confederates and two subjects) were either all males or all females at any one session. The experiment was run in a large room during evening hours and was said to be concerned with creativity. Subjects were told that the first session was a series of group activities designed to limber their minds up and get them in touch with their feelings. This group session provided the context for the manipulations. Subjects were told they would next work in an "individual session" in which they would all still be together, but in which each participant could choose the tasks he or she wanted to work at. Other sessions that would test creativity were alluded to later in the period. In other words, the first two sessions were presented as though they were designed to enhance creativity. Following these sessions, however, the two subjects actually completed the memory measure and questionnaire.

Conditions

A series of activities was developed for the three conditions. The manipulations for each condition during the group session lasted about 30 minutes. Male subjects only were run on Monday and Wednesday evenings and females on Tuesday and Thursday evenings, with the conditions being systematically rotated over evenings. The study occurred over an 8-week period. Subjects were phoned a day ahead of their appointment and asked to wear old clothes to the study.

Self-awareness condition. Subjects in the self-aware conditions were exposed to a group atmosphere and to activities designed to heighten feelings of individuality and self-awareness. In the waiting room the confederates acted fidgety and self-conscious. After the initial instructions the group was divided into a larger group (the confederates) and a smaller group (the two subjects) and taken to separate rooms to work at different tasks. It was stated that later all subjects would rotate in pairs through the one task, and in the meantime the others would complete some questionnaires. Once in the experimental room, the two subjects were asked to put on mechanics' long-sleeved coveralls to "protect their clothes" during the possibly messy later activities.

The two subjects were given name tags, seated at a large table, and then exposed to a series of manipulations designed to heighten their self-awareness. These were presented as ways of helping them get in touch with feelings and their inner self. Subjects, addressed

by name by the experimenter, first listened to music and answered questions about how it matched their personality characteristics. Next they wrote an essay about themselves, covering such topics as what makes them unique. They also listed their hometown, hobbies, personal motto, and other individuating information. They next read all the personal information to the other subject (being allowed to omit anything they felt was too personal). The subjects then answered questions about themselves compared to the other subject, for example, "Who do you think is probably more creative?"

At this point the research assistant brought in the six confederates who had ostensibly been working at a different task in another room. The assistant said that they had been told about the coveralls but that they preferred not to wear them. Since the phone caller who reminded them of their appointment had asked them to wear old clothes, they were not concerned if their clothes got dirty. Several confederates agreed that they preferred not to wear the coveralls and the experimenter, seemingly surprised, allowed them to proceed without coveralls if they weren't "worried if their clothes get dirty." Therefore, the two subjects stood out from the rest of the group because of their attire. The experiment next entered the "individual" phase of the study. At this point, the participants in the self-aware condition were dressed differently from the six confederates, had been exposed to a series of individuating tasks, and were part of a noncohesive group.

Non-self-aware condition. The confederates in this condition were pleasantly friendly but not outgoing. Subjects were exposed to a series of manipulation activities designed to focus their attention outward. The confederates and subjects performed the activities as a group throughout this condition and were all dressed in the coveralls. For example, they listened to music and rated its qualities and also rated jokes. They wrote how universities would be different in 100 years. Subjects also wrote a rebuttal to a letter that criticized university students. Next they worked on a series of challenging and fun games and puzzles. The emphasis in each activity was to relax and become more creative, and it was stated that none of the activities would be evaluated.

During the "individual" session these subjects were asked to press a pedal rhythmically throughout the session. Although this was said to be a trial activity to help judge the effects of rhythm on creativity, it was actually designed to be slightly distracting and thus reduce self-awareness (Duval & Wicklund, 1972) as the session continued. No names or name tags were used in this condition. Thus, this condition was designed to involve subjects in activities that would not accentuate individuality but would direct their attention outward away from themselves. No competition was created within the group, but neither was a spirit of group unity generated.

Deindividuation condition. In the waiting room the confederates were friendly and attempted to create a close group feeling in this condition. They spoke to the entire group, not to single individuals. The experimenter asked them to adopt a group name, and he thereafter called the group by this name. He said this name would be used to label their data, which were to be "treated as a group, not on an individual basis." The group was given the coveralls, and the confederates made putting them on a unifying and fun event. The first activities were designed specifically to draw the group together. Group members sang together and then did "elevation"—getting "psyched up" and together lifting an assistant high in the air. The group next formed a circle by interlocking arms and were told to prevent one of the assistants from breaking into the circle. The next activity, African dancing, was designed to continue the group-cohesive activities, but was also intended to produce the arousal, group coordination, and kinesethetic sensory experience that seems to be a part of many groups in which disinhibition occurs (Sargant, 1975). During the loud Burundi drum music, the group clapped in unison, swayed in unison, and danced around in a circle together. The lights in the room were dimmed. Thus, subjects in this condition were exposed to a variety of manipulations designed to create a unified group. To carry on the group feeling during the individual disinhibition session (in which subjects were seated around a large table) the confederates made occasional remarks to the group.

Measures

There were four major sets of measures: (a) activity disinhibition during the individual task, (b) speech disinhibition during the individual task, (c) memory measures taken after the "individual" session, and (d) a postexperimental questionnaire measuring group feeling, self-consciousness, altered state experiences, and so on.

Disinhibition tasks. Immediately after the manipulation session in which varying levels of group feeling and self-consciousness were induced, subjects moved to the "individual" session. They made their choices of the tasks they wanted to do, and each took a seat around a large square table. In front of each participant was a cue light, a microphone, and a list of the activity boxes available for this part of the experiment.

It was desirable to use tasks that were different in their social appropriateness but roughly equivalent in interest and fun. Two sets of tasks were created, disinhibited tasks in colorful boxes and inhibited tasks in dark brown boxes. In a pilot study 12 male and 13 female undergraduates had rated a large list of potential tasks on a scale for inhibition–disinhibition and a scale for boring/aversive versus interesting/fun. Twenty disinhibited tasks were selected and matched with 20 inhibited tasks that averaged the same score on interesting/fun. Since there were sex differences in the ratings, the 40 tasks were selected so that these differences were minimized on both the interest and disinhibition dimensions. On the 9-point disinhibition scale, the 20 inhibited tasks had a mean of 3.15, and the 20 disinhibited tasks averaged 6.45. However, the interesting/fun value of the tasks was similar (5.40 vs. 5.25). Examples of tasks and their ratings are shown in Table 1.

The tasks were grouped together in sets of four, and these tasks all came together in a box. For example, an inhibited box labeled *Intellectual challenges* contained

Table 1
Examples of Tasks and Ratings

Task	Interesting/fun rating[a]	Disinhibition rating[b]
Inhibited task		
Crossword puzzles	4.34	2.39
Reading about disarmament	6.64	2.00
Answering moral dilemma questions	5.57	4.01
Playing "Hi-Q"	3.83	3.71
Answering comprehension questions about a story	6.41	3.17
Disinhibited task		
Playing in mud	4.84	7.18
"Finger painting" with your nose	4.93	7.71
Writing the faults of your friends	6.24	5.03
Sucking liquids from baby bottles	6.34	6.84
Writing down all the obscenities you can think of	5.80	5.52

[a] Low = fun. [b] Low = inhibited.

four tasks such as "Hi-Q," pencil mazes, and so on. There were five types of boxes for the inhibited tasks and five types of boxes for the disinhibited ones. Each type of box had a label that broadly described the tasks inside (e.g., *Unusual foods*, *Desecration*, and *Public issues*). These box labels were also selected based on preratings so that they differed on rated disinhibition (7.9 vs. 4.16) but not on interesting/fun (4.2 vs. 4.38).

Subjects were told that during the individual session (lasting 20 minutes) they could choose any box they liked. They could see the boxes and their labels but did not know the contents (tasks) unless they selected that box. They could work in a box for as long or as short a time as they liked and do any number of tasks within it. When they tired of that box, they were to select another. Initial selections were made on a sheet of paper that contained a list of all the inhibited boxes in one column and all the disinhibited tasks in the other. Subjects were asked to make three ranked choices initially in case "we run out of a particular box." These three initial choices were weighted by their order to produce a measure of disinhibition of the initial choices. The experimenter and an assistant brought the new boxes when they were desired.

Subjects were told before the individual session that they should feel free to choose either type of box, since both "can probably enhance creativity." They were told that in the dark boxes were intellective tasks that could stimulate their minds and get them thinking quickly. They were reminded that quick thinking is an essential part of creativity. The experimenter said that the tasks in the colorful boxes could also enhance creativity by helping them get in touch with feelings and sensations. It was repeated that they should feel free to switch boxes whenever they liked and that they could choose any box or combination of boxes they wanted.

Two confederates seated at the far end of the table from the subjects recorded which boxes and tasks the subjects chose. Since these confederates were apparently working in boxes requiring writing, their writing did not seem strange and never attracted attention. The recording confederates also wrote down a description of the subject's behavior. These descriptions were later rated for the degree of disinhibition exhibited by the subject by raters who were blind as to the subject's condition. Each confederate chose the same boxes regardless of condition and did the same tasks in every condition. Four of the six confederates initially chose disinhibited boxes, and then they switched tasks in a preprogrammed regular way, the overall new choices being approximately half disinhibited boxes.

Speech measure. Subjects were asked to make several "creative statements" into a microphone during the individual session. These statements were to be made whenever a cue light came on in front of the person. The order of lights was always the same and allowed each subject to make three statements and each confederate to make two. The confederates made the same standard statements in all conditions, with some of their statements being rather unusual, creative, or disinhibited, and some being mundane and ordinary. An example of a statement is, "This reminds me of when I was young, exploring new things and acting crazy. I feel like a child again without a care in the world."

Subjects were told in the self-aware condition that their statements would be individually recorded. In the non-self-aware condition it was mentioned simply that the statements would be recorded. In the deindividuation condition it was stated that all the group would be recorded together and labeled with the group name. The recorded statements of the subjects were rated by two raters who scored the subject's three statements on inhibition–disinhibition. The ratings for the subjects were averaged across their three speeches. The two raters' scores were correlated across subjects, and since the interrater agreement was high ($r = .82$), the average score for the two raters was assigned to that subject.

Postexperimental questionnaire. After the individual session, subjects were taken to a different room and the research assistants administered the postexperimental

Table 2
Factor Loadings for Two Orthogonal Factors

Variable	Loading	
	Factor 1[a]	Factor 2[b]
Reported self-consciousness	−.73	
Tasks completed disinhibition score	.61	
Initial boxes choice (weighted)	.58	
Spontaneity versus conscious planning	.54	
Time seemed to go quickly	.53	
Lack of concern for what other subjects would think of you	.53	
Liking for the group	.51	
Feeling of group unity	.44	
Sum of unusual experiences reported		.77
Feeling of group unity		.61
Altered state self-report		.59
Liking for the group		.58
Subjective loss of individual identity		.44
Memory for confederate statements[c]	.07	−.01
Match statements to pictures[c]	.16	.07
Recognition of own statements[c]	.14	.03
Subjective feeling of anonymity[c]	.03	.39
Disinhibition of speech[c]	.37	.04

[a] Deindividuation. [b] Altered experiencing. [c] Did not load above .4 on either factor.

questionnaire and memory measures to them. The questionnaire contained a series of items, most rated on 9-point bipolar scales: (a) subjective feeling of time passage; (b) self-reported self-consciousness; (c) memory for the tasks the person next to them did (essay format); (d) perceived similarity to other participants and feeling of group unity and liking; (e) enjoyment of the session; (f) reports of altered perceptions, thinking, and emotions; (g) reports of an altered state of consciousness (yes or no answer) and other unusual experiences such as hallucinating; (h) the extent to which their behavior was guided by conscious planning versus flowing with the environment and making spontaneous choices; (i) memory measures for peripheral cues in the room such as the content of posters on the walls; and (j) suspicion measures, including suspicion of the confederates and guessing the purpose of the experiment.

Memory measures. Each subject was asked to select the photographs of persons who had been in their group and of persons not in their group. There were pictures of all six confederates and six filler photos of other undergraduates. Subjects were required to sort all of the pictures into two piles. Thus, a subject could score from 0 to 12 pictures correct. The same procedure was carried out for statements that had been made during the session. Ten correct confederate statements (out of the 12 made) written on cards were presented together with 10 filler statements that had not been said, and subjects were asked to sort them all into two piles. Next, the 6 veridical pictures and the 10 actual statements were given, and subjects were asked to match all the statements with the picture of the person who said them, even if they had to guess on some. Subjects could not keep track of how many statements each of the confederates made while they were involved in the task. Also, they did not necessarily receive two statements from each confederate. Thus, participants could match from zero up to several statements to a particular confederate. Therefore, the task was really one of matching each statement to the confederate who was most likely to have made it. Lastly, subjects were given six statements and asked to select the three they had made.

Ethical Precautions

Since the study involved deception and some unusual activities, a series of ethical precautions were used, and these are briefly described here. The initial instructions emphasized subjects' right to withdraw at any time and their right to refuse to do anything they did not wish to do in the experiment. Subjects were cautioned not to engage in any physical activities contraindicated by health problems. The confederates intervened if any subject's behavior became too extreme during the individual session (e.g., eating mud). The debriefing was conducted carefully and took 15–20 minutes. Each subject was introduced to facts about the study in a graduated way as recommended by Mills (1976), was promised anonymity, and his or her feelings about the study were discussed. In general, the ethical guidelines recommended by Diener and Crandall (1978) were followed. Many subjects afterward said it was the most interesting study in which they had participated.

Results

Factor Analysis

Sixteen of the major variables were subjected to a principal axis factor analysis with multiple correlations employed as the communality estimates. A solution with two factors was chosen based upon the scree of the eigenvalues (Cattell, 1966), the interpretability of the data, and the factor loadings. The solution was subjected to an orthogonal (varimax) rotation. For both factors the items with absolute weights above .4 are shown in Table 2.[1] The factor loadings reflect simple structure except for the case of group liking and group unity, which loaded substantially on both factors. The five variables that did not load above .4 on either factor are also shown in Table 2. The major factor contains the key elements of deindividuation, including group unity and liking, low self-awareness, spontaneity, and disinhibited behavior. The second factor also contained group unity and liking, but was otherwise composed of altered experience items. Thus, across subjects the key variables formed two major independent factors, one that can be labeled Deindividuation and the other that can be labeled Altered Experiencing. It appears that the memory items were mainly those that did not load on either factor. The first-order correlation between conscious planning and group unity was $r(124) = -.45$, $p < .001$. The correlation between the disinhibition score of tasks chosen and self-awareness was $r(124) = .38$, $p < .001$.

Effect of Conditions

The two factor scores for each subject were standardized, and a multivariate analysis of variance (MANOVA) was performed on the two scores. These data were collapsed across sex because this factor produced negligible differences on most variables. Thus, the MANOVA was computed on the three conditions with the two dependent variables. Pillai's Trace Criterion V transformed to a standard F revealed that the three conditions were significantly different, $F(4, 246) = 8.79$, $p < .001$. This method is one of several for calculating a MANOVA significance test, and is suitable for the present data (Olson, 1976). A multivariate extension of omega-square revealed that condition accounted for 23% of the variability in the dependent measures. The mean standardized factor scores and the univariate tests on the factors are shown in Table 3. Newman-Keuls comparison of means (significant differences designated by the subscripts in Table 3) reveals that the deindividuation group scored significantly higher on Factor 1 than the other two groups, who did not differ significantly from each other. On Factor 2 the deindividuation group was significantly higher than the self-aware group only.

The factors were also subjected to a discriminant analysis, and the discriminant functions were used to predict the probability that each subject was a member of each condition (Cooley & Lohnes, 1962). This procedure properly classified 74% of the deindividuation subjects (gave the largest probability that they were in the deindividuated group), 21% of the non-self-aware subjects, and 60% of the self-aware subjects. This procedure predicted that 33% of the non-self-aware subjects belonged to the deindividuated group and that 45% belonged to the self-aware group. It appears that some subjects in the non-self-aware group reacted in a manner similar to the deindividuation subjects and some in a way similar to the self-aware subjects.

Univariate data and analyses of variance on the individual variables are also shown in Table 3. Although sex of subject did produce significant main effect differences on a few of the individual variables, it did not result in significant interactions with condition, and these results tended not to be theoretically interesting. For this reason the sex data and analyses are omitted from Table 3. Although

---

[1] A stable factor-analytic solution depends on having a large number of subjects relative to the number of items to be factored. Therefore, a number of items were not included in the factor analysis, bringing the number of factored items to 16. The 10 items that were excluded either were redundant with other items or were of lesser importance to the theory of deindividuation. The excluded items were rated disinhibition of behavior; concern over what experimenter would think of you; similarity to the group; enjoyment of the group session; emotions, thinking, and perceptions altered from normal (3 items); memory for the confederate photographs, for peripheral cues, and for tasks neighbor did (3 items).

Table 3
Mean Scores for Factors and Individual Variables

Variable	Condition: Deindividuated	Condition: Non-self-aware	Condition: Self-aware	Univariate *F*
Factor 1 standardized scores (Deindividuation)	$.60_a$	$-.12_b$	$-.47_b$	11.46****
Factor 2 standardized scores (Altered experiencing)	$-.37_a$	.02	$.35_b$	4.69*
Rated disinhibition of behavior	5.9_a	4.0_b	2.4_c	30.72****
Tasks completed disinhibition score	$5.5_{a,x}$	4.7_y	4.5_b	5.63**
Initial boxes choice (weighted)	$6.7_{a,x}$	5.8_b	6.0_y	5.12**
Disinhibition of speech	6.5_a	$3.7_{b,x}$	$2.6_{b,y}$	42.78****
Reported self-consciousness	3.8_a	4.4	5.3_b	4.55*
Conscious planning not spontaneity[a]	6.0_a	4.3_b	4.5_b	7.24**
Loss of individual identity[a]	$24\%_x$	$7\%_y$	$7\%_y$	$\chi^2 = 7.00$*
Subjective feeling of anonymity[a]	4.0	4.5	4.8	1.75
Concerned other subjects think of you[a]	6.3	6.0	5.2	2.39
Concerned experimenter thinks of you[a]	7.2	6.6	6.8	.60
Liking for the group[a]	$2.9_{a,x}$	3.9_y	$4.9_{b,z}$	9.36***
Feeling of group unity[a]	3.2_a	$4.6_{b,x}$	$5.7_{b,y}$	14.68****
Similarity to group[a]	4.4_x	5.4_y	5.5_y	3.56*
Enjoyed group session[a]	3.1_a	4.5_b	4.3_b	6.42**
Emotions altered from normal	88%	76%	74%	$\chi^2 = 5.87$
Time seemed to go quickly[a]	3.2	3.3	3.9	2.00
Thinking seemed altered from normal	6.1	5.5	6.1	.83
Altered state self-report	39%	28%	21%	$\chi^2 = 2.83$
Sum of unusual experiences reported	1.2	1.0	.7	3.00
Altered perceptions	$56\%_x$	$28\%_y$	$34\%_y$	$\chi^2 = 6.52$*
Memory for confederate statements	13.9_x	14.8_y	15.0_y	3.38*
Memory for photographs	8.8_x	8.6	8.1_y	3.20*
Match statements to pictures	3.9	4.2	4.2	.35
Memory peripheral cues	$.9_x$	1.4	1.8_y	3.32*
Recognition of own statements	5.9	5.9	5.9	.04
Memory for tasks neighbor did	.9	.7	.6	1.55

Note. Multiple comparisons conducted only where main effect was significant. Means across the same row with the same subscripts in the same letter series are not significantly different (Newman-Keuls or χ^2) at $p < .01$ (abc) or $p < .05$(xyz).
[a] Items scored so that a low score indicated a greater magnitude of the item as labeled.
* $p < .05$. ** $p < .01$. *** $p < .001$. **** $p < .0001$.

the MANOVA was highly significant, the univariate alpha levels should be treated cautiously, since there are a large number of non-independent tests presented. In addition to the traditional *F* values shown in Table 3 for each variable, the results of Newman-Keuls multiple comparisons between the three means are shown in cases where the overall *F* was significant.

The variables shown in Table 3 are grouped together into classes of phenomena. Clearly, the largest differences between the experimental conditions came on the group feeling items and on the disinhibition items. The measures of disinhibition showed differences between the three conditions that were highly significant, and the Newman-Keuls analyses indicated that for these variables, all three groups were significantly different from the others. For the group feeling items, each of the three conditions also tended to be significantly different from the other conditions.

The self-consciousness and loss of identity items also showed differences between groups. The deindividuation group was significantly different from both of the other two groups on two of three key items. Although feelings of anonymity and concern over the evaluation of others did not reach an alpha level of .05, the deindividuation group did score in the predicted direction on these items. The smallest differences between conditions were on the altered state and memory items. It can be seen in Table 3 that most of the altered state items were not significant, although the deindividuation group was most extreme on these variables. For the memory items, the deindividuation group had the worst memory for the statements of the confederates and for cues in the room, but the best memory for the faces of the confederates, which is similar to the pattern found by Diener and Kasprzyk (Note 1). Finally, it can be noted that subjects' recognition of their own statements was almost perfect in each condition.

Analyses for Artifacts

There were several potential problems in the present study, and these should be analyzed. One question concerns the unit of statistical analysis used earlier. Individuals, not pairs of subjects, were employed in the analyses although the pairs of individuals might not be independent, since they could have influenced each other. This seems unlikely, since the majority of the group were confederates, and they set the tone of the group by acting similarly to each other. However, the possibility that subjects within the subject pairs influenced each other was checked by correlating within conditions each factor score for pairs of subjects run together. Thus, correlations for each of the conditions were derived that represented the similarity of the responses of pairs of subjects run together. These correlations were all near zero and averaged .06, revealing that the data for each subject were independent of the other subject's scores.

Another question concerns suspicion. When asked to guess the purpose of the study, fewer than 15% of the subjects had any idea beyond a study of creativity, and only a few subjects guessed anything related to the true purpose. Similarly, when asked about their suspicion of the confederates, the vast majority of subjects checked the choice saying they honestly never suspected any of the confederates. There were virtually no differences between conditions on the suspicion items. The two suspicion measures did not correlate with the number of weeks (rs = .079 and −.069), indicating that suspicion was not increasing during the semester as subjects were becoming more knowledgeable about psychology. This also suggests that leakage about the study into the subject pool was not occurring to any extent, which is consistent with past findings (Diener, Matthews, & Smith, 1972). Finally, the correlations between subjects' factor scores and their suspicion scores were low, ranging from −.06 to .08, with a mean of .01. Thus, there appeared to be low levels of suspicion, and what little suspicion there was did not seem to influence the data.

Discussion

The present results support the importance of the deindividuation construct, which relates social situational factors to lack of self-awareness and in turn relates lowered self-awareness to disinhibited behavior. The factor analysis revealed that characteristics such as lack of self-awareness, lack of conscious planning, group unity, disinhibited behavior, and lack of concern about what others would think of the subject all formed a single central factor. A comparison of factor scores revealed that subjects exposed to the deindividuation manipulation scored significantly higher than the other two groups on this factor. Thus, the two types of analyses converged in suggesting the validity of the deindividuation construct.

It should be noted that there was a significant first-order correlation between disinhibited behavior and lack of self-consciousness. This indicates that less self-aware participants tended to exhibit more disinhibited behaviors. The question arises as to whether low self-awareness caused disinhibited behavior or the reverse. This question cannot be answered with certainty because of the correlational nature of the findings. However, the correlation of −.35 between self-awareness and disinhibition of the initial activity choice (before any activities were done) suggests that the deviant

activity did not cause the lowered self-awareness. We know from the self-awareness literature that persons with lower self-awareness are less likely to adhere to norms, and therefore we can conjecture that the lower self-awareness exhibited by deindividuated subjects could certainly have had some causal role in disinhibiting behavior.

The present findings clearly support Diener's (1979) theory of deindividuation, which integrates the concepts of self-awareness and self-regulation with earlier conceptions of deindividuation. Lack of self-awareness and lack of conscious planning, group unity, and disinhibited behavior did occur together within individuals. Deindividuating conditions also produced these changes in subjects. The negative correlation between self-awareness and behavioral disinhibition also supports the theory of self-awareness (Wicklund, 1975, 1979). According to the theories of self-awareness and deindividuation, lowered self-awareness leads to lessened self-regulation because the person is less aware of behavior–norm discrepancies and less likely to plan for the future. The present results are compatible with these formulations.

The study provides a less clear answer to the second question addressed—whether deindividuation is similar to a lack of self-awareness occurring because of nonsocial causes. The deindividuation group was significantly higher on Factor 1 than the non-self-aware group, which did not differ significantly from the self-aware group. It can be noted that non-self-aware subjects, although somewhat less self-aware than self-awareness condition subjects, did not more frequently report a loss of identity or lower conscious planning. In addition, classification of subjects based on the discriminant function revealed that only a third of the non-self-aware subjects were grouped together with the deindividuation participants. Social causes of lowered self-awareness may be more likely to release dishibited behavior because the person can perform novel or antinormative behaviors and not become self-aware, whereas the person whose attention is focused outward by nonsocial factors can easily focus back on the self when confronted by norm-related choices or a novel situation. It is also possible that a unique characteristic of deindividuation is that the person not only loses a sense of individual identity but also goes along with the direction of group behavior.

The answer to whether deindividuation is accompanied by altered experiences is left uncertain by the present data. The deindividuation group differed significantly from the non-self-aware group on the second factor, which included several altered state items. About 40% of deindividuated subjects answered "yes" to the statement, "Felt like I was in an altered state of consciousness in which my perceptions or thinking were distorted compared to their normal functioning." It should be recalled, however, that a fairly large number of persons in the other groups also reported such experiences (about 25%) and that the differences between groups on these items were not significant in most cases. Thus, more research on this issue is necessary before firm conclusions can be reached.

One additional characteristic that has been proposed as part of deindividuation is a memory decrement (Festinger et al., 1952; Zimbardo, 1970). In the present study the memory for what had been said by others was poorest for the deindividuation group. However, subjects' memory for what they had said was uniformly high across conditions, indicating that even when self-awareness is low, the person can often still remember later what he or she has done. In addition, the deindividuation group had higher memory scores for pictures of the other group members, perhaps because they were more involved with the group. The deindividuated subjects were somewhat poorer at memory for peripheral cues. This pattern suggests that "memory" differences actually result from attentional differences in the various conditions and that deindividuation does not automatically lead to reduced memory or amnesia.

Alternative Explanations

There are a number of alternative explanations for the present results.[2] It is possible that

---

[2] One other alternative explanation for the present findings is in terms of demand characteristics. The major reason to discount a demand characteristics explanation of the findings is that it does not explain

subjects who conformed more to the disinhibited atmosphere in the deindividuated condition simply liked the group more, and they therefore complied with the demands of the setting. It should be noted that the conformity explanation does not contradict the idea of deindividuation. Indeed, compliance or conformity to group behavior in the initial stages of group activity may often set the stage for deindividuation and for the group to thereafter move in increasingly disinhibited directions. However, it does not appear that conformity is the sole explanation for the present findings. First, deindividuation subjects chose more disinhibited boxes immediately after the group manipulation and before they knew what other subjects had chosen. And second, deindividuation subjects showed a variety of responses that are not explained simply by conformity, for example, more frequently reporting a loss of identity, lower self-consciousness, and less conscious planning.

---

why the deindividuation items formed a coherent factor, as predicted by the theory. Another reason to discount this possibility is that few subjects guessed the true purpose of the study, and those few who did responded no differently from those who did not.

Reference Note

1. Diener, E., & Kasprzyk, D. *Causal factors in disinhibition by deindividuation*. Unpublished manuscript, University of Illinois at Urbana-Champaign, 1978.

References

Bandura, A. Self-reinforcement: Theoretical and methodological considerations. *Behaviorism*, 1976, *4*, 135–155.

Baron, R. S. Anonymity, deindividuation and aggression (Doctoral dissertation, University of Minnesota, 1970). *Dissertation Abstracts International*, 1971, *32*, 533A. (University Microfilms No. 71-18,681)

Carver, C. S. Facilitation of physical aggression through objective self-awareness. *Journal of Experimental Social Psychology*, 1974, *10*, 365–370.

Carver, C. S. Physical aggression as a function of objective self-awareness and attitudes toward punishment. *Journal of Experimental Social Psychology*, 1975, *11*, 510–519.

Carver, C. S. A cybernetic model of self-attention processes. *Journal of Personality and Social Psychology*, in press.

Cattell, R. B. The scree test for the number of factors. *Multivariate Behavioral Research*, 1966, *1*, 245–276.

Cooley, W. W., & Lohnes, P. R. *Multivariate procedures for the behavioral sciences*. New York: Wiley, 1962.

Diener, E. Effects of prior destructive behavior, anonymity, and group presence on deindividuation and aggression. *Journal of Personality and Social Psychology*, 1976, *33*, 497–507.

Diener, E. Deindividuation: Causes and consequences. *Social Behavior and Personality*, 1977, *5*, 143–155.

Diener, E. Deindividuation: The absence of self-awareness and self-regulation in group members. In P. Paulus (Ed.), *The psychology of group influence*. Hillsdale, N. J.: Erlbaum, 1979.

Diener, E., & Crandall, R. *Ethics in social and behavioral research*. Chicago: University of Chicago Press, 1978.

Diener, E., Dineen, J., Endresen, K., Beaman, A. L., & Fraser, S. C. Effects of altered responsibility, cognitive set, and modeling on physical agression and deindividuation. *Journal of Personality and Social Psychology*, 1975, *31*, 328–337.

Diener, E., Fraser, S. C., Beaman, A. L., & Kelem, R. T. Effects of deindividuating variables on stealing by Halloween trick-or-treaters. *Journal of Personality and Social Psychology*, 1976, *33*, 178–183.

Diener, E., Matthews, R., & Smith, R. E. Leakage of experimental information to potential future subjects by debriefed subjects. *Journal of Research in Personality*, 1972, *6*, 264–267.

Diener, E., & Srull, T. K. Self-awareness, psychological perspective, and self-reinforcement in relation to personal and social standards. *Journal of Personality and Social Psychology*, 1979, *37*, 413–423.

Diener, E., & Wallbom, M. Effects of self-awareness on antinormative behavior. *Journal of Research in Personality*, 1976, *10*, 107–111.

Diener, E., Westford, K. L., Dineen, J., & Fraser, S. C. Beat the pacifist: The deindividuating effects of anonymity and group presence. *Proceedings of the 81st Annual Convention of the American Psychological Association*, 1973, *8*, 221–222. (Summary)

Dion, K. L. Determinants of unprovoked aggression (Doctoral dissertation, University of Minnesota, 1970). *Dissertation Abstracts International*, 1971, *32*, 534A. (University Microfilms No. 71-18, 716)

Dipboye, R. L. Alternative approaches to deindividuation. *Psychological Bulletin*, 1977, *84*, 1057–1075.

Duval, S., & Wicklund, R. A. *A theory of objective self-awareness*. New York: Academic Press, 1972.

Festinger, L., Pepitone, A., & Newcomb, T. Some consequences of deindividuation in a group. *Journal of Abnormal and Social Psychology*, 1952, *47*, 382–389.

Ickes, W., Layden, M. A., & Barnes, R. D. Objective self-awareness and individuation: An empirical link. *Journal of Personality*, 1978, *46*, 146–161.

Kanfer, F. Self-regulation and self-control. In H. Zeier (Ed.), *The psychology of the 20th century*. Zurich: Kindler Verlag, 1977.

Maslach, D. Social and personal bases of individuation. *Journal of Personality and Social Psychology*, 1974, *29*, 411–425.

Mills, J. A procedure for explaining experiments involving deception. *Personality and Social Psychology Bulletin*, 1976, *2*, 3–13.

Olson, C. L. On choosing a test statistic in multivariate analysis of variance. *Psychological Bulletin*, 1976, *83*, 579–586.

Rule, B. G., Nesdale, A. R., & Dyck, R. Objective self-awareness and differing standards of aggression. *Representative Research in Social Psychology*, 1975, *6*, 82–88.

Sargant, W. *The mind possessed*. New York: Penguin Books, 1975.

Scheier, M. F., Fenigstein, A., & Buss, A. H. Self-awareness and physical aggression. *Journal of Experimental Social Psychology*, 1974, *10*, 264–273.

Watson, R. I. Investigation into deindividuation using a cross-cultural survey technique. *Journal of Personality and Social Psychology*, 1973, *25*, 342–345.

Wicklund, R. A. Objective self-awareness. In R. L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 8). New York: Academic Press, 1975.

Wicklund, R. A. Group contact and self-focused attention. In P. Paulus (Ed.), *The psychology of group influence*. Hillsdale, N. J.: Erlbaum, 1979.

Ziller, R. C. Individuation and socialization: A theory of assimilation in large organizations. *Human Relations*, 1964, *17*, 341–360.

Zimbardo, P. G. The human choice: Individuation, reason and order versus deindividuation, impulse and chaos. In W. J. Arnold & D. Levine (Eds.), *Nebraska Symposium on Motivation* (Vol. 17). Lincoln: University of Nebraska Press, 1970.

Received July 10, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 7, 1172–1178

Effects of Difficulty and Diagnosticity on Choice Among Tasks in Relation to Achievement Motivation and Perceived Ability

Ursula Buckert
Oberhausen, West Germany

Wulf-Uwe Meyer
Universität Bielefeld, West Germany

Heinz-Dieter Schmalt
Ruhr-Universität Bochum, West Germany

The experiment is a partial replication of a study conducted by Trope. It investigates the effects of two person characteristics (achievement motive and perceived own ability) and two task characteristics (difficulty and diagnostic value about own ability) on choice among achievement tasks. In accordance with the results of Trope, it was found that high-diagnostic tasks were preferred to low-diagnostic tasks, independent of their difficulty. Trope's finding that high resultant achievers choose high-diagnostic tasks over low-diagnostic tasks to a greater extent than low resultant achievers was not replicated. However, the perceived degree of own ability affected choice behavior: When easy and difficult tasks were both high in diagnosticity, subjects high in perceived ability preferred difficult over easy tasks, whereas subjects low in perceived ability preferred easy over difficult tasks. From this latter finding it is concluded that a self-informational conception of choice behavior has to include the subjective probability of success at tasks as a determinant of choice, in addition to objective difficulty and diagnostic value.

According to Atkinson's (1957, 1964) theory of achievement strivings, individuals high and low in resultant achievement motivation differ systematically in their choices of achievement tasks. High resultant achievers will select tasks that are intermediate in subjective probability of success, whereas low resultant achievers prefer to undertake tasks of high or low subjective probability (P_s). These predictions are based on a hedonic conception of achievement-related behavior. It is assumed that the choice of tasks intermediate in P_s maximizes positive affect (pride), prospectively connected with success, for individuals high in resultant achievement motivation, and that the selection of tasks high or low in P_s minimizes negative affect (shame), prospectively connected with failure, for persons low in resultant achievement motivation. In the relevant empirical studies undertaken by Atkinson and his co-workers, P_s is equated with the difficulty of the experimental tasks.

An alternate conception of achievement-related choice was developed by Weiner et al. (1971). This approach, which is based on the principles of attribution theory, explains the assumed differential task selection of the two motive groups in terms of the expected self-informational feedback of an outcome. It is assumed that high resultant achievers are motivated to obtain information about their capabilities; they therefore select tasks of intermediate difficulty because success and failure at these tasks yield the most information about the person undertaking these tasks. On the other hand, low resultant achievers are assumed to be motivated to avoid this information; they therefore select very easy or very difficult tasks. This is assumed because success at easy tasks and failure at difficult tasks (the usual outcomes) yield relatively

---

This research was conducted by the first author and was part of her diploma thesis.

Requests for reprints should be sent to Wulf-Uwe Meyer, Universität Bielefeld, Abteilung Psychologie, Postfach 8640, 4800 Bielefeld, West Germany.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3707-1172$00.75

little information about one's ability. This lack of self-informational feedback results from the fact that outcomes consistent with social norms lead to causal inferences about the task rather than about the person undertaking the task (Heider, 1958; Kelley, 1967, 1973). Thus, the same pattern of choices for the two motive groups is predicted from Atkinson's hedonic and Weiner et al.'s informational approaches to achievement behavior.

Two recent studies by Trope and Brickman (1975) and Trope (1975) unconfounded the factors that lead to these identical predictions that are derived from Atkinson's achievement motivation theory and Weiner et al.'s attributional conception. These authors argued that the relationship between task difficulty and expected self-informational value, postulated by Weiner et al. (1971), is not universally valid. The informational value also depends on the diagnosticity of tasks, which is operationally defined as the difference in the percentage of success at a given task between individuals high versus low in ability. For example, suppose that 100 people, 50 high in ability and 50 low in ability, first perform Task A and then Task B with the two possible outcomes, success or failure, at each. At each task 40 individuals succeed. The overall probability of success of both tasks then is .40. Suppose further that on Task A 40 from the group high in ability are successful and none from the group low in ability, whereas on Task B 22 out of the high-ability group have success and 18 from the low-ability group. It is clear that success at Task A distinguishes better between persons with different ability levels than success at Task B. Although the overall difficulty of both tasks is equal, Task A is more diagnostic of the ability level of a person than Task B.

Trope and Brickman pointed out that intermediate difficulty tasks are generally, but not necessarily, more informative (diagnostic) than easy or difficult tasks because tasks of intermediate difficulty mathematically can be made maximally diagnostic and therefore are ordinarily more efficient at discriminating between people with high and low ability. The authors conducted an experiment in which the variables, difficulty and diagnosticity of tasks, were independently manipulated. Groups of subjects were given a choice among easy, moderate, and difficult tests. The tasks of each test were highly diagnostic for some groups and low in diagnosticity for other groups. The results showed that tasks of high expected diagnostic value were preferred over tasks of low expected diagnostic value, independent of their difficulty. This result is interpreted by the authors as incompatible with Atkinson's theory because information seeking rather than maximizing positive affect (exhibited by a preference for intermediate difficulty tasks) seems to be the crucial determinant of choice behavior.

In a replication of Trope and Brickman's study by Trope (1975), resultant achievement motivation also was measured. The results again showed that high-diagnostic tasks were generally preferred to low-diagnostic tasks. Moreover, high resultant achievers preferred high over low-diagnostic tasks to a significantly greater extent than low resultant achievers. These results again are viewed by Trope as contradicting Atkinson's theory and giving some support to the position of Weiner et al. (1971) that high resultant achievers strive for ability-relevant information and that low resultant achievers avoid such information. Trope (1975), however, pointed out "that replication studies that employ other measures of achievement motive are definitely needed in order to firmly establish this conclusion" (pp. 1010–1011).

The present experiment is a partial replication of the studies conducted by Trope and Brickman (1975) and Trope (1975). It includes a well-validated German instrument for the assessment of resultant achievement motivation developed by Schmalt (1976). In addition to achievement motivation, the degree of ability that the subjects ascribe to themselves (perceived ability) is introducted here as a further independent variable; several authors have considered perceived ability as an important determinant of the achievement motive (Heckhausen, 1977; Meyer, 1976) or even equated differences in resultant achievement motivation with differences in perceived ability (Kukla, 1972, 1978; Meyer, 1973).

Method

Subjects

Subjects were 104 male students, 16 to 19 years old and attending Grades 11 and 12 of two German high schools (Gymnasium).

Design

The general features of the design and the procedure resembled those employed by Trope and Brickman (1975) and Trope (1975). The study included two experimental conditions with three tests varying in difficulty in each condition (easy ≈.70, moderate ≈.50, difficult ≈.30). In Condition 1, the test moderate in difficulty was high in diagnosticity, whereas the other two tests were low in diagnosticity. In Condition 2, the easy and the difficult test were high in diagnosticity, whereas the moderate difficult test had low diagnosticity. There were 51 subjects assigned to Condition 1 and 53 subjects to Condition 2.

Two other independent variables included in the study were the subjects' resultant achievement motivation and the degree of perceived ability that the subjects ascribed to themselves. Resultant achievement motivation was measured by the adult version of the Need Achievement (n Ach) Grid developed by Schmalt (1976). This instrument includes a series of depicted achievement situations with a fixed set of statements characteristic for hope of success and fear of failure orientation. The statements are taken from Thematic Apperception Test (TAT) scoring codes (Heckhausen, 1967) and are the same for each depicted situation. The subject has to mark those statements that would be characteristic for himself in each situation. The construct validity of the n Ach Grid is documented by a number of findings (see Schmalt, 1976).

Procedure

The experiment was run in groups ranging from 11 to 25 subjects, with each subject receiving a code to guarantee anonymity. First, the n Ach Grid was administered. The subjects then were told that the experimenter would administer a test called the Minnesota Multiphasic Integrative Orientation Test. To create the impression that the test really would be administered and that ability-relevant information could be gained by all subjects, the experimenter pointed out that she would visit the class three times. Subjects were told that in the first phase the test would be explained and that each subject would be given the opportunity to determine for himself the type of test items he would work on approximately 1 week later. In a second phase, they were told, the tasks selected by each individual would be administered. In a third phase, they were told, each subject would be given the opportunity to get information about the test result from the experimenter.

In accordance with the instructions used by Trope and Brickman (1975) and by Trope (1975), the subjects were then given a test booklet indicating that integrative orientation is characterized by being able "to generate new information from previously known information, to generalize, to find with relative ease new solutions to problems." It was emphasized that standard intelligence tests do not measure integrative orientation and that about half the students from high schools comparable in age were high in the ability and half were low. (For the rationale of this procedure, see Trope & Brickman, 1975.) Subjects were then told that the test consisted of three subtests, each consisting of 16 tasks. It was explained that subjects were to choose a total of 12 items and that they were free to decide for themselves how many items from each test they would like to work on.

Instead of leaving the tasks unspecified as Trope and Brickman did, examples of the tasks, supposedly used in the tests, were presented; they were taken from the Intelligenz-Struktur-Test (Amthauer, 1973). Two rows, with five cubes in each, were shown, with each cube carrying different symbols on the three sides presented. The cubes in the two rows were the same, but in the lower row they were turned, tipped over, or both, and had changed their positions in relation to the upper row. It was explained that pairs of identical cubes had to be identified. Following this presentation, subjects had to estimate the degree of their integrative ability on a 9-point scale. The scale was anchored at odd numbers with the following labels: very high (1), high (3), average (5), low (7), very low (9). Up to this point, the instructions were the same for all subjects.

Manipulation of difficulty and diagnosticity. To create the experimental conditions, subjects received different booklets. In these booklets the information about difficulty and diagnosticity was given in the form of bar diagrams that showed the percentage of students high and low in integrative ability who succeeded on the three subtests. The diagrams were accompanied by an explanation: For the high diagnosticity tests, the subjects were told that there was a relatively large difference between the successes of students high and low in integrative ability. Success or failure therefore would indicate the ability level. For low diagnosticity tests, subjects were told that this difference was relatively small, success and failure therefore would not indicate the ability level. Furthermore, the overall percentage of success indicated whether the test was easy, moderate, or difficult. Table 1 gives the reported percentages of success for high- and low-ability students in the two experimental conditions.

Measurement of dependent variables. After the above descriptions, subjects were given another booklet indicating that they were to choose a total of 12 items from the tests. It was stressed that the items selected were to be performed by the subject during the next test session. In the booklets again some short information about difficulty and diagnosticity of the three subtests was presented in the form of bar diagrams. The subjects were then asked to indicate how many items from each test they would like to be given at the next test session with the total number of items selected not exceeding 12. Furthermore, subjects had to indicate the degree of preference for each test in comparison with the other two tests on a 10-point bipolar

scale ranging from very weak preference to very strong preference.

Results

Since the analyses of item selection and preference ratings yielded the same results, the latter are not presented. The mean number of items chosen from each test in Conditions 1 ($n = 51$) and 2 ($n = 53$) is shown in Figure 1. Figure 1 reveals that in both conditions, more items are chosen from the high-diagnostic tests than from the low-diagnostic tests. The means for each test differ significantly between the experimental conditions, easy, $t(102) = 5.34$; moderate, $t(102) = 10.95$; difficult, $t(102) = 4.61$; $ps < .001$.[1] In contrast to the results from the experiments of Trope and Brickman and of Trope, in Condition 2 there was no preference for easy tasks over difficult ones.

To test the effect of the achievement motive on choice, subjects were divided at the median on the Net Hope (NH) scores in groups high (NH ≥ 5, $n = 52$) and low (NH < 5, $n = 52$) in resultant achievement motivation. There were no significant differences in preference between the motive groups. Thus, Trope's finding of a more pronounced preference of high-diagnostic over low-diagnostic tasks in the high than in the low motive group was not replicated.

To examine the influence of perceived ability, which is unrelated to achievement motivation ($r = -.12$), on item selection, groups high (Scale Values 1–4, $n = 56$) and low (Scale Values 5–9, $n = 48$) in perceived ability were separated, with 26 highs and 25 lows in Condition 1 and 30 highs and 23 lows in Condition 2. The mean number of items selected is shown in Figure 2. In Condition 1 there were no significant differences in the mean preferences of the two groups. In Condition 2, however, individuals low in perceived ability preferred easy test items to a significantly greater extent than individuals high in perceived ability, $t(51) = 3.79$, $p < .01$, whereas for the difficult test in this condition the reversed preference pattern is given, $t(51) = 2.27$, $p < .05$. The difference between the groups at the moderately difficult test failed to reach statistical significance, $t(51) = 1.92$. Within the groups high and low in perceived ability, there were significant differences between the experimental conditions at each level of test difficulty (t test, all ps at least $< .05$).

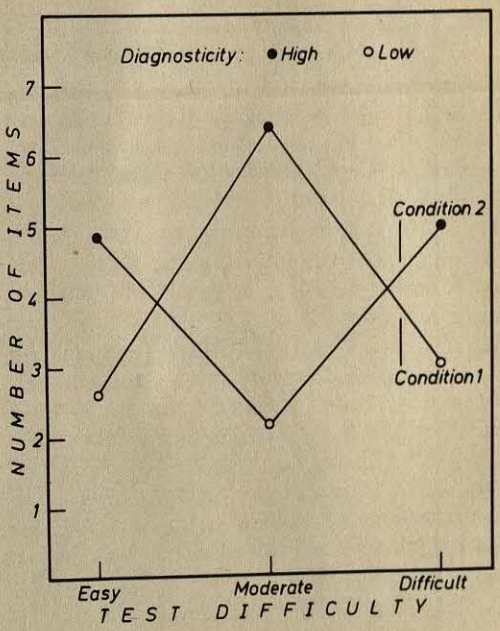


Figure 1. Mean number of items chosen from tests varying in difficulty in two experimental conditions.

Table 1
Percentages of Success Among High- and Low-Ability Students on Three Tests Reported in Two Experimental Conditions

Student ability	Test difficulty: Easy (.70)	Moderate (.50)	Difficult (.30)
	Condition 1		
High	73	75	33
Low	66	28	27
	Condition 2		
High	90	55	51
Low	50	48	10

Note. Criterion values are in parentheses.

[1] A comparison of the means for the tests within each experimental condition was not possible because the number of tasks selected from each test was not independent.

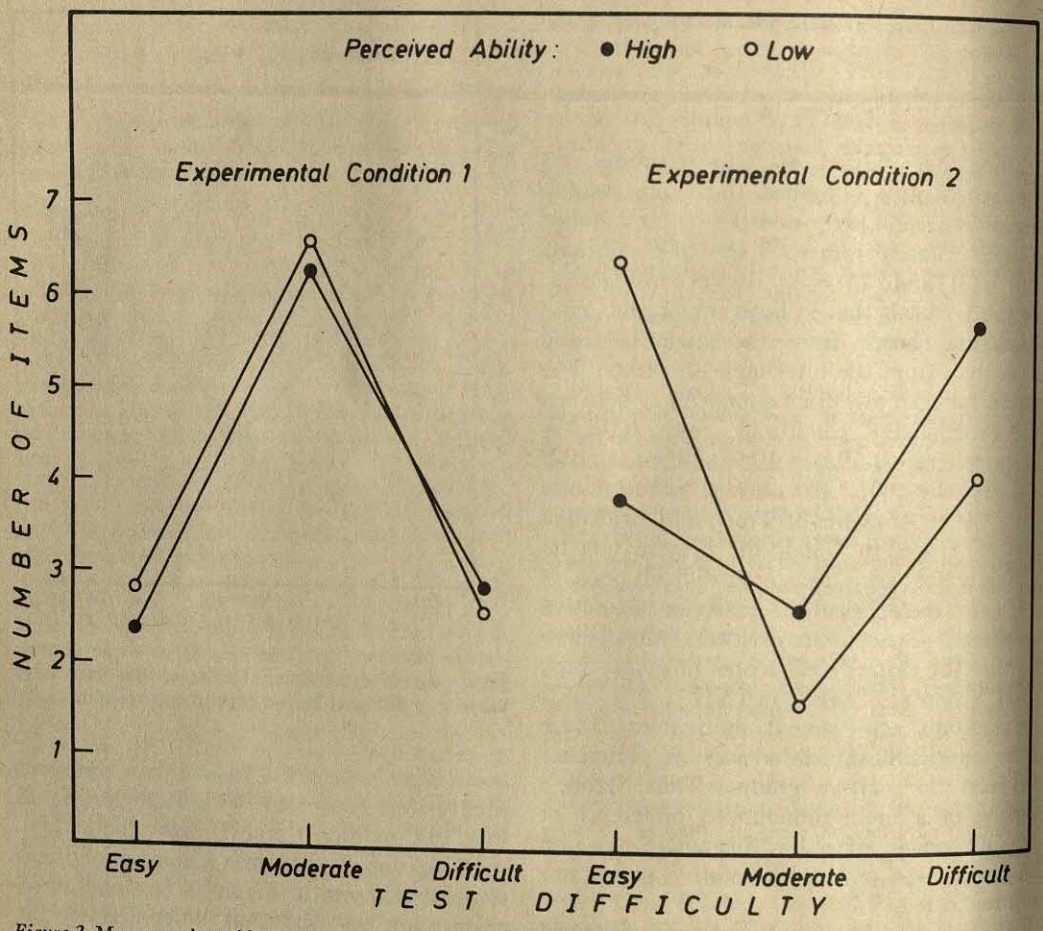


Figure 2. Mean number of items chosen from tests varying in difficulty in two experimental conditions by subjects high and low in perceived ability.

Discussion

The results of this experiment partially replicated the findings reported by Trope and Brickman (1975) and Trope (1975): It was found that high-diagnostic tasks were preferred to low-diagnostic ones, independent of their difficulty. This supports the view that choice behavior is guided by a tendency to attain self-relevant information.

But, in contrast to previous studies, a general preference for easy tasks over more difficult ones was not found. Furthermore, Trope's finding that high resultant achievers choose high-diagnostic tasks over low-diagnostic tasks to a significantly greater extent than low resultant achievers was not replicated. This lack of any differential choice behavior between the motive groups is consistent with the results of Meyer, Folkes, and Weiner (1976), which demonstrated that both high and low achievers selected tasks in a pattern that is best described as normally distributed around the difficulty level of .50.

The disparate results in the experiment reported here and in Trope's (1975) study can scarcely be attributed to the different instruments used for the assessment of the achievement motive in both studies. Meyer et al. (1976) and Trope used the Mehrabian (1969) scale, with Trope finding marked differences in choice between high and low achievers, which was not the case with Meyer et al. Furthermore, in a related study, Starke (Note 1) employed, in accordance with this experiment, the Grid technique. He found that high achievers select tasks that provide ability-relevant feedback, whereas low achievers

select tasks providing ability-irrelevant feedback. Thus, altoghether, it seems clear that individuals high in achievement need—independent of the disparate situational conditions in the experiments—strive for information about their capabilities. However, it remains unclear under what situational conditions individuals low in achievement need strive for or avoid such information.

Another important finding of this study is the difference in choice behavior between subjects high and low in perceived ability. In Condition 2 the high-ability group preferred difficult over easy tasks, whereas the low-ability group preferred easy over difficult tasks, with the easy and difficult tasks being high in diagnosticity. In Condition 1 the two groups correspondingly preferred tasks from the intermediate difficulty test, which is the only test high in diagnosticity in this condition.

This overall pattern of choice is neither completely consistent with Weiner et al.'s (1971) position that choice is a function of task difficulty, nor with Trope and Brickman's (1975) position that choice is a function of the expected diagnostic value of tasks. Thus, there must be at least one other determinant of choice behavior in addition to these two variables. It appears that an informational conception of choice behavior has to include the subjective probability of success at tasks as a determinant of choice. Meyer (1973, 1976) has contended that given tasks with which a person has had a history of involvement, information gain is maximal when P_s at a task equals .50 rather than when the objective difficulty of a task, as defined by the proportion of successful individuals, equals .50. Assume, for example, that Individuals A and B are confronted with a task intermediate in difficulty, both having some familiarity with the general type of tasks. A perceives own task-specific ability to be very high and estimates P_s to be 1.00, whereas B perceives own ability to be very low, estimating P_s at the task to be .00. It is clear that the expected success for A and the expected failure for B prospectively provide no new information concerning their abilities at this intermediate difficulty task, but only corroborate the established ability percepts. Individuals will expect most information about their ability from tasks in which own success and failure are maximally uncertain and whose P_s therefore is .50. Positive results at tasks with higher values of P_s and negative results at tasks with lower values of P_s have less self-informational value, inasmuch as these results either provide information about the difficulty of the task or confirm already established perceptions of own ability.

A number of studies correspondingly show that individuals high in perceived ability estimate the P_s of tasks at all difficulty levels as higher than individuals low in perceived ability (Meyer & Hallermann, 1977; Starke, Note 1; Buckert, Note 2; Schulz, Note 3). Thus, with increasing perceived ability, objectively more difficult tasks should be chosen to get self-relevant information at tasks for which P_s equals .50. Although P_s was not controlled here, it can be assumed, extrapolating from the studies just mentioned, that the perceived ability also influenced P_s: For the high-ability groups, the objectively difficult test (.30) should be nearer to $P_s = .50$ than the easy one (.70), whereas for the low-ability group, the easy test should be nearer to $P_s = .50$. It therefore can be assumed that from tasks equally high in diagnosticity (Condition 2), those are preferred whose subjective probability is closest to .50. Thus, although individuals high and low in perceived ability *phenotypically* select different tasks, *genotypically* they are both striving to attain ability-relevant information.

In sum, it seems false to us to conceive the informational properties of choice in general as a function of task difficulty (Weiner et al., 1971) or as a function of expected diagnosticity, which is usually dependent on task difficulty (Trope & Brickman, 1975). This may be true for situations in which the tasks to perform at are not known to the individual or in which the tasks are unspecified, as in the experiments by Trope and Brickman (1975), Trope (1975), and Meyer et al. (1976, Experiments I and II), or when the individual is completely unfamiliar with the tasks. Under these conditions, when the individual is unable to estimate own task-specific abilities, it might be the most efficient strategy to choose intermediate-difficulty (i.e., highly-diagnostic) tasks to obtain information about one's own ability. But when the tasks are known to the in-

dividual, as in the experiment reported here, and the individual already has an estimate of own ability, which is, however, subjectively not fully valid, then the perceived self-informational value of choice behavior at tasks that the individual perceives as diagnostic, with respect to own abilities, should be a function of P_s with maximal expected information at tasks intermediate in P_s.

Reference Notes

1. Starke, E. *Informationseinholung über die eigene Leistung in Abhängigkeit von Leistungsmotiv und wahrgenommener eigener Begabung.* Unpublished diploma thesis, Psychologisches Institut, Ruhr-Universität Bochum, West Germany. 1975.
2. Buckert, U. *Informationseinholung über die eigene Leistung in Abhängigkeit von Aufgabenschwierigkeit und wahrgenommener eigener Begabung.* Unpublished diploma thesis, Psychologisches Institut, Ruhr-Universität Bochum, West Germany, 1976.
3. Schulz, U. *Leistungen und Kausalattribuierungen bei Erfolgs- und Mißerfolgsmotivierten in Abhängigkeit von Leistungsresultat und Aufgabenschwierigkeit.* Unpublished diploma thesis, Psychologisches Institut, Ruhr-Universität Bochum, West Germany, 1975.

References

Amthauer, R. *Intelligenz-Struktur-Test 70 (IST 70).* Göttingen, West Germany: Hogrefe, 1973.

Atkinson, J. W. Motivational determinants of risk-taking behavior. *Psychological Review*, 1957, *64*, 359–372.

Atkinson, J. W. *An introduction to motivation.* Princeton, N.J.: Van Nostrand, 1964.

Heckhausen, H. *The anatomy of achievement motivation.* New York: Academic Press, 1967.

Heckhausen, H. Achievement motivation and its constructs: A cognitive model. *Motivation and Emotion*, 1977, *1*, 283–329.

Heider, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

Kelley, H. H. Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska symposium on motivation, 1967.* Lincoln: University of Nebraska Press, 1967.

Kelley, H. H. The processes of causal attribution. *American Psychologist*, 1973, *28*, 107–128.

Kukla, A. Foundations of an attributional theory of performance. *Psychological Review*, 1972, *79*, 454–470.

Kukla, A. An attributional theory of choice. In L. Berkowitz (Ed.), *Advances in experimental social psychology.* (Vol. 11). New York: Academic Press, 1978.

Mehrabian, A. Measures of achieving tendency. *Educational and Psychological Measurement*, 1969, *20*, 445–451.

Meyer, W.-U. *Leistungsmotiv und Ursachenerklärung von Erfolg und Misserfolg.* Stuttgart, West Germany: Klett, 1973.

Meyer, W.-U. Leistungsorientiertes Verhalten als Funktion von wahrgenommener eigener Begabung und wahrgenommener Aufgabenschwierigkeit. In H.-D. Schmalt & W.-U. Meyer (Eds.), *Leistungsmotivation und Verhalten.* Stuttgart, West Germany: Klett, 1976.

Meyer, W.-U., Folkes, V., & Weiner, B. The perceived informational value and affective consequences of choice behavior and intermediate difficulty task selection. *Journal of Research in Personality*, 1976, *10*, 410–423.

Meyer, W.-U., & Hallermann, B. Intended effort and informational value of task outcome. *Archiv für Psychologie*, 1977, *129*, 131–140.

Schmalt, H.-D. *Die Messung des Leistungsmotivs.* Göttingen, West Germany: Hogrefe, 1976.

Trope, Y. Seeking information about one's own ability as a determinant of choice among tasks. *Journal of Personality and Social Psychology*, 1975, *32*, 1004–1013.

Trope, Y., & Brickman, P. Difficulty and diagnosticity as determinants of choice among tasks. *Journal of Personality and Social Psychology*, 1975, *31*, 918–926.

Weiner, B., et al. *Perceiving the causes of success and failure.* Morristown, N.J.: General Learning Press, 1971.

Received July 14, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 7, 1179-1185

Role of Foreseen, Foreseeable, and Unforeseeable Behavioral Consequences in the Arousal of Cognitive Dissonance

George R. Goethals
Williams College

Joel Cooper and Anahita Naficy
Princeton University

Subjects delivered a counterattitudinal speech supporting an unwanted policy. Two groups were given information about the consequences of making the speech before agreeing to make it. One was explicitly informed of the possibility of an unwanted consequence occurring. The second group was given a general description of the consequences that was designed to make the specific unwanted consequence retrospectively foreseeable if made known later. In addition, there was a third group that was told nothing about the consequences of making the speech at the time of the decision to make it. After the speech was delivered, half of the subjects in each group were informed of a specific unwanted consequence of their act and half were given no further information. As predicted, self-justificatory attitude change was found only in the two conditions in which subjects were informed prior to making the speech of the specific unwanted consequence and in the condition in which subjects were given a general description of the consequences beforehand and specific information about the unwanted consequence after the speech. The results are discussed in terms of the relation of personal responsibility to cognitive dissonance arousal.

A series of recent experiments has made it relatively clear that unwanted behavioral consequences that become known to actors only after they have performed a behavior can lead to dissonance-produced self-justificatory attitude change (Cooper & Worchel, 1970; Cooper, Zanna, & Goethals, 1974; Goethals & Cooper, 1972). There is also considerable evidence leading to the conclusion that such consequences produce dissonance only when their occurrence is foreseen at the time of the decision to engage in the behavior (Cooper, 1971).

The latter point, that unwanted consequences must be foreseen if they are to produce dissonance, is consistent with arguments by Cooper (1971), and also Hoyt, Henley, and Collins (1972), regarding the role of personal responsibility in cognitive dissonance. Cooper proposed that counterattitudinal behavior will produce dissonance if the actor feels personally responsible for the consequences of the behavior. Responsibility is accepted if (a) the person had free choice in performing the behavior and (b) the person was able to foresee its unwanted consequences. However, Cooper's analysis does not tell us whether the unwanted consequences must be foreseen or only foreseeable in order for them to lead to the acceptance of responsibility. The distinction between foreseen and foreseeable consequences is an important one, both for psychological processes and legal decisions.

Consider the following example: A man enters a gambling establishment. According to his already existing cognitions, he would find it undesirable to donate his money to the

The authors wish to thank Marie-Claire Kamin and Mark P. Zanna for their help with this research.

Requests for reprints should be sent to George R. Goethals, Department of Psychology, Williams College, Williamstown, Massachusetts 01267.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3707-1179$00.75

Mafia. In the first situation, the person knows full well that the establishment is run by the Mafia and that any loss of money will enter into mafioso hands. We can say that losing money in this situation will result in a *foreseen* consequence. At the other extreme, a person with the same attitudes enters a church to play bingo. He later learns that the "church" was merely a building borrowed from a movie set and that it was, in fact, a front for the Mafia. He, too, contributed money to the Mafia, but this time the consequence was completely *unforeseeable*.

In the middle lies the situation in which the person does not have firm evidence as to who runs the gambling house. When he learns that it is the Mafia who pocketed his money, he may be tempted to say, "I didn't know." However, he may have had to walk past several bodyguards to get into the place; he may have have had to pass the statue of Al Capone; he may have had to whisper a password to a gun-holding stranger to enter the locked room. It can be argued that he certainly could have known that the Mafia received his money, had he given it some reflection. We can say that the consequence in this situation was *foreseeable*. The difference between the last two situations is that in the former, there was no reasonable way in which the person could foresee the consequence of his behavior. In the latter, a reasonable person could have been expected to be aware of the possible occurrence of the consequence.

A more formal set of definitions may be helpful. By foreseen consequences we mean those whose possible occurrence the actors are explicitly aware of at the time of decision. The occurrence of these consequences need only be possible, not definite, for them to be regarded as foreseen. The critical point is that they be in the actors' consciousness as possible occurrences. By foreseeable consequences we refer to those results of behavior that were not in the actors' awareness at the time of decision but that they feel they—or any reasonable person—could have anticipated in light of the information they were explicitly given. That is, these are consequences the actors were not aware of, but feel in retrospect that they reasonably could be expected to have been aware of them. Finally, unforeseeable consequences are those that the actors were not aware of and, furthermore, feel that there was no way that a reasonable person could have anticipated.

In the present article, we are testing the proposition that both foreseen and foreseeable consequences can lead to the arousal of cognitive dissonance. Unforeseeable consequences will not. Moreover, there will be a major difference between the foreseen and foreseeable situations. Past research (e.g., Goethals & Cooper, 1975) suggests that the explicit mention of an unwanted event prior to a decision to act will result in dissonance arousal regardless of whether subjects ever learn if the event took place. Although this is true of a foreseen consequence, we do not expect it to be true of a foreseeable consequence. When the unwanted consequence is not explicitly in an individual's awareness, it can only lead to dissonance arousal if its occurrence is made explicit after the behavior. Only then will subjects be able to think back to whether they could have been reasonably expected to anticipate the occurrence of the consequence.

In the current study, subjects agreed to participate in attitude-discrepant behavior. The degree of foreseeability of an unwanted behavioral consequence was separated into foreseen, foreseeable, and unforeseeable variations. In the foreseen variation, the unwanted consequence was explicitly mentioned to the subjects as one of the events that could occur as a result of their behavior. It was not a definite outcome, but the subject was made directly aware of its possible occurrence. In the foreseeable variation, the events to follow from the behavior were described in general terms at the time of decision. The specific unwanted consequence was not explicitly mentioned, but it was an instance of the general class of consequences that was described. In the unforeseeable variation, no mention was made of anything bearing any relation to the unwanted consequence that was to result from the attitude-discrepant behavior.

In addition, the information that subjects were given following their counterattitudinal behavior was varied. In the *not-informed* variation, there was no further mention of consequences after the subjects performed the behavior. In the *informed* conditions, subjects

were told explicitly that the unwanted consequence definitely was to occur.

It was predicted that self-justificatory attitude change would occur in the foreseen conditions, regardless of whether subsequent information about the occurrence of the consequence was ever given. On the other hand, subjects in the foreseeable variation were predicted to manifest attitude change only in the informed condition. It was expected that when these subjects received specific post-behavioral information of the occurrence of the consequence, they would feel that they could have been aware of it. Finally, subjects in the unforeseeable conditions were not expected to change their attitudes. Regardless of any subsequent information they might receive about the consequence of their behavior, it was expected that they would absolve themselves of the responsibility for those consequences, since they had no way of knowing about them at the time they agreed to their attitude-discrepant act.

Method

Overview

Under the guise of participating in a study on psycholinguistics, subjects delivered a counterattitudinal speech supporting an unwanted policy on campus. One group was explicitly informed of the possibility of an unwanted consequence following the attitude-discrepant act. A second group was given a general description of the consequences of the behavior, designed to make a specific unwanted consequence retrospectively foreseeable. A third group was not told anything about the consequences of the act.

After the counterattitudinal speech was delivered, half of the subjects in each group were informed of a specific unwanted consequence of their act, and half were given no further information. Finally, the subjects gave a rating of their opinion about the issue at hand, after which they were debriefed.

Subjects

Sixty Princeton undergraduates came in response to an advertisement for a psychology experiment. Subjects who asked for more information about the experiment were told that the study was on psycholinguistics.

Procedure

Each subject was run individually by two experimenters. One experimenter conducted the first half of the procedure and the other experimenter the second half. This technique was chosen to reduce the impression management concerns of the subjects. Both experimenters were female. One was a native of France and the other a native of Iran. Both spoke fluent English with quite noticeable French and Persian accents, respectively. This factor played a part in the selection of the cover story and lent it considerable credence.

When the subjects arrived the first experimenter greeted them and explained that the experiment, being conducted in conjunction with foreign language departments, concerned psycholinguistics and was investigating the different linguistic devices used in oral and written communication. It was explained that subjects would either write an essay or give a speech arguing for one side of a particular issue. The experimenter added that an issue was chosen that most students would be familiar with, namely a pending university decision to double the size of the freshman class and therefore over time, the number of undergraduates at the university. The experimenter explained that she knew how most undergraduates felt about the issue and that she had plenty of speeches and essays opposing the plan. Now, she said, she needed some favoring the proposed policy.

At this point the procedure varied for subjects in different conditions. Those in the unforeseeable conditions were simply told that participation in the experiment was voluntary and that now that they had heard what the experiment was about, they could decide whether they wanted to continue. The experimenter emphasized that the choice was theirs and that they could leave immediately and be paid. All subjects agreed to stay.

For the subjects in the foreseen and foreseeable variations, the experimenter explained that she should tell the subjects that several other groups on campus besides the psychology and language departments were interested in the speeches and essays collected in the study. She said that they were interested in things such as the linguistic devices used, the issues raised, and the tactics taken in composing the communications. In the foreseeable conditions the experimenter simply explained that the subject's speech would be sent to one of the interested groups on the basis of a random selection procedure. Then she proceeded to the discussion of free choice. In the foreseen conditions, before mentioning the random selection procedure the experimenter elaborated on the interested "groups" by deliberately reading from a note pad the names and interests of the groups. She stated:

> The groups are first, a group of graduate students in a linguistic strategies seminar (this group is interested in the linguistic devices used in the speeches and essays collected in this study) and second, the debating team. This group is interested in the fact that we have chosen topics that have two sides to them, and have asked for speeches and essays on both sides. Thus the interest of this group lies in the tacks taken by different students in composing the communication. And finally, the board of admissions of Princeton University is also interested in this study, for as I mentioned before, the University is considering an increase to 1,600 in the size of the freshman class.

When the subject had agreed to write the essay, the experimenter led her to another room where the subject was introduced to the second experimenter who was "helping out" with the project and who would record the speech and ask some final questions. The first experimenter said goodbye and left. The remainder of the procedure was conducted by the second experimenter. She gave the subject some paper to make notes on, told the subject to take 5 minutes to prepare a 1-minute speech, and left the room. She returned in 5 minutes and recorded the subject's speech. She began recording by speaking the subject's name and the topic and direction of the speech into the microphone. Then she handed the microphone to the subject who delivered the speech.

After the speeches had been recorded, the experimenter told the subjects that they seemed very persuasive. Then in the foreseen-informed condition, she reminded the subjects that a random process was to be used to determine which of the three other interested groups would get the speech. She asked the subject to look at a letter on the back of the piece of scratch paper, saw that it was the letter "B," and pointed to a chart showing that speeches recorded from pieces of scratch paper with the letter "B" would be sent to the board of admissions.

In the foreseeable-informed condition, the experimenter reminded the subjects that other groups were interested in the speeches and told the subjects what the three groups were, exactly as the subjects in the foreseen group had been told prior to the speech. Then she explained the random assignment procedure and informed the subject that the speech was to be sent to the board of admissions.

In the unforeseeable-informed condition, the experimenter explained to the subjects after the speeches were recorded that there were other interested groups, and so forth. Then the actual groups were identified, the random assignment procedure was explained, and it was revealed to subjects that their speech would be sent to the admissions board.

None of these postspeech events occurred for the subjects in the three not-informed conditions. All subjects then were asked to answer a questionnaire that contained a question asking the subjects to indicate, on an 11-point scale, the extent to which they agreed or disagreed with the idea of doubling the size of the freshman class. When the experimenter gave the questionnaire to the subjects, she explained she had to leave briefly to get another questionnaire. She told the subjects to complete the first questionnaire and put it in an envelope while she was gone. This questionnaire served as the main dependent measure of the study. When the subjects had placed the first questionnaire in the envelope, the experimenter returned with the second questionnaire that contained several questions about the experiment, including how free subjects felt to refuse to participate in the study and a check on the manipulations. The manipulation check asked subjects, on an 11-point scale, whether their speech was to be used for linguistic research only or other purposes as well. After the subjects had completed the second questionnaire, they were fully debriefed, sworn to secrecy, and paid $1.50 for their participation.

Table 1
Means on Question Regarding Uses of Speech

Condition	Foreseen	Foreseeable	Unforeseeable
Informed	10.7_c	10.2_c	10.3_c
Not informed	10.7_c	5.9_b	1.6_a

Note. Higher numbers indicate subjects believed that speech was to be used by "other groups"; lower numbers that speech was to be used for "linguistic research only." Means not sharing a common subscript differ at the 5% level by the Newman-Keuls test.

Results

Effectiveness of the Manipulations

Subjects were asked whether their speeches were to be used for linguistic research only or whether they would also be used by other groups. The means on this measure are presented in Table 1. An analysis of variance showed significant main effects for the informed factor, $F(1, 54) = 60.55$ $p < .001$; the foreseeability factor, $F(2, 54) = 24.36$, $p < .001$; and a significant interaction, $F(2, 54) = 20.34$, $p < .001$. Subjects in all three informed conditions were well aware that the speech would be used for other purposes. In addition, subjects in the foreseen–not-informed conditions knew that the speech would be used for other purposes. Subjects in the foreseeable–not-informed condition were, surprisingly, less aware that the speech would be used for other purposes than subjects in the aforementioned conditions, but more aware than subjects in the unforeseeable–not-informed condition who believed that the speech would be used for research purposes only. The foreseeable–not-informed condition subjects' lesser awareness of other uses of the speech makes sense in terms of their being less fully informed about the other groups interested in the speeches. Overall, the results suggest that the manipulations had their intended impact. In addition, all subjects indicated that they felt free to decline participating in the speech-making task.

Attitudes Toward Increases in Class Size

Subjects were asked whether they thought that the freshman class size should be increased

to 1,600 students. They responded on a scale labeled *strongly agree* (scored 1) and *strongly disagree* (scored 11) at the end points. The results are presented in Table 2. As expected, subjects in the unforeseeable–not-informed group were strongly opposed to an increase in the size of the freshman class. This group, which made a speech but never believed any unwanted consequence would occur, was expected to reflect the prevalent attitude on campus ($M = 10.1$)

An analysis of variance performed on the data of Table 2 resulted in main effects for both the informed and foreseeability factors. Informed subjects evidenced more favorable attitudes toward the proposal than not-informed subjects, $F(1, 54) = 5.85$, $p < .02$. In addition, as the degree of prior awareness of the consequences increased from unforeseeable to foreseen, subjects indicated greater agreement with the counterattitudinal position, $F(2, 54) = 12.26$, $p < .001$. Most important, our hypotheses led to the prediction of an interaction between the two variables. This was supported by the analysis of variance, which found the interaction to be reliable, $F(2, 54) = 3.50$, $p < .05$. A Newman-Keuls analysis summarized in Table 2 adds further support. When the consequences were foreseen prior to the decision to act, subsequent information made no difference in the magnitude of attitude change. Both the informed and not-informed variations evidenced substantial changes in belief toward the proposal. Similarly, when the consequences of the behavior were completely unforeseeable prior to the decision, the level of subsequent information again made no reliable difference in final opinions. However, when the consequences were foreseeable—those that one might reasonably have been expected to anticipate but were not explicitly made known prior to the act—then the level of subsequent information made an important difference. If not informed of the unwanted consequence, foreseeable subjects reported anti-class increase attitudes ($M = 10.2$), but if they were subsequently informed, they demonstrated considerable moderation ($M = 7.4$). The difference between the two foreseeable variations was significant ($p < .05$).

Discussion

The results are supportive of the original predictions. First, when subjects perform an act under conditions of high choice that they realize has the potential to produce unwanted consequences, they show the effects of experiencing dissonance. This is true whether or not the subjects are informed that the unwanted consequence did actually occur. That is, even if subjects are not told the outcome of their behavior, they still show manifestations of dissonance. This latter finding is entirely consistent with previous research (Goethals & Cooper, 1975) suggesting that the mere possibility that one's behavior will produce unwanted consequences is sufficient to produce dissonance if subjects anticipate receiving no further information regarding the final outcome.

Second, if one's behavior produces unwanted consequences that are totally unforeseeable, such consequences produce no dissonance. If a subject performs what appears to be innocuous behavior that later turns out to have unwanted consequences that could not possibly have been foreseen, the subject will not accept responsibility for those consequences and will not manifest dissonance.

Finally, and most importantly, the study has demonstrated that consequences made known to the subject after he or she has performed a behavior can and do produce dissonance if the subject perceives, retrospectively, that they were foreseeable; that is, that he or she could have foreseen that such an unwanted consequence could occur.

One important aspect of these results is that they reaffirm the idea that personal responsibility for consequences is a necessary precondition for dissonance arousal. Furthermore, the study has helped specify under what conditions subjects will accept responsibility. They will accept responsibility if they feel the consequences were foreseeable, in the sense that they *could have been* foreseen by one who was being careful and thoughtful about his or her behavior. Because the consequences could have been foreseen, subjects will feel that they should have foreseen them and are responsible

Table 2
Mean Attitudes Toward Increasing the Size of the Freshman Class

Condition	Foreseen	Foreseeable	Unforeseeable
Informed	7.1_a	7.4_a	9.6_b
Not informed	7.4_a	10.2_b	10.1_b

Note. Higher numbers indicate less favorable attitudes toward increasing class size. Means not sharing a common subscript differ at the 5% level by the Newman-Keuls test.

for the unwanted consequences even though they did not, in fact, foresee them.

It seems as if the subjects in our experiment accepted responsibility for their actions in much the same way that the law assigns responsibility to people whose behavior leads to undesirable consequences. Drivers who seriously harm pedestrians in auto accidents are held responsible by the law consistent with the principles of choice and foreseeability. Assuming that the drivers were driving as they were of their own volition, choice is clearly present. But their lawyers might argue that the outcome was not foreseen, or their clients clearly would have driven differently. However, the consequence was foreseeable. A reasonable man (a well-accepted legal fiction) would have foreseen that driving in this way could lead to a serious accident and injury. That is, drivers can be held negligent (i.e., responsible) even if they did not foresee what would occur, if they should have anticipated what might occur.

The demonstration that dissonance can be produced by behavior that was performed where no unwanted consequences were explicitly foreseen is important in terms of the external validity of dissonance research. Most studies have shown that if subjects are coerced or induced to perform behavior that they know has unwanted consequences, they will experience dissonance. Yet one wonders how often people perform such behaviors in real life. Of course, the inducements and pressures that dissonance experimenters use to elicit compliance are not absent from the real world. Still, it seems that people more often perform behavior not realizing that it will produce negative consequences but subsequently find that it did. They then may feel that they should have known. For example, people may not buy a house if they realize that the roof is going to leak. However, they may find that it does leak and that they neglected to check it before they bought the house. Here the consequence was not foreseen but probably would be regarded as foreseeable, and the people will feel responsible for the situation's having occurred. They should reduce dissonance by believing more strongly in how nice their house is. These kinds of events—people behaving in ways that produce unanticipated but foreseeable events—are probably more common in arousing dissonance than those in which people knowingly behave in ways that bring about unwanted consequences.

Finally, these data are important in relation to the overall understanding of the idea of responsibility and its relation to dissonance arousal. Wicklund and Brehm's (1976) review of the development of the responsibility concept suggests that although choice and foreseeability are necessary for responsibility and thus dissonance reduction, "unforeseen consequences can arouse dissonance under special conditions" (p. 71). The present study suggests one special condition for which the unforeseen consequence is retrospectively foreseeable. This special condition seems to apply to all of the studies reviewed by Wicklund and Brehm that appear to contradict the proposition that foreseeability is a necessary prerequisite for accepting responsibility for behavioral outcomes.

For example, in studies conducted by Pallak, Sogin, and Van Zante (1974) and by Sogin and Pallak (Note 1), subjects participated in a task that involved selecting random numbers. They were told afterwards that their numbers could not be used because of the way they had performed the task. Although they did not know beforehand that their numbers might not be used on the basis of the experimenter's criterion, it was probably forseeable that their performances could be found wanting in some way. Subjects were simply told that their numbers "should be usable." This seems to leave open the possibility that on some basis or other, admittedly unknown, they also might not be usable.

Wicklund and Brehm (1976) discuss a study by Aronson, Chase, Helmreich, and Ruhnke (1974) in which subjects wrote counter-attitudinal essays and were told before or after writing them that their essays would be shown to a persuasible or nonpersuasible audience. There were two other conditions (lie conditions) in which subjects were promised that the essays would not be shown to other people. These subjects also discovered after writing their essays that they would be shown to a persuasible or nonpersuasible audience. Aronson et al. found attitude change in the persuasible audience-after condition, in which they argued that the unwanted consequence was unforeseeable. However, in all conditions of the experiment, subjects were told that their essays would be used "for an attitude change study," and so an element of foreseeability of the unwanted consequence may well have been present. Aronson et al. pointed out that there was greater attitude change in the persuasible audience-lie condition than expected. Perhaps this finding can be explained in terms of this degree of foreseeability. Incidentally, Wicklund and Brehm also noted that this study may have involved foreknowledge on the part of the subjects.

We would like to be able to argue that not only the present study, but also the research reviewed by Wicklund and Brehm (1976), shows conclusively that self-justificatory attitude change does not occur when actors produce unwanted consequences that are truly unforeseeable. We think that a plausible case can be made that retrospective foreseeability of the kind created in the present experiment did exist in all of the studies that were designed to show dissonance effects without foreseeability. However, it is difficult to make this point definitively. Unlike the present experiment, previous studies have been concerned only with a foreseen versus unforeseen distinction rather than a retrospectively foreseeable versus unforeseeable distinction within an unforeseen variation. In examining the procedures of these studies, it may be reasonable to argue that they resemble more clearly our foreseeable variation than our unforeseeable variation, but this argument cannot be drawn conclusively. We believe that dissonance effects have not been demonstrated in conditions in which the unwanted consequences are truly unforeseeable, but further research would be required to support our interpretation of earlier studies.

Reference Note

1. Sogin, S. R., & Pallak, M. S. *Responsibility, bad decisions, and attitude change: Volition, foreseeability, and locus of causality for negative consequences.* Unpublished manuscript, University of Iowa, 1974.

References

Aronson, E., Chase, T., Helmreich, R., & Ruhnke, R. Feeling stupid and feeling guilty—Two aspects of the self-concept which mediate dissonance arousal in a communication situation. *International Journal of Communication Research*, 1974, *3*, 340–352.

Cooper, J. Personal responsibility and dissonance: The role of foreseen consequences. *Journal of Personality and Social Psychology*, 1971, *18*, 354–363.

Cooper, J., & Worchel, S. Role of undesired consequences in arousing cognitive dissonance. *Journal of Personality and Social Psychology*, 1970, *16*, 199–206.

Cooper, J., Zanna, M. P., & Goethals, G. R. Mistreatment of an esteemed other as a consequence affecting dissonance reduction. *Journal of Experimental Social Psychology*, 1974, *10*, 224–233.

Goethals, G., & Cooper, J. Role of intention and post-behavioral consequence in the arousal of cognitive dissonance. *Journal of Personality and Social Psychology*, 1972, *23*, 293–301.

Goethals, G. R., & Cooper, J. When dissonance is reduced: The timing of self-justificatory attitude change. *Journal of Personality and Social Psychology*, 1975, *32*, 361–367.

Hoyt, M. F., Henley, M. D., & Collins, B. E. Studies in forced compliance: Confluence of choice and consequence on attitude change. *Journal of Personality and Social Psychology*, 1972, *23*, 205–210.

Pallak, M. S., Sogin, S. R., & Van Zante, A. Bad decisions: Effect of volition, locus of causality, and negative consequences on attitude change. *Journal of Personality and Social Psychology*, 1974, *30*, 217–227.

Wicklund, R. A., & Brehm, J. W. *Perspectives on cognitive dissonance.* New York: Wiley, 1976.

Received July 24, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 7, 1186–1195

Focus of Attention, Chronic Expectancy, and Responses to a Feared Stimulus

Charles S. Carver and Paul H. Blaney
University of Miami

Michael F. Scheier
Carnegie-Mellon University

An attentional model of fear-based behavior is proposed and a study that tested the model is reported. It was predicted that among subjects with moderate fear of snakes, heightened self-attention during an approach attempt would cause increased awareness of existing anxiety, followed by one of two courses of events: Subjects who believed that they could do the behavior in spite of their fear were expected to redirect their attention to the behavior–goal comparison and exhibit no behavioral deficit. Subjects who doubted their ability to do the behavior were expected to divert their attention from the behavior–goal comparison and to withdraw behaviorally from the approach attempt. The results of the study support this reasoning. Discussion centers on relationships between the proposed model and previous theory.

Recent approaches to understanding anxiety and anxiety-related behavior have emphasized the role of cognitive processes in such experiences (see, e.g., Sarason, 1972a, 1972b; Wine, 1971). Some of this theorizing, for example, Wine's summary and integration of the literature of test anxiety, has treated attentional focus as an important variable in the fear response. Almost universally, however, direction of attention has been considered only as a response—a reaction to fear-inducing situations. Very little consideration has been given to the *consequences* of shifts in focus of attention.

An exception to this general rule comes from research on fear-based withdrawal that was conducted by Carver and Blaney (1977a, 1977b), investigating the effects of bogus arousal feedback on fearful subjects' approach behavior. In two of those studies (Carver & Blaney, 1977a, Experiment 3; 1977b) subjects were tested who had moderate fear of nonpoisonous snakes, but who varied in their self-rated expectancy of being able to approach, pick up, and hold a snake. Confident subjects were defined as those who believed that they would be able to do those behaviors, even though doing so would cause them some personal discomfort; doubtful subjects were those who reported being not at all sure that they could do the behaviors. Subjects later attempted the behavioral sequence in question while simultaneously receiving bogus heartbeat feedback. For some subjects, the feedback indicated autonomic quiescence, for others it indicated gradually increasing arousal (the heartbeat acceleration was coordinated with the subject's approach behavior). Doubtful subjects (i.e., those with negative expectancies) responded to the arousal feedback by attending less to the comparison between their behavior and their goal (according to self-reports; Carver & Blaney, 1977b) than did subjects in the constant heartbeat control group and by withdrawing earlier in the approach sequence. However, confident subjects (i.e., those with positive expectancies) responded to the arousal

Portions of the material contained in this article were presented at a symposium on "Attention and Social Behavior" at the August 1978 meeting of the American Psychological Association.

The authors are grateful to Daniel Castellanos for his able service as experimenter.

Requests for reprints should be sent to Charles S. Carver, Department of Psychology, P.O. Box 248185, University of Miami, Coral Gables, Florida 33124.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3707-1186$00.75

feedback by increasing the attention they paid to the comparison between their behavior and their goal and by bringing the one into closer conformity with the other.

Carver and Blaney (1977b) suggested the following interpretation for their findings, based on Duval and Wicklund's (1972) self-awareness theory: An accelerating heartbeat might, as a sign of fear, lead to greater self-focus than a constant heartbeat. This would be consistent with theorizing that holds self-focus to be a response to fear-inducing situations (e.g., Wine, 1971). However, Carver and Blaney further suggested that subsequent behavioral and focus-of-attention differences between confident and doubtful subjects might have resulted from two different modes of responding *to* that self-focus. That is, it has been demonstrated repeatedly in other behavioral contexts that one consequence of self-directed attention is a heightened tendency to conform to salient behavioral standards (e.g., Carver, 1974, 1975; Froming, in press; Gibbons, 1978; Scheier, Fenigstein, & Buss, 1974; Wicklund & Duval, 1971). This is an effect of self-attention that appears to occur whenever such conformity is possible. This phenomenon seemed to Carver and Blaney to be implicit in their findings among confident subjects who had been presented with arousal feedback. That is, these subjects reported greater attention to the behavior–goal comparison than did those presented with constant feedback, and they failed to display any behavioral deficit. These subjects thus seemed to be attempting, successfully, to match their behavior to the salient standard (i.e., picking up and holding the snake).

On the other hand, at least one previous study (Duval, Wicklund, & Fine, 1972) had shown that the realization that one cannot conform to the behavioral standard caused self-focus to enhance behavioral withdrawal. That previous demonstration seemed to Carver and Blaney (1977b) to be consistent with their findings among doubtful subjects. That is, arousal feedback caused doubtful subjects to withdraw earlier from the approach attempt and to report reduced focus on the behavior–goal comparison.

More recent data suggest, however, that Carver and Blaney's (1977b) interpretation of their results was not entirely accurate. Specifically, Fenigstein and Carver (1978) have shown that the presence of a sound identified to subjects as their heartbeats causes increased self-attention, as reflected by performance on a color–word test (cf. Geller & Shaver, 1976) and by increased self-attribution of hypothetical outcomes (cf. Buss & Scheier, 1976; Duval & Wicklund, 1973). More importantly, Fenigstein and Carver's results failed to indicate that an accelerating heartbeat sound leads to greater self-focus than does a constant heartbeat sound. Carver and Blaney's (1977b) interpretation of their findings as reflecting responses to self-focus is thus called into question.

A Sequential Model of Fear-Based Behavior

There is, however, a way to reconcile certain aspects of Carver and Blaney's reasoning with the Fenigstein and Carver (1978) finding. A model of approach and withdrawal that does so, which represents a refinement of Carver and Blaney's (1977b) theory, has been presented elsewhere (Carver, in press) as part of a more comprehensive analysis of behavior regulation. The model (see Figure 1) is described by the following assumptions:

1. When a behavioral standard is salient (in this case, to approach and handle the feared stimulus), self-attention leads to increased attempts to conform behaviorally to the standard. If fear arousal cues never become salient during the approach, this process continues until the task is completed. This would seem to reflect what occurred among Carver and Blaney's (1977a, 1977b) subjects who were presented with a constant heartbeat sound. Those subjects apparently accepted the nonarousal information contained in the feedback as veridical. Thus, although they may have been self-attentive (as is implied by Fenigstein & Carver's data, 1978), their behavior reflected only conformity to the standard.

2. If arousal becomes strong enough to be salient, the approach attempt is momentarily interrupted. This interruption leads to an assessment of the likelihood of being able to continue the behavior. The combination of self-focus plus salient arousal information

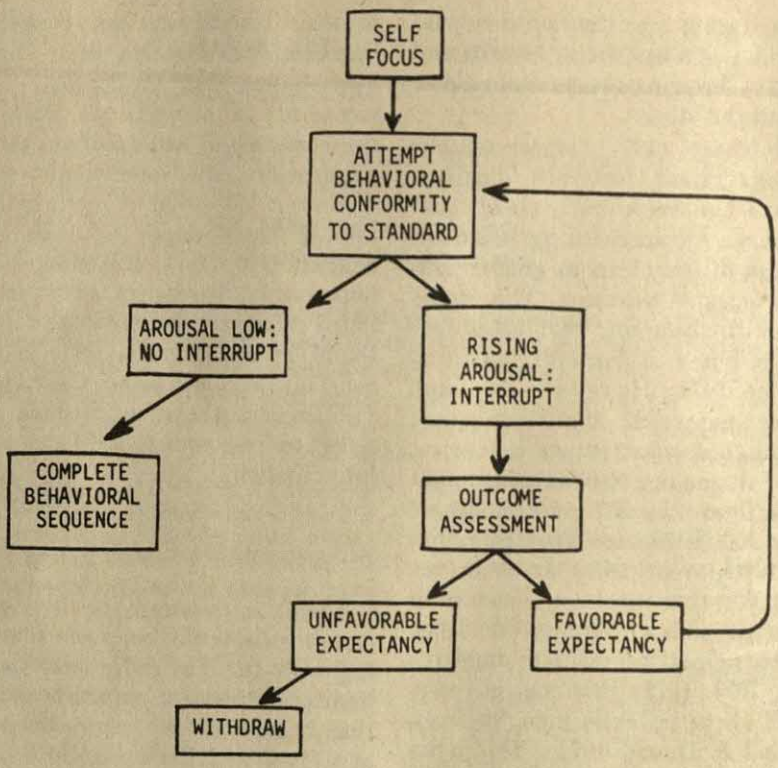


Figure 1. Postulated sequence following self-attention in a fear-provoking behavioral context.

would seem to correspond to the subjective experiences of Carver and Blaney's subjects who had been presented with an accelerating heartbeat.

3. If the outcome of the assessment is a positive expectancy (e.g., if the person judges that he or she can cope with the fear), the result is a return to the matching-to-standard attempt, with perhaps even increased concentration on that attempt (as reflected by self-report focus-of-attention data from Carver and Blaney's, 1977b, subjects). If the outcome is a negative expectancy (if the person thinks that his or her limits have been reached), the consequence is behavioral withdrawal. The expectancy judgment thus is a kind of psychological watershed. All behavioral responses following this judgment seem ultimately to fall into one of those two categories, that is, renewed efforts or withdrawal.

To summarize, in this model the following variables are considered relevant to the understanding of approach behavior in a fear-eliciting situation: the degree of self-focus experienced during the task, the level of anxiety experienced during the task, and one's degree of confidence in probable success at the task. Self-focus initially increases the salience of task completion as one's goal; but if anxiety arises, self-focus also increases one's awareness of the anxiety. The greater the subjective experience of anxiety, the greater the likelihood that one will pause to evaluate one's chances of successful completion. If the resulting expectancy judgment is favorable, the person returns to the approach attempt; if the judgment is unfavorable, the person withdraws. Note that any renewed approach can be interrupted again, later in the sequence, if perceived anxiety increases again.

Present Research

The study reported below was conducted as a test of this model. In this study, dispositional anxiety was held constant and both self-focus and dispositional expectancies were

varied.[1] Specific predictions were derived as follows: Subjects with any substantial degree of chronic fearfulness with regard to some stimulus (e.g. a snake) presumably experience some veridical fear-related autonomic arousal at some point during an attempt to approach that stimulus. There is considerable evidence that in the absence of bogus feedback, experimentally heightened self-focus causes an increase in awareness of one's internal affective states (Scheier & Carver, 1977). Thus we predicted that in a fear-provoking context, heightened self-focus should lead to enhanced perceptions of rising fear arousal among all fearful subjects, regardless of their chronic expectancies. The subjective perception of arousal, above some quasi-threshold, is assumed to cue the self-assessment process. Thus we predicted that the enhanced awareness of anxiety experienced by the self-attentive subjects would lead them to undertake self-assessment earlier in the approach sequence than the less self-attentive subjects, for whom the arousal would be less salient. Subsequent behavioral responses to that early self-assessment, however, should vary with subjects' expectancies. That is, doubtful subjects (those with negative expectancies) should respond by withdrawing earlier in the approach sequence if self-focus is high than if it is low. Confident subjects (those with positive expectancies) should exhibit no such behavioral deficit; if anything, confident subjects should respond with renewed efforts. Thus we predicted an interactive influence of self-attention on subjects' approach/withdrawal tendencies. A similar interactive effect was expected on subjects' self-reports of what they had attended to during the approach attempt.

Method

Subjects

Potential subjects, undergraduates at the University of Miami, completed a two-item pretest at the beginning of the semester. The items required respondents (a) to estimate the amount of anxiety aroused in them by nonpoisonous snakes (very little, slight, moderate, marked, or strong) and (b) to indicate what they thought their response would be to being asked to pick up a nonpoisonous snake ("I'd do it without any discomfort"; "I'd do it, but I'd feel a little queasy about it"; "I'm not at all sure that I'd do it"; or "I'd be too fearful even to try").

Our predictions depended on an interaction between self-attention and expectancies, with chronic fearfulness held constant. To select subjects who differed appropriately with regard to expectancies, we chose persons who reported moderate fear of snakes. Confident subjects were defined as those who also indicated that they could pick up a snake even though feeling queasy about it. Doubtful subjects were those who indicated being not at all sure that they would do it. These were the same criteria for subject selection as were used by Carver and Blaney (1977b).

Subjects completed the experimental procedures described below, in individual sessions several weeks after having completed the questionnaire. Subjects were randomly assigned to either the mirror present or the mirror absent condition by a second experimenter, who had no contact with subjects. The primary experimenter had no knowledge of subjects' pretest scores or of the hypotheses being tested until after the sessions had been completed. Assignment of subjects to conditions led to the following distribution: Each group of doubtful subjects was comprised of 10 males and 6 females, the confident–mirror present group had 10 males and 5 females, and the confident–mirror absent group had 9 males and 4 females.

Procedure

When the subject arrived at the experimental site, he or she was given typed instructions explaining that the study concerned the physical aspects of anxiety. The subject was to attempt to approach, pick up, and hold a live nonpoisonous snake while his or her heartbeat ostensibly was monitored by the experimenter via a microphone. (This explanation and the following procedures were adapted from those of Carver & Blaney, 1977a, 1977b; the microphone was used to make the procedures as similar as possible to those used previously, but the present study did not use any bogus feedback.)[2] Each subject was encouraged to go as far as he or she could (thus establishing approach as the behavioral standard), but the subject was also told that he or she could stop at any point simply by indicating a desire to go no farther.

After reading the instructions, the subject strapped a microphone to his or her chest at a position where he

---

[1] What is critical in the model is the person's *situational* expectancy. However, on the basis of a substantial amount of data from previous research (Carver & Blaney, 1977a, 1977b), we assumed that subjects' dispositional expectancies would provide a good basis for inferring their situational expectancies. The results of the present study lend additional credence to that assumption.

[2] It might be argued that strapping on a microphone would lead to heightened self-attention. However, data from Fenigstein and Carver (1978) indicate that such is not the case.

or she had located a strong heartbeat sound (using a stethoscope provided for this purpose). The subject then was given gloves and was escorted to the test corridor. At the far end of the corridor (which was slightly less than 1 m wide and approximately 4 m long) was a glass cage with a sliding top, standing on a pedestal about .5 m off the floor. Inside this cage was a boa constrictor slightly over 1 m in length. The experimenter reminded the subject that his or her task was to open the cover, reach in and pick up the snake, and hold it until the experimenter said to return it to the cage. After cautioning the subject not to begin until a verbal signal had been given, the experimenter returned to the observation room, ostensibly to turn on the heartbeat recorder.

Self-attention manipulation. Self-attention was varied in this experiment by the presence or absence of a mirror. Evidence that this stimulus does increase self-focus has been discussed at length elsewhere (Carver & Scheier, 1978). For subjects in the mirror absent condition, the corridor was generally featureless. For subjects in the mirror present condition, however, a .5 m × .75 m wall mirror had been hung behind the cage at approximately chest level. The angle of the mirror gave the subject a view of his or her face and shoulders; the snake, however, was not visible in the mirror at any point except the final stage of the approach task (discussed later). In the mirror present condition, before leaving the subject the experimenter pointed out the mirror's presence, indicated that it was part of the experiment, and said that its purpose would be explained later. No subject ever expressed concern or suspicion about the mirror's presence, either at that point or during postexperimental debriefing.

It will be noted that this method of introducing a self-awareness manipulation differs from usual procedures. Typically, the mirror is treated as belonging to a different experiment (e.g., Carver, 1974, 1975) or as being necessary for a subsequent task (e.g., Wicklund & Duval, 1971). We decided to point the mirror out to subjects in the present context, however, for the following reason: A live snake is an attention-attracting stimulus even for people who are not afraid of snakes. Inasmuch as all of the present subjects had reported being moderately fearful of snakes, it seemed likely that they would be very attentive to the snake's presence while in the test corridor. This would tend to make the mirror less salient as a stimulus, thus minimizing its impact on subjects. (See Scheier et al., 1974, for a further discussion of the role of stimulus salience in effectiveness of self-attention manipulations.) For this reason, it was decided to draw subjects' attention explicitly to the mirror as a part of their behavioral context.[3]

Approach task. After the experimenter returned to the observation room, he called to the subject to begin approaching. Nine levels of approach had been predefined: (1) the subject's appearing in the experimenter's view, (2) coming to within 6 inches (15 cm) of the cage, (3) touching the cage, (4) putting a hand inside the cage, (5) touching the snake, (6) grasping the snake, (7) lifting the snake a few inches, (8) lifting it from the cage, and (9) holding it for 15 sec. Each subject's approach score was defined as the highest level that he or she attained.

After the subject had completed the approach task or had indicated an unwillingness to continue, the experimenter told the subject to return to the other end of the corridor. The subject then was escorted to the original room, where he or she was given a postexperimental questionnaire, then debriefed and dismissed.

Postexperimental Measures

The postsession questionnaire contained several items intended to gain indirect evidence on hypothesized mediating processes. (With one exception, noted later, all of these questions were answered on 7-point scales anchored with "none at all" and "very much.") The first item, bearing on the subject's awareness of affect, was "How much anxiety did you experience in the presence of this snake, as indicated by your bodily arousal?"

The presence of noticeable affect was expected to lead to a self-assessment process. This process should be reflected by a momentary sense of concern as to whether one could do the behavior in question. Thus a second question posed to subjects was "When you were approaching the snake, how much did you notice a momentary sense of inadequacy and fearfulness?"

The following three items were included to try to determine what subjects' attention had been focused on during the approach attempt: "When you were approaching the snake, how much attention were you paying to your 'chronic' level of fearfulness toward snakes?"; "When you were approaching the snake, how much attention were you devoting to assessing whether or not your bodily arousal was increasing?"; and "When you were approaching the snake, how much attention were you paying to your goal—that is, picking up and holding the snake—and how well you were doing compared to that goal?"

Finally, subjects were asked to rate their general level of fearfulness toward nonpoisonous snakes. This

---

[3] It might be argued that this experimental setting has the particular demand characteristic of reminding subjects that they had previously made a prediction about whether or not they could do the approach task. If this were the case, however, one would expect a behavioral difference between confident and doubtful subjects in both mirror present and mirror absent conditions. As will be detailed later, this did not occur. It might be argued, on the other hand, that heightened self-focus makes subjects more sensitive to preexisting demand characteristics. This is evidence from three separate studies, however (Gibbons, Carver, Scheier, & Hormuth, in press; Scheier, Carver, & Gibbons, in press), that self-focus can exert precisely the opposite influence. The nature of demand characteristics in this type of research is an interesting theoretical problem, but full treatment of the problem is beyond the scope of this article.

tem was identical to the first pretest item and had the same five response options as had the pretest item.

It will be noted that all of these self-report data were collected after the behavioral task had been completed. Thus results on these items should be interpreted with some caution.

Results

Awareness of Affect

An analysis of variance was conducted on self-reported degree of anxiety felt in the presence of the snake (see Table 1) to determine whether the manipulation of self-focus had increased subjects' awareness of their bodily anxiety responses.[4] This analysis, as predicted, yielded only a main effect for self-awareness condition, $F(1, 52) = 9.62$, $p < .004$, with greater anxiety being reported by subjects in the mirror present than in the mirror absent condition. Group comparisons using the error term from the overall analysis of variance (Winer, 1962, p. 65ff.) indicated that this effect was most reliable among doubtful subjects, $t(52) = 2.92$, $p < .01$, the comparable comparison not attaining significance among the confident subjects ($t = 1.41$, *ns*).

Table 1
Self-Reported Anxiety Experienced in the Presence of the Snake, Self-Reported Momentary Sense of Inadequacy and Fearfulness During the Approach Task, and Actual Level of Approach Toward the Snake

Group	Condition: Mirror present	Condition: Mirror absent
Self-reported anxiety[a]		
Confident	4.33	3.46
Doubtful	5.25	3.56
Self-reported sense of inadequacy[a]		
Confident	4.20	3.85
Doubtful	5.38	3.63
Approach behavior[b]		
Confident	8.47	8.23
Doubtful	6.38	8.31

[a] Larger numbers indicate greater anxiety and felt inadequacy. [b] Larger numbers indicate greater approach.

It was expected that awareness of anxiety and consequent self-assessment would be reflected in self-reports of having experienced a momentary sense of "inadequacy and fearfulness." An analysis of variance of those data (see Table 1) revealed that again there was a significant main effect for self-awareness condition in the expected direction, $F(1, 52) = 6.05$, $p < .02$. As with the anxiety self-reports, this effect was most reliable among doubtful subjects, $t(52) = 2.85$, $p < .01$; among confident subjects, $t < 1$. No other component of this analysis was statistically significant.

Approach Behavior

It was predicted that heightened self-focus would interact with subjects' expectancies of being able to do the approach task, such that heightened self-attention would cause earlier withdrawal among doubtful subjects but not among confident subjects. An analysis of variance of approach scores yielded a marginally significant interaction, $F(1, 52) = 3.67$, $p < .06$, of the expected form (see Table 1). Subsequent comparisons indicated that doubtful subjects withdrew reliably earlier in the approach sequence when the mirror was present than when it was absent, $t(52) = 2.48$, $p < .03$. The approach scores of the confident subjects were unaffected by the presence of the mirror, however.

There was a considerable range of within-cell variances on the approach measure. Although this diversity was not sufficient to render an analysis of variance entirely inappropriate by F_{max}, we judged it desirable to gain additional information on approach scores through a nonparametric test. Therefore, a comparison between approach scores of the two doubtful groups was conducted by means of a Kruskal–Wallis "analysis of variance by ranks" (see, e.g., Hays, 1963, p. 637ff.). When adjusted for ties in rank, this test also indicated that the scores of these groups did differ as predicted, $H' = 10.04$, $p < .01$.

[4] All analyses reported here included sex of subject as a variable. Because there was only one significant effect involving sex in the results of the study, however, data presented in all tables have been combined across gender.

Table 2
Self-Reports of Attention Paid to Chronic Level of Fearfulness, Self-Assessment of Bodily Arousal, and Behavior–Goal Comparison

Group	Condition: Mirror present	Condition: Mirror absent
Chronic level of fearfulness		
Confident	4.67	4.23
Doubtful	4.81	3.56
Assessing degree of bodily arousal		
Confident	3.60	4.69
Doubtful	4.31	3.13
Behavior–goal comparison		
Confident	5.93	5.62
Doubtful	5.06	6.06

Note. Larger numbers indicate greater attention.

Focus of Attention

Three postexperimental items had been included to assess subjects' recollections about what they had focused their attention on while attempting the approach task. Analysis of subjects' self-reports of attention paid to their chronic level of fearfulness toward snakes yielded no significant difference among subject groups (Table 2), with only the self-awareness main effect component approaching significance ($p = .10$).

An analysis of reports of attention paid to assessing whether or not bodily arousal was increasing (see Table 2) yielded two interactions. Of greatest interest was the interaction between pretest group and self-awareness condition, $F(1, 52) = 5.72$, $p < .02$. This interaction indicated that confident subjects tended to report devoting less attention to assessing their arousal when their approach had occurred in the presence of a mirror than if there had been no mirror and that the opposite tendency had occurred among doubtful subjects. Neither of these tendencies was significant by itself, however (ts = 1.56 and 1.81, respectively). In addition, there was an unanticipated Sex × Pretest interaction, $F(1, 52) = 10.99$, $p < .003$, of the following form: Confident males (overall) reported devoting greater attention to assessing arousal than did confident females, $t(52) = 2.18$, $p < .05$, and doubtful males (overall) reported devoting less attention to assessing arousal than did doubtful females, $t(52) = 2.64$, $p < .05$. We have no ready explanation for this gender interaction.

Analysis of self-rated focus on the behavior-goal comparison (see Table 2) revealed an interaction between pretest group and self-awareness condition that approached significance, $F(1, 52) = 3.12$, $p < .08$. In line with previous findings (Carver & Blaney, 1977b), confident subjects tended to report more attention to this comparison, and doubtful subjects reported less attention, when the mirror had been present than when it had been absent. The fact that this interaction took a form opposite to that of the Self-Awareness × Pretest group interaction on attention devoted to assessing arousal suggests that directing attention to the behavior–goal comparison and directing attention to assessing arousal might be complementary processes. When self-rated attention to the behavior–goal comparison was reanalyzed as a proportion of the total attentional focus reported by subjects, the Self-Awareness × Pretest group interaction did achieve an acceptable significance level, $F(1, 52) = 4.62$, $p < .04$. Subsequent comparisons indicated that doubtful subjects in the mirror present condition reported having paid proportionally less attention to the behavior–goal comparison than did doubtful subjects in the mirror absent condition, $t(52) = 2.40$, $p < .04$; the opposite tendency among confident subjects was not reliable ($t < 1$). However, the overall correlation between these two self-reports was quite low ($r = -.10$). This casts some doubt on the notion that the two represent mutually exclusive responses among our subjects.

These analyses of variance all used self-awareness condition as a predictor of the objects of subjects' attention. It is also informative to examine briefly the relationship between perceived anxiety and those objects of attention. Among confident subjects, awareness of anxiety did not predict how much attention was devoted to searching oneself for arousal cues, $r(28) = -.07$; among doubtful subjects, in comparison, there was a clear positive relationship between these variables,

$r(32) = .38$, $p < .02$. These correlations differed reliably from each other ($z = 1.71$, $p < .05$, one-tailed). Additionally, confident subjects displayed a weak but positive relationship between perceived anxiety and attention devoted to the behavior–goal comparison, $r(28) = .19$, $p = .16$, whereas these variables were inversely related among doubtful subjects, $r(32) = -.28$, $p < .06$. These two correlations also differed significantly from each other ($z = 1.65$, $p < .05$, one-tailed). This overall pattern of correlations is quite consistent with our reasoning and with the results of the above analyses of variance.

Postsession Fear

An analysis of subjects' postexperimental ratings of their general levels of fearfulness toward nonpoisonous snakes revealed only a marginally significant tendency for doubtful subjects to report more fearfulness ($M = 3.34$) than confident subjects ($M = 2.82$), $F(1, 52) = 3.21$, $p < .08$. These means represent deviations of .34 and −.18, respectively, from the pretest levels of self-report fearfulness. Correlational analysis provided some evidence that these slight shifts in general fear of snakes were mediated by subjects' observations of their autonomic and physical behavior during the approach task. That is, postsession fearfulness was positively correlated with self-reports of perceived anxiety during the approach task, overall $r(60) = .55$, $p < .001$, and was negatively correlated with actual approach level, overall $r(60) = -.63$, $p < .001$.

Discussion

The results of this study were quite consistent with our hypotheses. According to subjects' self-reports, heightened self-focus led to overall increases in awareness of anxiety-based arousal and in momentary sensations of fearfulness and inadequacy. On the other hand, *behavior* was not directly influenced by self-focus, but rather by self-focus in interaction with expectancies. That is, doubtful subjects withdrew earlier in the approach sequence when self-focus was heightened than when it was not, but no such deficit was exhibited by confident subjects. Self-focus also exerted an interactive influence on the self-reported content of attention during the approach attempt. Confident subjects, though aware of fear, apparently responded to that awareness by attempting to divert their attention from internal fear indices to their behavioral goal. In effect, they showed evidence of increased attempts to match their behavior to the salient standard, even though the approach task quite obviously placed a ceiling on behavior.[5] Doubtful subjects, in contrast, reported increased focus on assessing their bodily arousal and decreased focus on the behavior–goal comparison as functions of heightened self-awareness. These cognitive effects were consistent with the doubtful subjects' early withdrawal from the approach attempt. In sum, the results fit the pattern that had been predicted. Moreover, all of these findings conceptually replicate previous results of Carver and Blaney (1977b).

It is worthy of some additional note that subjects' self-reports of their fearfulness did not allow prediction of their behavior in this task. Our subject groups did not differ from each other in their ratings of their chronic fearfulness of snakes. Yet they differed behaviorally from each other. The ratings of chronic fearfulness were validated to some degree by the finding that reports of situationally experienced fear did not differ reliably between the pretest groups (Table 1). But self-reported fear—even this situational fear—simply did not allow accurate prediction of behavior. Nor did even the self-reported expectancy difference between subject groups predict behavior accurately, except when self-focus was heightened. Apparently these expectancies were utilized by subjects only when attention was self-directed. This pattern of findings is exactly as would be predicted by the sequential analysis presented in the introduction of this article.

---

[5] Enhanced behavioral conformity is somewhat difficult to observe in this context because the desired behavior is clearly prescribed and because there is such a ceiling on performance. However, as discussed in the introduction, heightened conformity to standards as a function of self-attention has been demonstrated previously many times in other contexts (e.g., Carver, 1974, 1975; Scheier et al., 1974; Wicklund & Duval 1971).

The present theory and research have a number of other theoretical implications that deserve mention. The fact that our model is framed in terms of expectancies invites a comparison between it and Bandura's (1977) recent analysis of fear-based behavior. There are some similarities between these models, but there are differences as well.[6] According to Bandura, expectancy judgments make use of several sources of information: the person's prior performances in similar behavioral situations; vicarious experiences, such as the observation of the behavior of another person in the situation; concurrent verbal persuasions, such as externally provided suggestions or self-instructions; and present arousal perceptions. This position on the contribution of arousal perceptions to expectancies was qualified elsewhere (Bandura, Adams, & Beyer, 1977) by noting that attributional considerations play a role in the perceived meaning of the arousal. However, there still appears to be a difference between Bandura's position and ours with regard to the role assumed for arousal perceptions. Our model and the data from the present study suggest that arousal perceptions may cue an assessment process, but that the outcomes of that assessment (expectancy judgments and eventual behavior) depend largely on the other sources of information.[7]

It should be noted, however, that we dealt here with an approach task of only moderate difficulty, which in turn produced only moderate levels of anxiety. It is conceivable that at extremely high levels of arousal, perceptions of autonomic activity can eventually have an impact on the behavior of even confident persons via changes in their expectancies. Indeed, a hint of this more complicated function of arousal perceptions may be contained in the present finding that post-session self-rated fear of snakes was influenced by the anxiety cues that had been perceived during the approach attempt as well as by the level of approach attained. Perhaps subsequent expectancies might also have been altered by one or the other of these sources of information.

One final point should be addressed briefly. A mirror was used to heighten self-attention in this experiment because it has been widely accepted as the "purest" manipulation of self-focus, that is, the least contaminated by other influences. For example, the presence of a videotape camera may raise in a subject's mind the implication that unknown others either are observing now or may observe later. It will be noted, however, that the use of a mirror to induce self-attention in a fear-arousing context leaves open the possibility that any consequent heightened awareness of fear may occur via facial cues rather than internal fear cues. Although this possibility cannot be entirely discounted in the present study, other research (Scheier & Carver, 1977) indicates that enhancement of emotional experience occurs among subjects high in the disposition to be self-attentive as well as among subjects confronted with a mirror manipulation. This convergence of evidence suggests that the primary influence of the mirror, even in an emotional context, is to remind the person of himself or herself. This reasoning also suggests, however, that an important direction for future research will be to test our model's predictions by means of individual differences in self-attention.

---

[6] There are also some similarities between both of these models and Stotland's (1969) analysis of "hope."

[7] Other more subtle conceptual differences between our analysis and that of Bandura are discussed elsewhere (Carver, in press).

References

Bandura, A. Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 1977, *84*, 191–215.

Bandura, A., Adams, N. E., & Beyer, J. Cognitive processes mediating behavioral change. *Journal of Personality and Social Psychology*, 1977, *35*, 125–139.

Buss, D. M., & Scheier, M. F. Self-consciousness, self-awareness, and self-attribution. *Journal of Research in Personality*, 1976, *10*, 463–468.

Carver, C. S. Facilitation of physical aggression through objective self-awareness. *Journal of Experimental Social Psychology*, 1974, *10*, 365–370.

Carver, C. S. Physical aggression as a function of objective self-awareness and attitudes toward punishment. *Journal of Experimental Social Psychology*, 1975, *11*, 510–519.

Carver, C. S. A cybernetic model of self-attention processes. *Journal of Personality and Social Psychology*, in press.

Carver, C. S., & Blaney, P. H. Avoidance behavior and perceived arousal. *Motivation and Emotion*, 1977, *1* 61–73. (a)

Carver, C. S., & Blaney, P. H. Perceived arousal, focus of attention, and avoidance behavior. *Journal of Abnormal Psychology*, 1977, *86*, 154–162. (b)

Carver, C. S., & Scheier, M. F. Self-focusing effects of dispositional self-consciousness, mirror presence, and audience presence. *Journal of Personality and Social Psychology*, 1978, *36*, 324–332.

Duval, S., & Wicklund, R. A. *A theory of objective self-awareness*. New York: Academic Press, 1972.

Duval, S., & Wicklund, R. A. Effects of objective self-awareness on attribution of causality. *Journal of Experimental Social Psychology*, 1973, *9*, 17–31.

Duval, S., Wicklund, R. A., & Fine, R. L. Avoidance of objective self-awareness under conditions of high and low intra-self discrepancy. In S. Duval & R. A. Wicklund, *A theory of objective self-awareness*. New York: Academic Press, 1972.

Fenigstein, A., & Carver, C. S. Self-focusing effects of heartbeat feedback. *Journal of Personality and Social Psychology*, 1978, *36*, 1241–1250.

Froming, W. J. The relationship of moral judgment, self-awareness, and sex to compliance behavior. *Journal of Research in Personality*, in press.

Geller, V., & Shaver, P. Cognitive consequences of self-awareness. *Journal of Experimental Social Psychology*, 1976, *12*, 99–108.

Gibbons, F. X. Sex guilt and reactions to pornography: Enhancing attitude-behavior consistency through self-focused attention. *Journal of Personality and Social Psychology*, 1978, *36*, 976–987.

Gibbons, F. X., Carver, C. S., Scheier, M. F., & Hormuth, S. Self-focused attention and the placebo effect: Fooling some of the people some of the time. *Journal of Experimental Social Psychology*, in press.

Hays, W. L. *Statistics for psychologists*. New York: Holt, Rinehart & Winston, 1963.

Sarason, I. G. Experimental approaches to test anxiety: Attention and the uses of information. In C. D. Spielberger (Ed.), *Anxiety: Current trends in theory and research*. New York: Academic Press, 1972. (a)

Sarason, I. G. Test anxiety and the model who fails. *Journal of Personality and Social Psychology*, 1972, *22*, 410–413. (b)

Scheier, M. F., & Carver, C. S. Self-focused attention and the experience of emotion: Attraction, repulsion, elation, and depression. *Journal of Personality and Social Psychology*, 1977, *35*, 625–636.

Scheier, M. F., Carver, C. S., & Gibbons, F. X. Self-directed attention, awareness of bodily states, and suggestibility. *Journal of Personality and Social Psychology*, in press.

Scheier, M. F., Fenigstein, A., & Buss, A. H. Self-awareness and physical aggression. *Journal of Experimental Social Psychology*, 1974, *10*, 264–273.

Stotland, E. *The psychology of hope*. San Francisco: Jossey-Bass, 1969.

Wicklund, R. A., & Duval, S. Opinion change and performance facilitation as a result of objective self-awareness. *Journal of Experimental Social Psychology*, 1971, *7*, 319–342.

Wine, J. Test anxiety and direction of attention. *Psychological Bulletin*, 1971, *76*, 92–104.

Winer, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

Received July 24, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 7, 1196–1203

Sharing Secrets: Disclosure and Discretion in Dyads and Triads

Ralph B. Taylor
Center for Metropolitan Planning and Research
Johns Hopkins University

Clinton B. De Soto and Robert Lieb
Johns Hopkins University

To develop a more comprehensive picture of the variables that influence disclosure patterns, the impact of group size on sharing secrets was explored. Given Derlega and Chaikin's suggestion that the existence of a closed dyadic boundary is a prerequiste for intimate self-disclosure, it was hypothesized that subjects would be more willing to disclose intimate information in a dyad than in a triad. The results of Experiment 1, which used a role-playing methodology, confirmed the hypothesis. The main effect of group size was observed over a range of roles and items of information. In addition to the main effect, group size interaction effects also indicated that the difference between dyad and triad disclosure rates increased with more intimate items of information and with more intimate roles. These interaction effects suggested that the importance of a closed dyadic boundary depends in part on the expected confidentiality of the interchange. In Experiment 2 the conversations of groups of acquaintances were recorded and rated for intimacy. As predicted, the conversations of dyads were more intimate than those of triads. Suggestions for understanding the intimate quality of dyads are discussed.

Self-disclosure is a topic that has recently generated an enormous amount of research (cf. Cozby, 1973), and substantial support for what Jourard (1971) called the "dyadic effect: disclosure begets disclosure" (p. 66) has been obtained. Two hypotheses—modeling and social exchange—have been offered to explain this effect. The latter hypothesis has received the lion's share of empirical support (Certner, 1973; Davis, 1976; Davis & Skinner, 1974; Derlega, Harris, & Chaikin, 1973; Ehrlich & Graeven, 1971; Jones & Archer, 1976; Worthy, Gary, & Kahn, 1969), and reciprocity has become a dominant theme in self-disclosure research.

However, several limitations preclude drawing conclusions on the generality of reciprocal disclosure. Research has been conducted largely with pairs of strangers, and these interactions may not be prototypical of interactions between acquaintances or friends (cf. Derlega, Wilson, & Chaikin, 1976).[1] Second, almost all disclosure research has been conducted with dyads, and little is known about disclosure patterns in larger groups. For example, when interacting with more than one other, a person may find it difficult to maintain reciprocal disclosure vis-à-vis each other person in the group.

The term dyadic effect seems to suggest that there is something special, apart from reciprocity, about disclosure in dyads. For example, dyads may be more intimate because each person has the undivided attention of the other and can give undivided attention to the

Portions of this research were presented at the meeting of the Eastern Psychological Association in New York, April 1976. This research was conducted while the first author was supported by a predoctoral fellowship from Johns Hopkins University.

The authors are indebted to E. Scott Geller, B. von Haller Gilmer, Gary Gottfredson, Steve Gottfredson, Jack Hamilton, and several reviewers for helpful comments on this article. Glenn Ferguson helped in rating the transcripts.

Requests for reprints should be sent to Ralph B. Taylor, Center for Metropolitan Planning and Research, Johns Hopkins University, Baltimore, Maryland 21218.

[1] For a contrary viewpoint, see Rubin (1974, 1975).

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3707-1196$00.75

other. Although this interpretation of the dyadic effect has intuitive appeal, the little work that has been conducted does more to undermine than support this appeal. Rubin (1976) found that the presence or absence of a third party had no influence on subjects' responses to disclosures by an experimenter. Spinner (Note 1) in his laboratory experiment also found no main effect of group size on depth (intimacy) of disclosure. Drag (1969) found that the two-person discussion groups self-disclosed more than eight-person groups but not more than four-person groups. Group size interaction effects have indicated that people may disclose more or less readily in groups larger than dyads, depending on the composition of the audience (Chelune, 1976) and the mode of communication (Spinner, Note 1).

In sum, although results are mixed, evidence suggests that disclosure patterns in dyads, as compared to other groups, are not unique. Nonetheless, before concluding that there are no effects of group size on depth of disclosure, group size as a source of social influence should be examined.

Blake (1958, p. 229) has suggested that situational sources of social influence include both (a) the "central stimulus" or the "immediate focus of attention" and (b) "context" factors. Group size qualifies as a context factor. In a self-disclosure situation, the recipient of information is the central stimulus. Research has indicated that recipient characteristics such as physical attractiveness (Brundage, Derlega, & Cash, 1977) and a reflective or aggressive style (Ellison & Firestone, 1973) do influence disclosure patterns.

Context factors have received little attention in self-disclosure research, although they can be influential. For example, Johnson and Dabbs (1976) and Rubin and Shenker (1978) both observed that proximity influenced amount of disclosure on low- and medium-intimacy topics.

We wished to further investigate the relationship between group size and disclosure patterns, since (a) context factors in general may exert a widespread influence on social behaviors (Blake, 1958) and (b) group size has been established as a potent influence of social interaction (Bales & Borgatta, 1955; Hackman & Vidmar, 1970; O'Dell, 1968; Slater, 1958; Thomas & Fink, 1963). Exploration of the relationship between group size and self-disclosure uncovers two relevant theoretical considerations. First, Derlega and Chaikin (1977) have suggested that self-disclosure may be viewed as an interpersonal boundary regulation process (Altman, 1975; Altman & Taylor, 1973) and that a precondition of intimate disclosure is the existence of a closed dyadic boundary, that is, "a boundary within which it is perfectly safe to disclose to the invited participant and across which the self-disclosure will not pass" (p. 104). The importance of a closed dyadic boundary has been alluded to before. Simmel (see Wolff, 1950) suggested that dyads have a special quality of intimacy that is not present in larger groups, and thus, dyads could share secrets more often and with more security.

The second theoretical consideration is the layperson's perception of disclosing. De Soto (1960; De Soto & Kuethe, 1959) found that "confides in" was perceived as a pair-forming relationship in contrast to "likes," which was perceived as a group-forming relationship. Confides in was viewed by subjects as symmetric and nontransitive or "essentially a pair-wise interchange" (De Soto & Kuethe, 1959, p. 193). This view of disclosing is congruent with the notion of a closed dyadic boundary.

To test Derlega and Chaikin's (1977) suggestion that a closed dyadic boundary is a prerequisite for intimate disclosure, self-disclosure between acquaintances was examined in dyads and triads. Since a closed dyadic boundary does not exist in a triad, we hypothesized that subjects would be less willing to make intimate disclosures in a triad than in a dyad. Although our hypothesis may seem intuitively obvious to some, it is important to bear in mind that previous research has by and large failed to yield main effects of group size on intimacy of disclosure.[2]

---

[2] Predictions contrary to ours might be drawn from a deindividuation perspective (Diener, Fraser, Beaman, & Kelem, 1976; Zimbardo, 1969), which suggests that the presence of a group serves as a releaser for certain behaviors. Contrary predictions could also be drawn from a sensitivity training or encounter group perspective, which is based on the working assumption

We also expected, in light of Chelune (1976) and Spinner (Note 1), that dyad/triad differences in disclosure rates would increase when more confidential interchanges were expected.

Although intimate disclosures between strangers can be elicited, we decided to focus on intimate disclosures between acquaintances. Given that intimate interchanges occur more frequently between acquaintances than strangers, this focus might enhance the ecological validity of the results.

Experiment 1

Method

Subjects. A specially designed self-disclosure questionnaire, described later, was administered to 21 undergraduate volunteers enrolled in psychology courses at Johns Hopkins University and Towson State University.

Procedure. To manipulate group composition and information content in a controlled fashion, a role-playing technique was used. Willingness versus reluctance to disclose was measured. Subjects were asked to predict how they would behave in hypothetical situations, not how they had behaved in past situations, to assure similarity of situations across subjects. Items contained specific information rather than general topics.

Six items of information were presented on separate pages of the questionnaire with individual roles or pairs of roles listed underneath. The subject was asked to rate how quickly he/she would disclose the item to occupants of various roles or pairs of roles. Each item was encountered twice in the questionnaire—once to assess disclosure to single roles and once to assess disclosure to pairs of roles. The instructions exhorted subjects to think of particular individuals that they knew to fill each role. Disclosure was assessed across several role pairs (sister and best friend of the opposite sex, boy/girlfriend and best friend of the same sex, two liked professors, and acquaintance and roommate) to assess the generality of any effects of group size. Role pairs were chosen on the basis of pilot subjects' suggestions about what role pairs they would be likely to encounter in everyday settings. Finally, disclosure was assessed across six items of information (see Table 1).[3]

A four-category scale ranging from 1 ("I will tell the person [them] as soon as I see him or her [them]") to 4 ("I will probably never tell the person [them]") was used to measure disclosure. Questions concerning disclosure to pairs of individuals were presented separately from those concerning disclosure to single individuals. Order of presentation was randomized across subjects and order of items was randomized within each type of question. At the end of the questionnaire, subjects rated the intimacy of each item of information and the closeness of their relationship with occupants of each of the specific roles.

Results and Discussion

The analysis of variance used a 2 × 4 × 6 (Group Size × Role Pairs[4] × Information) within-subjects factorial design. Although the use of a repeated measures design may possibly have aroused subjects' suspicions, in debriefing no subjects expressed any awareness concerning the group size hypothesis.

The hypothesized main effect of group size was observed, $F(1, 20) = 11.4$, $p < .005$, ($\hat{\omega}^2 = .01$). Subjects indicated that they would be more reluctant to disclose in a triad than in a dyad (see Table 1). This main effect was qualified by two interactions.

A modest though significant Group Size × Information interaction occurred, $F(5, 100) = 2.5$, $p < .05$ ($\hat{\omega}^2 < .01$). The difference in dyad versus triad disclosure rate increased as the intimacy of the information increased (see Table 1). (The rank-order correlation between dyad versus triad disclosure rate and intimacy ratings of the items was .74.) This interaction further supports the hypothesis. As the information under consideration became more intimate, the triad, in comparison to the dyad, was perceived as a less appropriate disclosure setting. Furthermore, relative reluctance to disclose in a triad was a general phenomenon, since it occurred with five out of the six items of information.

Also, a significant Group Size × Role Pairs interaction was obtained, $F(3, 60) = 4.4$, $p < .01$ ($\hat{\omega}^2 = .01$). The difference in dyad versus triad disclosure rate increased as the intimacy rating of the role pairs increased. There was a perfect rank-order correlation

that people are willing to reveal intimate information while in a sizable group.

[3] Only negative items were used, since pilot subjects quickly disclosed items of positive information, regardless of group size or group composition. Although Jones and Wortman (1973) have pointed out that disclosure of positive information can pose problems for the discloser (e.g., being perceived as a bragger), in our samples this was not the case.

[4] To determine disclosure to a role pair when the members of the pair were encountered individually, the mean of the disclosure rate to the two individuals was used.

Table 1
Relationships Among Group Size, Disclosure Rates, and Information in Experiment 1

Information	Intimacy rating	Disclosure rate	
		Dyad	Triad
You have lost a large sum of money from your wallet	3.29	2.24	2.21
You have been rejected from the college or graduate school of your choice	5.14	2.13	2.21
During vacation you got into a car accident due to excessive speed and the person who was with you suffered a broken leg	5.19	2.32	2.63
Your parents are getting divorced	6.66	2.40	2.71
Your brother has been committed to a mental hospital	7.29	2.72	2.98
You have discovered that you have incurable leukemia	7.29	2.89	3.24

Note. On intimacy ratings, a higher score indicates a more intimate item; on disclosure rates, a higher score indicates a slower disclosure rate (i.e., more reluctance to disclose).

between these two measures. Although it was based only on four pairs, this correlation does strongly suggest that a triad, compared to a dyad, became a less appropriate setting for disclosure when the individual was interacting with more intimate contacts.[5]

Consistent with prior self-disclosure research (Derlega & Chaikin, 1975; Jourard & Lasakow, 1958), main effects for role pairs (information was revealed more quickly to more intimate role pairs) and information (more intimate information was disclosed less readily) and a Role Pairs × Information interaction (difference in disclosure rate to acquaintance and roommate and to two liked professors decreased as the intimacy of the items increased) were obtained (all ps $< .001$, ω^2s $= .22, .07$, and $.01$, respectively).

As predicted, subjects thought they would be more reluctant to disclose intimate information in a triad than in a dyad. This main effect and the group size interactions lend support to Derlega and Chaikin's (1977) suggestion concerning the importance of a closed dyadic boundary. Furthermore, the group size interaction effects, although small in terms of magnitude, do suggest that the presence of a closed dyadic boundary became more important as the expected confidentiality of the interchange increased.

Experiment 2

The results of Experiment 1 are limited in that they were based on a role-playing methodology. Subjects were asked to describe how they thought they would act in various situations. Behavioral observations of actual disclosure patterns in dyads and triads would increase our confidence in these findings. We hypothesized that in a fairly unstructured leaderless discussion, conversations among dyads would be more intimate (i.e., contain greater depth of self-disclosure) than conversations among triads.[6]

[5] The results of this experiment were almost wholly replicated with a second larger sample ($n = 26$), using different items and role pairs. In the replication the predicted main effect of group size was again obtained ($p < .005$). A significant Group Size × Information interaction ($p < .05$), similar in interpretation to the one obtained in the first study, was also obtained.

[6] In Experiment 2 an instructional variable (discuss intimate vs. discuss nonintimate topics) could have been included to replicate the group size interaction effect in Experiment 1. Although such an instructional variable was considered, we decided against it because (a) such an instructional variable would make the conversations less natural and (b) we felt it was appropriate to replicate the group size main effect before seeking to replicate any interaction effects.

Method

Subjects and procedure. Freshmen in on-campus housing at Johns Hopkins University were solicited as volunteers by an experimenter. He explained that he was conducting a survey to find out how groups of freshmen felt about the quality of life on campus. The freshmen were randomly asked to contact either one or two of their acquaintances and arrange a time for a group discussion. The experimenter returned to the volunteer's room with a tape recorder at the time appointed for the discussion.

A total of 35 freshmen (19 males and 16 females) agreed to participate. Subjects were grouped into seven same-sex dyads (2 male and 5 female) and seven same-sex triads (2 female and 5 male). Each grouping was composed of acquaintances (i.e., friends the original contact had brought with him/her to the group discussion).[7]

When he arrived at the dorm room, the experimenter explained that the purpose of the discussions was to find out how groups of subjects felt about the quality of life on campus. Fully informed consent was obtained from the subjects.

The experimenter set the general outline for the discussions: "Talk about any problems concerning undergraduate life that you have encountered here as freshmen. Your discussion should, as much as possible, focus on sharing specific personal experiences that each of you has had related to the topic." After answering any further questions the experimenter started the tape recorder and left the room. He was not present during the discussions so that he would not disturb the natural atmosphere of the interactions.

After 9–12 minutes of conversation, the experimenter returned to the room and terminated the discussion. To compensate for increased resource input, triads were allowed a longer discussion time (M = 10.5 minutes) than dyads (M = 9.3 minutes). A short postdiscussion questionnaire was handed out to the subjects. The questionnaire assessed subjects' perceptions of various aspects of the group discussion, subjects' level of acquaintance with other group members, and subjects' general level of disclosure of their personal worries to other role figures (roommate, parents, best friend of the opposite sex, and best friend of the same sex). All questions used a 7-point rating scale, except for the question on level of acquaintance, which used a 4-point scale. Rate of disclosure across the four role pairs was summed to establish a general rate of disclosure of intimate information for each subject. After completing the questionnaire, subjects were fully debriefed.

Results and Discussion

The unit used in all analyses was the group mean. One-way analyses of variance were used to test the effects of group size on two sets of dependent variables: the intimacy of the group discussions and subjects' perceptions of the discussions.[8] The effects of group size were first assessed with no variables held as covariates and then with sex composition, mean level of acquaintance, and mean disclosure habits held as covariates.[9]

Intimacy of discussions was strongly influenced by group size, $F(1, 12) = 21.2$, $p < .001$ ($\hat{\omega}^2 = .59$). As predicted, the discussions of triads were less intimate than the discussions of dyads (see Table 2). The

---

[7] Some previous self-disclosure research has observed that females disclose more readily than males (cf. Cozby, 1973). Thus, one might argue that by overrepresenting females in the dyad condition we are biasing the sample in favor of our hypotheses. However, in this experiment sex composition was not correlated with any of the questionnaire or behavioral dependent variables (all ps $> .05$). This internal analysis indicates that although females were slightly overrepresented in the dyad condition, this did not actually influence the results.

[8] The conversations were transcribed. In the transcripts, there was no indication of the number of people in a particular group. Using a 4-point intimacy scale, each conversation was rated independently by one of the experimenters and by an individual who was totally unfamiliar with the experiment and its purposes. Interrater reliability was assessed using the intraclass correlation (Winer, 1962). Reliability between the two sets of ratings was .96. The mean ratings averaged across the two raters were used in the analysis. The estimated reliability of these mean ratings was .98.

The intimacy scale used was as follows: 1 = little or no disclosure (discussants shift focus away from selves), 2 = superficial or conventional disclosure (reveal trite or peripheral aspects of selves), 3 = personal disclosure (reveal specific experiences concerning more personal topics), 4 = intimate disclosure (specific experiences concerning more personal topics are discussed, and subjects clearly discuss their responses to these experiences).

[9] The analysis of covariance (ANCOVA) used here represents, in the terminology of Evans and Anastasio (1968), Usage 2 of ANCOVA: " 'Adjustment' of treatment means for differences between intact groups, when the covariate is unrelated to the treatments" (p. 227). The covariates used in this analysis (sex composition, level of acquaintance, and disclosure habits) were all statistically independent of the treatment (all rs $< .30$, all ps $> .10$). Furthermore, these covariates were statistically independent of questionnaire (amount learned) and behavioral (rated intimacy of conversations) treatment effects (all rs $< .40$, all ps $> .05$). These correlations are small. Evans and Anastasio note that "small correlations such as might occur . . . in the context of Usage 1 or 2, might not have serious consequences" (p. 233). Thus, in this analysis the usage of ANCOVA was appropriate and was not seriously biased by the small violations of the assumptions of independence.

Table 2
Behavioral and Perceived Effects of Group Size in Experiment 2

Condition	Dependent variable	
	Intimacy of discussion	Amount learned about others
Dyad	2.79	3.36
Triad	1.43	2.24

Note. Intimacy of discussion means were derived from ratings of the transcripts; a higher mean indicates a more intimate discussion. Amount learned means were obtained from subjects' answers to the question: "As a result solely of this discussion, how much do you think you have learned about the other people in the group?," with a higher mean indicating more learned.

influence of group size remained strongly significant with sex composition, acquaintance level, and disclosure habits held as covariates, $F(1, 9) = 12.0$, $p < .01$.

The covariate of acquaintance influenced the intimacy of the discussions, although the effect was much smaller than the effect of group size. Better acquainted groups had more intimate conversations, $F(1, 9) = 5.5$, $p < .05$.

On the postdiscussion questionnaires, the amount subjects learned about other group members was influenced by group size, $F(1, 12) = 5.4$, $p < .05$ ($\hat{\omega}^2 = .30$), with members of dyads learning more about others in the group than members of triads (see Table 2). This result is consistent with the finding that dyads had more intimate discussions than triads. The effect of group size on amount learned persisted, with the three covariates held constant, $F(1, 9) = 7.2$, $p < .05$. Furthermore, amount learned due to the discussion was solely a function of group size. (For all covariates, $Fs < 1$.) Amount learned due to the discussion was independent of other dimensions assessed by the questionnaire. The highest correlation between amount learned and other questions about the interaction was .32.

Members of dyads tended to perceive their discussions as less awkward and more intimate than did members of triads. These effects, although in the hypothesized direction, did not attain the accepted .05 level of statistical significance.

On the questionnaire results, there were two significant effects due to covariates. First, groups composed of members with high disclosure habits perceived the group discussions as less awkward, $F(1, 9) = 14.4$, $p < .01$. Second, the perceived intimacy of the topic discussed was influenced by the level of acquaintance of group members, with better acquainted groups perceiving their discussion topic as more intimate, $F(1, 9) = 14.0$, $p < .01$.

General Discussion

Two experiments using dissimilar methodologies have supported the hypothesis that individuals, when interacting with acquaintances, are more likely to disclose intimate information in a dyad than in a triad. These results support Derlega and Chaikin's (1977) suggestion that intimate disclosure depends on a closed dyadic boundary, that is, the revealer perceiving that his/her message is safe with the recipient. Disclosure was less intimate in triads in which the closed dyadic boundary did not exist. Furthermore, both experiments provided strong evidence that disclosing behavior closely corresponds to the perceived nontransitive nature of "confides in" (De Soto & Kuethe, 1959). Thus, sharing secrets is not only perceived as a pairwise interchange dividing people into dyads but, behaviorally, confiding actually appears to operate in this fashion. (See also Rubin & Shenker, 1978.)

Although there are other differences between dyads and triads in terms of role conflict (Brown, 1965), unanimity of mood (Heider, 1958), and coalition formation (Mills, 1953), the pattern of findings in Experiment 1 (see also Footnote 5) strongly supports our interpretation of the findings. With more intimate items of information and more intimate role pairs, the difference between disclosure rates in dyads and triads increased. Although the magnitude of these interaction effects was not large, the effects were congruent with prior research (Chelune, 1976; Spinner, Note 1). Furthermore, these interaction effects help clarify the role of the closed dyadic boundary. It appears that the salience of the boundary

increases as the confidentiality of the interchange increases.

In a public setting in which strangers are interacting and disclosure is likely to be nonintimate, our theoretical rationale would lead us to expect no effect of group size on disclosure. This was the result obtained by Rubin (1976) in his field experiment. Although it is hazardous to make inferences from negative findings, Rubin's failure to obtain a group size effect in a low disclosure setting is in line with the perspective presented here.

One limitation of the present study is that the effects of group size were explored only over a narrow range. However, this limited range of group sizes was a legitimate focus, since the differences between dyads and triads are more dramatic than the differences between other pairs of group sizes (cf. Hackman & Vidmar, 1970; O'Dell, 1968).

The results of the present study go some distance toward understanding the inherent quality of intimacy that Simmel (Wolff, 1950) attributes to dyads. However, the reasons for this quality are not as yet fully illuminated. Our expectation is that the intimacy of dyads is due to particular perceptual properties (e.g., perceived climate) and structural properties (e.g., role complexity and intensity) that covary with group size.

Furthermore, the present study highlights a host of questions that should be attended to by researchers in self-disclosure. What is the relationship between reciprocity and group size? Although reciprocal disclosure is straightforward in a dyad and involves only one input and one output channel, it is probably more complex in larger groups. In a triad each individual receives two sets of incoming messages and can send two different sets of messages, although everyone "overhears" everyone else's messages. In a triad, people may avoid the difficulties of maintaining two separate channels by disclosing at a uniform low level of intimacy. The apparent robustness of Jourard's dyadic effect and the reciprocity underlying it may in part be due to the past research focus on dyads.

Also, given that context factors such as group structure and central stimulus characteristics such as qualities of the target both influence disclosure patterns, what is the relative impact of each of these on disclosure? Consideration of this question should lead researchers to a more comprehensive understanding of the determinants of self-disclosure.

Reference Note

1. Spinner, B. *Privacy maintenance and self-disclosure.* Paper presented at the 86th annual convention of the American Psychological Association, Toronto, Canada, August 1978.

References

Altman, I. *The environment and social behavior.* Monterey, Calif.: Brooks/Cole, 1975.

Altman, I., & Taylor, D. A. *Social penetration: The development of interpersonal relationships.* New York: Holt, Rinehart & Winston, 1973.

Bales, R. F., & Borgatta, E. F. Size of group as a factor in the interaction profile. In A. P. Hare, E. F. Borgatta, & R. F. Bales (Eds.), *Small groups: Studies in social interaction.* New York: Knopf, 1955.

Blake, R. R. The other person in the situation. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior.* Stanford, Calif.: Stanford University Press, 1958.

Brown, R. *Social psychology.* New York: Free Press, 1965.

Brundage, L. E., Derlega, V. J., & Cash, T. F. The effects of physical attractiveness and need for approval on self-disclosure. *Personality and Social Psychology Bulletin*, 1977, *3*, 63–66.

Certner, B. C. Exchange of self-disclosures in same-sexed groups of strangers. *Journal of Consulting and Clinical Psychology*, 1973, *40*, 292–297.

Chelune, G. J. Self-Disclosure Situations Survey: A new approach to measuring self-disclosure. JSAS *Catalog of Selected Documents in Psychology*, 1976, *6*, 111–112. (Ms. No. 1367)

Cozby, P. C. Self-disclosure: A review. *Psychological Bulletin*, 1973, *79*, 73–91.

Davis, J. D. Self-disclosure in an acquaintance exercise: Responsibility for level of intimacy. *Journal of Personality and Social Psychology*, 1976, *33*, 787–792.

Davis, J. D., & Skinner, A. E. G. Reciprocity of self-disclosure in interviews. *Journal of Personality and Social Psychology*, 1974, *29*, 779–784.

Derlega, V. J., & Chaikin, A. L. *Sharing intimacy: What we reveal to others and why.* Englewood Cliffs, N.J.: Prentice-Hall, 1975.

Derlega, V. J., & Chaikin, A. L. Privacy and self disclosure in social relationships. *Journal of Social Issues*, 1977, *33*, 102–115.

Derlega, V. J., Harris, M. S., & Chaikin, A. L. Self-disclosure reciprocity, liking and the deviant. *Journal of Experimental Social Psychology*, 1973, *9*, 277–284.

Derlega, V. J., Wilson, M., & Chaikin, A. L. Friendship and disclosure reciprocity. *Journal of Personality and Social Psychology*, 1976, *34*, 578–582.

De Soto, C. B. Learning a social structure. *Journal of Abnormal and Social Psychology*, 1960, *60*, 417–421.

De Soto, C. B., & Kuethe, J. L. Subjective probabilities of interpersonal relationships. *Journal of Abnormal and Social Psychology*, 1959, *59*, 290–294.

Diener, E., Fraser, S. C., Beaman, A. L., & Kelem, R. T. Effects of deindividuation variables and stealing among Halloween trick-or-treaters. *Journal of Personality and Social Psychology*, 1976, *33*, 178–183.

Drag, R. M. Experimenter behavior and group size as variables influencing self-disclosure (Doctoral dissertation, University of Florida, 1968). *Dissertation Abstracts International*, 1969, *30*, (5-B), 2416. (University Microfilms No. 69-17,016)

Ehrlich, H. J., & Graeven, D. B. Reciprocal self-disclosure in a dyad. *Journal of Experimental Social Psychology*, 1971, *7*, 389–400.

Ellison, C. W., & Firestone, I. J. Development of interpersonal trust as a function of self-esteem, target status, and target style. *Journal of Personality and Social Psychology*, 1973, *29*, 655–663.

Evans, S. H., & Anastasio, E. J. Misuse of analysis of covariance when treatment effect and covariate are confounded. *Psychological Bulletin*, 1968, *69*, 225–234.

Hackman, J. R., & Vidmar, N. Effects of size and task type on group performance and member reactions. *Sociometry*, 1970, *33*, 37–54.

Heider, F. *The psychology of interpersonal relations*. New York: Wiley, 1958.

Johnson, C. F., & Dabbs, J. M., Jr. Self-disclosure in dyads as a function of distance and the subject–experimenter relationship. *Sociometry*, 1976, *39*, 257–263.

Jones, E. E., & Archer, R. L. Are there special effects of personalistic disclosure? *Journal of Experimental Social Psychology*, 1976, *12*, 180–193.

Jones, E. E., & Wortman, C. B. *Ingratiation: An attributional perspective*. Morristown, N.J.: General Learning Press, 1973.

Jourard, S. M. *The transparent self*. New York: Nostrand, 1971.

Jourard, S. M., & Lasakow, P. Some factors in self-disclosure. *Journal of Abnormal and Social Psychology*, 1958, *56*, 91–98.

Mills, T. M. Power relations in three person groups. *American Sociological Review*, 1953, *18*, 351–357.

O'Dell, J. W. Group size and emotional interaction. *Journal of Personality and Social Psychology*, 1968, *8*, 75–78.

Rubin, Z. Lovers and other strangers: The development of intimacy in encounters and relationships. *American Scientist*, 1974, *62*, 182–190.

Rubin, Z. Disclosing oneself to a stranger: Reciprocity and its limits. *Journal of Experimental Social Psychology*, 1975, *11*, 233–260.

Rubin, Z. Naturalistic studies of self-disclosure. *Personality and Social Psychology Bulletin*, 1976, *2*, 260–263.

Rubin, Z., & Shenker, S. Friendship, proximity and self-disclosure. *Journal of Personality*, 1978, *46*, 1–22.

Slater, P. E. Contrasting correlates of group size. *Sociometry*, 1958, *21*, 129–139.

Thomas, E. J., & Fink, C. F. The effects of group size. *Psychological Bulletin*, 1963, *60*, 371–384.

Winer, J. S. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

Wolff, K. H. (Ed.). *The sociology of Georg Simmel*. New York: Free Press of Glencoe, 1950.

Worthy, M., Gary, A. L., & Kahn, G. M. Self-disclosure as an exchange process. *Journal of Personality and Social Psychology*, 1969, *13*, 59–63.

Zimbardo, P. G. The human choice: Individuation, reason and order versus deindividuation, impulse and chaos. In W. D. Arnold & D. Levine (Eds.), *Nebraska Symposium on Motivation* (Vol. 17). Lincoln: University of Nebraska Press, 1969.

Received July 26, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 7, 1204–1210

Expressive Tendencies and Physiological Response to Stress

Clifford I. Notarius
Catholic University of America

Robert W. Levenson
Indiana University

This study assessed the effects of natural expressive tendencies on physiological response to stress. Male undergraduates were unobtrusively observed while watching a stressor videotape. On the basis of the subjects' facial responsiveness to the film, a group of 23 natural expressers and 22 natural inhibitors were selected and exposed to a threat of shock situation during which heart rate, respiration rate, skin conductance, and facial expressions were monitored. In accord with the discharge model of emotion, natural inhibitors were less facially expressive and more physiologically reactive to the shock threat than were natural expressers. The results also demonstrated that overt expressivity is stable over time and situation. On personality measures, natural expressers scored significantly higher on Mehrabian's empathic tendency scale, thus supporting the efficacy of this paper-and-pencil instrument as a measure of nonverbal responsiveness. The two groups did not differ on measures of self-esteem, introversion–extraversion, or locus of control. The results are discussed in terms of the discharge model as a descriptive metaphor and not a causal theory.

Understanding the relationship between overt emotional expression and physiological response is necessary for a comprehensive theory of emotion as well as for an adequate conceptualization of behavior change following any of several expressive therapies (Nichols & Zax, 1977). Facial display is one component of emotional expression that has been shown to influence both physiological reaction patterns and the subjective experience of emotion (e.g., Izard, 1978; Laird, 1974). Although there is agreement that facial display is an important component of emotional response, there is conflicting evidence about whether facial display attenuates (Buck, Miller, & Caul, 1974; Buck, Savin, Miller, & Caul, 1972; Lanzetta & Kleck, 1970) or augments (Lanzetta, Cartwright-Smith, & Kleck, 1976) physiological response to emotionally arousing situations.

Lanzetta and Kleck (1970), Buck et al. (1972), and Buck et al. (1974) used an encoding/decoding paradigm to study the relationship between facial displays of emotion and physiological response to arousing situations. In these studies, a sender subject was exposed to an emotionally arousing event while an observer subject watched the subject's face on a video monitor and attempted to decode the sender's expression. The sender's skin conductance or heart rate or both were monitored throughout the stimulus presentation. The results of these studies indicated that the observer subjects were most accurate at decoding the facial expressions of sender subjects who were least physiologically reactive to the eliciting stimuli; conversely, observer subjects were least accurate at decoding the facial expressions of sender subjects who were most physiologically reactive. The results of these studies, which imply that reduced decoding accuracy was due to relatively fewer facial expressions, have been interpreted as support for

This article is based on data collected as part of the first author's doctoral dissertation.

The authors would like to thank Chrys Ruybal, Kitty Kane, and Michael Reed for serving as coders, and Alyne Robin and George Doxas for serving as experimenters. Howard J. Markman provided valuable comments on an earlier draft of this article.

Requests for reprints should be sent to Clifford I. Notarius, Department of Psychology, Catholic University of America, Washington, D.C. 20064.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3707-1204$00.75

the discharge model of emotion according to which facial expressions are associated with attenuated physiological response to emotional stimuli.

Unfortunately, the encoding/decoding paradigm does not provide a strong empirical test of the discharge model. In the studies that used this paradigm, the assessment of overt expressivity was dependent on the measure of decoding accuracy and the implication that this measure reflects overt expressivity. A more direct test of the relationship between facial expressiveness and physiological reactivity would be to place subjects in an arousing situation with trained raters recording the extent of their facial expressiveness. This methodology would enable an objective assessment of a subject's tendency to facially display a response to an emotionally arousing situation. Comparison of the objectively rated facial display data with recorded physiological response data would allow a test of the discharge model in situations in which natural response tendencies would most likely be evidenced.

An alternative approach is possible in which an attempt is made to directly manipulate the extent of facial responsiveness. Lanzetta et al. (1976) tested the discharge model in a well-designed study incorporating experimentally manipulated facial displays and independent assessment of expressivity during exposure to electric shock. Lanzetta et al. instructed subjects to either pose an intense expressive reaction or to pose no reaction to electric shock and found that subjects who were instructed to pose no reaction were significantly less *physiologically* reactive to the shock than subjects who had posed an intense reaction. These results are opposite to predictions based on the discharge model. Lanzetta et al. interpreted their results as support for proprioceptive feedback models of emotion (Gellhorn, 1964; Izard, 1971).

The discrepancy between Lanzetta et al.'s (1976) findings and previous support for the discharge model (Block, 1957; Buck et al., 1974; Buck et al., 1972; Jones, 1950; Lanzetta et al., 1970; Learmonth, Ackerly, & Kaplan, 1959) may stem from differences across studies in the operationalization of emotional expressivity. With the exception of Lanzetta et al.'s (1976) study, all previous investigations that found support for the discharge model allowed subjects access to their natural emotional responses during exposure to an eliciting situation. Thus, as Lanzetta et al. suggest, short-term, experimentally manipulated control of emotional expression may lead to a positive relationship between facial expression and physiological reactivity, whereas natural response patterns may be characterized by the discharge model.

One possible mediator of this short-term, positive relationship may be patterns of general somatic activity that are activated by instructions to "be responsive." To the extent that subjects become more physically active, one would expect such activity to be accompanied by increases in heart rate and skin conductance. Conversely, if instructions to hide responses produce a decline in somatic activity and a more general state of relaxation, decreases in cardiac and electrodermal activity might follow. In any event, it seems likely that results derived from experiments that involve natural expressive tendencies will differ from the results obtained from direct manipulation of expressivity, insofar as the two methodologies reveal different aspects of the relationship between facial expressivity and physiological responsiveness.

The purpose of the present study was to determine the relationship between facial displays of emotion and physiological reactivity to stress in subjects whose natural expressive styles were unconstrained. The experimental design incorporated a behavioral coding of facial expressivity (Mehrabian, 1972) and multiple measures of physiological reactivity (heart rate, respiration rate, and skin conductance). Based on the consistency of findings from studies in which subjects were allowed their natural emotional responses, it was predicted that natural expressers would be less physiologically reactive to an emotionally arousing situation than natural inhibitors.

A secondary purpose of the study was to explore personality correlates of expressive tendencies. Buck et al. (1974) reported that expressers were higher in self-esteem and were more extraverted than inhibitors. In the present study, subjects were assessed on self-esteem, introversion–extraversion, and empathic tendency.

Method

Subject Selection

To test the experimental hypotheses, it was necessary to preselect a group of natural inhibitors (persons who show little or no facial responsivity to an emotional situation) and a group of subjects who were natural expressers (people who show high levels of facial responsivity to an emotional situation). Seventy-six male undergraduates enrolled in introductory psychology classes were recruited for the preselection phase of the experiment; they were fulfilling a course requirement.

To establish natural expressive tendencies, subjects were shown an industrial accident film that had been demonstrated to be emotionally arousing (Lazarus, Opton, Nomikos & Rankin, 1965). The subjects were told that they would view a brief film and then be asked to complete several questionnaries. The questionnaires were Mehrabian's empathic tendency scale, Eysenck's Introversion–Extraversion Scale, Rotter's Locus of Control Scale, and Janis and Field's Self-Esteem Inventory.

Subjects viewed the industrial accident film in groups of four to six while three coders (two undergraduates and the first author) observed through a one-way mirror. Throughout the film, coders counted the number of facial expressions occurring in 30-sec time periods and entered this number on a coding sheet. Criteria for determining the occurrence of a facial expression were based on a procedure described by Mehrabian (1972). Using this procedure, changes from a neutral display to a nonneutral display and back to a neutral display constituted one facial expression. Gestural behaviors, such as slight movements of the eyebrows or touching the face with the hand, were not counted as facial expressions. Although more elaborate qualitative coding schemes exist for classifying facial displays in terms of discrete emotional states (Ekman & Friesen, 1975), it was decided to use a relatively simple quantitative coding system in this study to determine if a general relationship between natural facial expressiveness and physiological reactivity could be detected using a stress paradigm.

In addition to recording the number of facial expressions, coders assigned a subjective rating on a 10-point scale to each subject after the film, with 0 indicating little overt expressivity and 10 indicating a high level of expressivity. For all groups of four to six subjects, each coder observed three subjects, thus generating three independent codings of at least one subject during each screening of the film. This coding overlap permitted continuous reliability estimates throughout the subject selection procedure. Correlations between observers' codings averaged .88 (range, .79 to .95 for the number of expressions) and averaged .91 for the subjective index (range, .85 to .95).

Subjects who received a behavioral rating of 9 expressions or more and a subjective rating of 5 or more were classified as natural expressers, whereas subjects receiving a behavioral rating of 3 expressions or less and a subjective rating of 1 or less were classified as natural i-' ›rs. On this basis, 23 natural inhibitors and 22 natural expressers were selected for continuation in t experiment.

Procedure

A single laboratory appointment was arranged f each of the 45 subjects selected for continuation in t experiment. As each subject arrived at the laborator he was seated in a comfortable chair in front of a vid monitor and an unobtrusively placed video camera. T subject was told that he would be viewing a film on t monitor and that the experimenters were interested people's physiological reactions to the film. Physiolog cal sensing devices to detect heart rate, respiration rat and skin conductance were then attached to the subjec and he was told that the experimenter (one of two r search assistants blind to the experimental hypothese was going to set up the film and check to see that th physiological recording equipment was workin properly.

Just prior to leaving the subject, the experimente informed the subject that he would soon see a "digita voltmeter" displayed on the monitor via a close circuit TV camera placed in the adjoining room. Th subject was told that the "voltmeter" indicated th voltage flowing through his body as measured by a cli (actually a ground clip) attached to his ear. Th experimenter explained that the meter had an intern circuit to sense rapid increases in voltage and that if th internal circuit was activated, "9999" would begin t flash as a warning that a dangerously high level o voltage was present and that a strong shock coul result. The subject was then asked to monitor th voltmeter and to signal the experimenter immediatel if 9999 began to flash by ringing a buzzer conveniently placed to the subject's right. The experimenter asked if the subject had any questions and then left, tellin the subject that he had to check the recordin equipment.

In reality, the monitor displayed a videotape th was identical for all subjects and on which was record a small digital light display of four numbers. The pre programmed display presented "0000" for 4 min, lead ing to small increases over the next 30 sec until 999 began to flash on and off for 1 min. The flashing display of 9999 constituted the threat of shock. As soon as the subject signaled the onset of 9999, the experimente began to jiggle the wires in the adjoining room as if to attempt to locate the "problem." A loud thump on the wall was produced to accompany the return of 0000 to the video monitor; then 0000 remained on the monito for 4.5 min (the poststimulus period). The experimente entered the subject room and apologized for the flashing 9999 by explaining that this was the first time this had happened, that the cause was located and corrected, and that the subject was in no danger. The subject was then shown a second stressor film and exposed to an additional arousing situation before being debriefed.[1]

[1] Only data gathered in response to the threat of shock are reported. Although all subjects watched the film and were exposed to a third eliciting situation, it

Monitoring physiological responses. Skin conductance was measured by passing a constant voltage between two Beckman electrodes placed on the volar surfaces of the middle segments of two fingers of the right hand. Heart rate was recorded bipolarly using Beckman surface electrodes. Respiration rate was assessed by stretching a mercury strain gauge 10 inches (25.4 cm) in length above the subject's waist. Physiological signals were passed through a Grass Model 7 polygraph and routed through the analog-to-digital converter of a PDP-11 computer for on-line processing.

Monitoring facial expressivity. Facial expressivity was coded by the same coders using a similar procedure to that used in the preselection of subjects, with the exception that only one coder rated each subject.[2] The coder watched the subject's face on a closed-circuit video monitor located in a separate room and pressed a button for the duration of each nonneutral facial display. In the case of expressions maintained over 5 sec, the computer automatically counted an additional expression. Coders were blind as to which group the subject was assigned.

Videotape functions. The videotape served two functions, (a) to present the threat of shock stimulus to the subject and (b) to signal the end of the trial periods to the computer. An inaudible signal (17,000 kHz) placed on the videotape at spaced intervals (30 sec during stimulus presentations and 90 sec during baseline periods) momentarily closed a decoding switch, signaling the computer to begin a new trial period. The videotape continued for the duration of the experiment to ensure identical timing of the experiment for each subject.

Dependent measures. The computer was used to obtain the following dependent variables: (a) heart rate (interbeat interval)—time, in msec, between heart beats; (b) respiration rate (intercycle interval)—time, in msec, between inspiration and expiration cycles; (c) skin conductance—skin conductance, in micromhos; and (d) facial expression—the number of facial expressions coded.

Results and Discussion

Physiological reactivity was assessed by comparing average responses occurring in a 4.5-minute prestimulus period with average responses occurring in a 1-minute stimulus period. A 4.5-minute poststimulus period was compared with the prestimulus period to determine the extent of return to resting levels following exposure to an emotionally arousing situation. Dependent variables were entered into a Groups × Trials repeated measures analysis of variance with two groups (natural inhibitors and natural expressers) and three trials (prestimulus, stimulus, and poststimulus). Planned comparisons by t test were used to test specific reaction patterns of natural expressers and natural inhibitors.

Due to equipment malfunction, heart rate data were lost for one natural inhibitor across all stimulus periods, respiration rate data were lost for one natural expresser during the stimulus period, and skin conductance data were lost for one natural expresser across all stimulus periods and for one natural expresser during the prestimulus period. The remaining available data from these subjects were included in the analyses.

Stability of the inhibitor–expresser dimension. The preselection of subjects was intended to yield two groups of subjects who differed in the tendency to display facial expressions in response to emotionally arousing situations. In the absence of evidence concerning the stability of the tendency towards overt facial displays of emotion, it was hoped that a selection procedure based on behavioral observations of expressions in response to a stressor situation would be most likely to reliably identify the two groups. Examination of the mean number of facial expressions displayed in response to the threat of shock stimulus indicated that natural expressers averaged significantly more facial expressions than did natural inhibitors, $t(86) = 2.53$, $p < .025$; the mean numbers of expressions were 1.34 and .87, respectively. This result confirms the preselection classification of subjects and demonstrates stability of expressive tendencies across two independent situations (film and threat of shock) and across time (approximately 9 weeks separated the preselection film and the threat of shock situations).

Assessment of the arousing situation. Since the threat of shock was a novel stressor, it was important to determine whether subjects found the situation arousing. A significant trials effect

---

was decided that the within-subject design used in this study did not permit valid interpretation of the physiological data gathered following the threat of shock.

[2] Even though several months had passed since the preselection phase of the experiment, every attempt was made to ensure that a given subject's facial expressions were coded by a different rater in the preselection and laboratory phases of the research. Use of separate groups of coders in both phases of the study and having two coders per subject in the laboratory phase (to allow continuation of reliability monitoring) would be desirable in future work.

Table 1
Physiological Reactivity Patterns for Inhibitors and Expressers

Subjects	Measure	Prestimulus	Shock threat	Poststimulus
Inhibitors	Heart rate[a]	865	818*	882
	Respiration rate[b]	3648	3298*	3621
	Skin conductance (micromhos)	12.7	16.6*	15.7*
Expressers	Heart rate[a]	823	820	838
	Respiration rate[b]	3593	3530	3687
	Skin conductance (micromhos)	10.8	14.5*	12.9*

[a] Time in msec between heart beats.
[b] Time in msec between inspiration and expiration cycles.
* Significantly different from prestimulus level ($p < .001$).

was found on all dependent measures, indicating that the threat of shock was a physiologically arousing situation; for facial expression, $F(2, 86) = 35.12$, $p < .001$; for heart rate, $F(2, 84) = 11.21$, $p < .001$; for respiration rate, $F(2, 85) = 10.61$, $p < .001$; and for skin conductance, $F(2, 83) = 49.49$, $p < .001$. Anecdotal data gathered during the debriefing also indicated that the subjects found the situation arousing and believable. One subject, for example, related, "When the nines began to flash I really thought I was in for it."

Reactivity patterns in natural expressers and natural inhibitors. On measures of heart rate and respiration rate, natural expressers were less reactive to the threat of shock stimulus than were natural inhibitors. Natural inhibitors showed a significant heart rate increase, $t(84) = 3.76$, $p < .001$, and a significant respiration rate increase, $t(85) = 4.34$, $p < .001$, from the prestimulus to the stimulus period, whereas natural expressers showed no significant change from prestimulus levels in heart rate, $t(84) = .28$, *ns*, or respiration rate, $t(85) = .78$, *ns*. During the poststimulus period, heart rate and respiration rate of both groups returned to prestimulus levels. Heart rate and respiration rate means are presented in Table 1. These results are consistent with the discharge model.

Analysis of the skin conductance measure indicated that both groups were similarily reactive to the threat of shock; natural inhibitors showed a significant increase in skin conductance from prestimulus levels, $t(83) = 7.11$, $p < .001$, as did natural expressers, $t(83) = 6.67$, $p < .001$. During the poststimulus period, each group maintained a significantly higher skin conductance level relative to the prestimulus period, $t(83) = 5.49$, $p < .001$, for natural inhibitors; $t(83) = 3.89$, $p < .001$, for natural expressers. Skin conductance means are presented in Table 1. Skin conductance reactivity patterns did not discriminate natural expressers from natural inhibitors, as would be predicted on the basis of the discharge model.

Resting levels of natural inhibitors and natural expressers. Although reactivity patterns from the prestimulus to the stimulus period are the critical test of the discharge model, baseline differences occurring in the prestimulus period were also assessed. Natural expressers displayed a faster heart rate during the prestimulus period than did natural inhibitors, $t(84) = 3.35$, $p < .001$; whereas natural inhibitors displayed a significantly higher skin conductance level than did natural expressers, $t(83) = 3.49$, $p < .001$. The two groups did not display any prestimulus respiration rate differences, $t(85) = .68$, *ns*.

These baseline differences, observed in subjects who had just entered the laboratory and who were told only to monitor a "voltmeter," were unexpected. Furthermore, the pattern of the baseline differences indicated that one group was not simply more aroused to the general laboratory situation than the other group, since heart rate was elevated for natural expressers whereas skin conductance was elevated for natural inhibitors. These findings,

together with the lack of correspondence between skin conductance reactivity and heart rate and respiration rate reactivity, suggest that expressive tendencies may be associated with differences in individual physiological response stereotypy (Roessler & Engel, 1974) between expressers and inhibitors. This interpretation is offered tentatively, pending replication and extension of the results.

Personality correlates of expressive tendencies. Natural expressers were shown to be significantly more empathic on Mehrabian's empathic tendency questionnaire than were natural inhibitors, $F(1, 43) = 5.37$, $p < .025$. In a series of studies, Mehrabian (1972) presented evidence that overt responsiveness is the primary characteristic of subjects who score high on this questionnarie. Thus, the empathic tendency questionnaire appears to have discriminative validity.

The two groups did not significantly differ on measures of introversion–extraversion, self-esteem, or locus of control. These results are contrary to the findings of Buck et al. (1974).

Conclusion

The results of the present study are consistent with the discharge model; expressive subjects were significantly less physiologically reactive to an emotional stressor than were nonexpressive subjects. Furthermore, the stability of the inhibitor–expresser dimension over time and situation suggests an enduring attribute of affective style. However, it would be inappropriate, on the basis of these data, to conclude that facial expression functions as a causal agent in determining the parameters of physiological response to stress. More appropriate experimental designs for examining the causal relationship in which expressiveness is directly manipulated (e.g., Lanzetta et al., 1976) have produced results that are discrepant with the discharge model. This discrepancy between the results of studies in which subjects were allowed their natural response patterns and those in which subjects were asked to control their emotional expression underscores the fact that the discharge model is a descriptive metaphor for clinical and empirical findings that expressive individuals are less physiologically reactive to stress than are natural inhibitors. The fact that facial expressions are not mediating the relationship between expressivity and physiological reactivity leads to speculation concerning a third factor or factors that are responsible for the observed relationship.

Speculatively, the causal factors mediating the observed relationships between expressivity and physiological reactivity may involve the subject's cognitive appraisal of the eliciting situations. Given personality differences between expressive and inhibited subjects (Block, 1957; Jones, 1950; Learmonth et al., 1959), it would not be unexpected to find characteristic differences between the groups in the cognitive processes that function to shape an emotional response out of an environmental stressor. According to Lazarus, Averill, and Opton (1970), the outcome of these cognitive, subjective processes is an appraisal of the eliciting situation (along such dimensions as perceived danger, threat, or security) that determines a complex emotional response, which includes both overt expression and physiological response. Prestimulus differences in physiological levels observed in the present study may reflect the operation of differential appraisal from the moment the subjects entered the laboratory environment.

Attempts to refine our understanding of the determinants of different physiological reactivity patterns between expressive and nonexpressive individuals, whether or not cognitive appraisal is the primary mediator, must acknowledge the influence of several factors that have yet to be addressed. The first of these is the possible role of stimulus specificity. Although the facial responses in the present study were similar for the accident film and the threat of shock, it is conceivable that other stimuli might produce different patterns of facial and physiological response. In this respect, it would be important to examine stressors of a more interpersonal nature as well as situations in which positive affect is elicited. Second, continued research must recognize that natural expressive tendencies cannot be equated with instructionally manipulated displays of emotion. Short-term control of expression does not seem to be the equivalent of an expressive response from an individual's natural repertoire. Third, a more qualitative coding of facial ex-

pressiveness would be useful in determining whether the discharge model is uncritical as to the nature of facial display or whether such dimensions as appropriateness of affect, duration of display, or specific emotion displayed improve the predictive power of the model. Finally, further refinements in the scope of physiological response analysis may be important. Although it is generally agreed that monitoring one physiological system (e.g., skin conductance) is insufficient for adequately assessing physiological response, there is less consensus as to how data from multiple physiological systems should be handled in light of such factors as stimulus specificity and individual response stereotypy, which may strongly influence the nature of interrelationships among facial and physiological indices of expressivity.

References

Block, J. A study of affective responsiveness in a lie-detection situation. *Journal of Abnormal and Social Psychology*, 1957, *55*, 11–15.

Buck, R. W., Miller, R. E., & Caul, W. F. Sex, personality, and physiological variables in the communication of affect via facial expression. *Journal of Personality and Social Psychology*, 1974, *30*, 587–596.

Buck, R. W., Savin, V. J., Miller, R. E., & Caul, W. F. Communication of affect through facial expressions in humans. *Journal of Personality and Social Psychology*, 1972, *23*, 362–371.

Ekman, P., & Friesen, W. V. *Unmasking the face*. Englewood Cliffs, N.J.: Prentice-Hall, 1975.

Gellhorn, E. Motion and emotion: The role of proprioception in the physiology and pathology of emotions. *Psychological Review*, 1964, *71*, 457–472.

Izard, C. E. *The face of emotion*. New York: Appleton-Century-Crofts, 1971.

Izard, C. E. *Human emotions*. New York: Plenum Press, 1978.

Jones, H. E. The study of patterns of emotional expression. In M. L. Reymert (Ed.), *Feelings and emotions*. New York: McGraw-Hill, 1950.

Laird, J. D. Self-attribution of emotion: The effects of expressive behavior on the quality of emotional experience. *Journal of Personality and Social Psychology*, 1974, *29*, 475–486.

Lanzetta, J. T., & Kleck, R. E. Encoding and decoding of nonverbal affect in humans. *Journal of Personality and Social Psychology*, 1970, *16*, 12–19.

Lanzetta, J. T., Cartwright-Smith, J., & Kleck, R. E. Effects of nonverbal dissimulation on emotional experience and autonomic arousal. *Journal of Personality and Social Psychology*, 1976, *33*, 354–370.

Lazarus, R. S., Averill, J. R., & Opton, E. M. Towards a cognitive theory of emotion. In M. A. Arnold (Ed.), *Feelings and emotions*. New York: Academic Press, 1970.

Lazarus, R. S., Opton, E. M., Nomikos, M. S., & Rankin, N. O. The principle of short circuiting of threat: Further evidence. *Journal of Personality*, 1965, *33*, 622–635.

Learmonth, G. J., Ackerly, W., & Kaplan, M. Relationships between palmar skin potential during stress and personality variables. *Psychosomatic Medicine*, 1959, *21*, 150–157.

Mehrabian, A. *Nonverbal communication*. Chicago: Aldine, 1972.

Nichols, M., & Zax, M. *Catharsis in psychotherapy*. New York: Gardner Press, 1977.

Roessler, R., & Engel, B. T. The current status of the concepts of physiological response specificity and activation. *International Journal of Psychiatry in Medicine*, 1974, *5*, 359–366.

Received August 11, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 7, 1211-1220

The Cognition–Emotion Process in Achievement-Related Contexts

Bernard Weiner, Dan Russell, and David Lerman
University of California, Los Angeles

Two experiments were conducted to examine the linkages between cognitions and emotions. In the first investigation, subjects reported a "critical incident," in which they succeeded or failed an exam for a particular reason (e.g., help from others, lack of long-term effort). They then recounted three affects that were experienced. The data revealed prevalent affects linked with success and failure regardless of the attribution for the outcome. But many emotions were identified that are associated with specific attributions (e.g., luck-surprise; others-gratitude and others-anger). In addition, dimensions of causal attributions, such as locus, also influence recollected feeling states, particularly esteem-related emotions. Hence, it was proposed that in achievement-related contexts there are three sources of affect elicited by disparate cognitions. In the second experiment, it was demonstrated that individuals can use emotional cues to infer why a success or a failure has occurred. The proposed cognition–emotion and emotion–cognition couplings appear to be symmetrical.

The research reported in this article was guided by the simple presumption that a variety of cognitions, particularly causal attributions, influence emotional reactions in achievement-related contexts. This belief recently found suggestive support in an investigation conducted by Weiner, Russell, & Lerman (1978). The questions raised in Weiner et al. provide the foundation for the research reported here.

Weiner et al. compiled a dictionary list of approximately 250 potential affective reactions to success and failure in an academic context. The dominant causal attributions for achievement performance were also identified (Elig & Frieze, 1975; Frieze, 1976). Then a cause for success or failure was given within a brief story format, the success- or failure-related affects that had been identified were listed, and the subjects reported the intensity of the affective reactions that they thought would be experienced in this situation. Responses were made on simple rating scales. A typical story follows.

> Francis studied intensely for a test he took. It was very important for Francis to record a high score on this exam. Francis received an extremely high score on the test. Francis felt that he received this high score because he studied so intensely [his ability in this subject; he was lucky in which questions were selected; etc.]. How do you think Francis felt upon receiving this score? (Weiner et al., 1978, p. 70)

A number of provocative findings emerged from this investigation. First, there were a group of "outcome dependent – attribution independent" affects that were rated as vividly and equally experienced, regardless of the perceived attribution or the "why" of success. Examples of affects given success were pleasure, happiness, satisfaction, and goodness, whereas for failure the outcome dependent – attribution independent affects included uncheerfulness, displeasure, and being upset.

In addition, for both success and failure many affects were discriminably related to specific attributions. Table 1 shows a subset of the causal attributions for success and failure and a label that best describes their distinguishing linked emotion.

This research was supported by Grant MH25687-04 to Bernard Weiner from the National Institute of Mental Health.

Requests for reprints should be sent to Bernard Weiner, Department of Psychology, University of California, Los Angeles, Los Angeles, California 90024.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3707-1211$00.75

Table 1
Attributions and Dominant Discriminating Affects for Success and Failure

Attribution	Success	Failure
Ability	Competence	Incompetence
	Confidence	—
Unstable effort	Activation	Guilt
	Augmentation	Shame
Stable effort	Relaxation	Guilt, shame
Personality	Self-enhancement	Resignation
Others	Gratitude	Aggression
Luck	Surprise	Surprise

Note. Table 1 based on Weiner, Russell, & Lerman (1978).

Table 1 reveals that for success, attributions to ability were thought to give rise to feelings of confidence and competence; unstable effort ascriptions produced heightened activation and high potency emotions (e.g., being uproarious and delirious), whereas stable effort attributions brought about relaxation; personality attributions resulted in self-enhancement (e.g., conceit and pride); attributions to others were associated with gratitude; and ascriptions to luck were linked with surprise. Given failure, ability attributions were perceived as giving rise to feelings of incompetence; effort ascriptions generated reports of shame and guilt; personality attributions were linked with resignation; ascriptions to others caused vindictiveness and aggression; and bad luck produced surprise.

Many of the findings summarized in Table 1 seem obvious—as is often the case, "the truth is not new." Yet there were also unexpected results and unanticipated revelations. For example, the data suggest that whether success is followed by activation or calmness depends on a discrimination between long- and short-term effort as the perceived cause of the outcome. Experiments by Lazarus (1966) also suggest that arousal (i.e., activation) is a product of a cognitive appraisal process. This position contrasts with the well-known contention of Schachter (1964) that the labeling of internal arousal precedes and defines one's emotional experience.

Table 1 reveals some interesting comparisons between emotional reactions to success and failure. If personal ability is the perceived cause, then, as might be anticipated when the outcomes are diametric, the reported experiences for success and failure are in direct opposition: respectively, competence in contrast to incompetence. This antithesis is also evident when attributions are to others (the generated emotions for success and failure, respectively, are gratitude vs. hostility). But if luck is the causal attribution, then the reactions to both success and failure include surprise. And, given effort ascriptions, the feelings associated with success and failure—high activation or calmness versus shame and guilt—are unrelated.

In spite of the systematic and significant findings reported by Weiner et al. (1978), there is reason to be skeptical about the data. These investigators cautioned that

> This procedure is fraught with danger, even as a starting point. First of all, we assumed that individuals would project their own emotional experiences, or those observed in others, upon the characters in the stories. Second, we assumed that our labels reflect the "real" experiences of the subjects. Finally, we assumed that repression and surpression, memory distortion, response sets, experimenter demands, and individual differences in affective labeling as well as in the subjective meaning of these labels would not render our results meaningless (see Davitz, 1969). In sum, there obviously are many limitations of this initial study. (Weiner et al., 1978, p. 70)

Later it was warned that

> The reactive and respondent measure undoubtedly encourages subjects to report affects that are not experienced, or are experienced in a manner not truly captured by the labels. (Weiner et al., 1978, p. 71)

A series of follow-up studies was therefore initiated with the purpose of replicating the prior findings with altered methodologies and extending the research in new directions. The two investigations reported here introduce the changes given below.

(a) In Experiment 1, a "critical incident" technique (Davitz, 1969) is used in which subjects relate a personal experience of success or failure, given a designated causal attribution. Free responses, rather than rating scales of specific affects, are the dependent indicators of emotions. Thus, the procedure uses operant methods of affective assessment, whereas Weiner et al. (1978) used a respondent methodology (McClelland, 1971). In addition, the methodology is not simulational (although it is retrospective).

Table 2
Percentage of Subjects Recounting a Critical Incident as a Function of Outcome and Causal Attribution

Outcome	Attribution					
	Ability	Unstable effort	Stable effort	Personality	Others	Luck
Success	94	87	68	54	75	66
Failure	73	81	33	42	56	56

(b) In Experiment 2, affects are used as the independent variables to examine whether they can be used as information to infer thought processes (causal attributions).

In both experiments, a major empirical and theoretical goal is to determine whether the emotions linked with internal causal ascriptions cluster together and differ from the emotions associated with external ascriptions. In prior writings (e.g., Weiner et al., 1971) it was contended that the dimension of locus of causality (internal vs. external) mediates between success and failure and the emotional reactions to these outcomes, with affects heightened given feelings of internal responsibility. Weiner et al. (1978) isolated two sources of emotion—outcome and causal attributions—but failed to identify the role of locus of causality as a third determinant of affect (see review in Weiner, 1977).

Experiment 1

Method

The subjects were 79 male and female college students enrolled in introductory psychology at the University of California, Los Angeles, and participating to satisfy a course requirement. They were tested in groups ranging in size from 5 to 16.

Each subject completed a questionnaire in which 12 achievement conditions were represented. Each condition consisted of an outcome (doing well or poorly on a test) determined by one of six causes (ability, unstable effort, stable effort, personality, other people, and luck). These causes were selected because they showed the clearest relationships to different qualities of emotional experience in the Weiner et al. (1978) investigation. The subjects were asked, for example, to

> Think of a time when you did well on a test in a school subject that was very important to you. You felt that the reason you did well was that the teacher (or other people, such as friends) helped you a great deal.

The order of the 12 conditions (two outomes × six causes) was randomly varied between subjects.

The subjects were asked to try to recall a time when they had been in such a situation. If they could recall this type of personal experience, then they were asked for a brief description, including details such as the academic course, when the event occurred, and so on. It was hoped that this prompting would enhance the salience of the event. Finally, the subjects were asked how they felt in the situation, using three affective labels to characterize their emotional reactions. Short spaces for three words were provided on the questionnaire.

Research has indicated that subjects often do not have words readily available to describe their experiences. As an aid to subjects in characterizing and communicating their affective reactions, examples of six affects known from the Weiner et al. (1978) study to be associated with success (calmness, competence, conceit, exhilaration, gratitude, and pride), six affects known to be associated with failure (depression, fear, guilt, incompetence, resignation, and vindictiveness), and one associated with both success and failure (surprise) were included with the instructions. These were listed as "words that describe different types of emotion." Thus, the procedure was a mixture of free and guided recounting of emotional experience.

Subjects worked through the experimental conditions at their own pace. When the questionnaire was completed by all the subjects in a group, the experiment was discussed openly.

Results and Discussion

For all the reported analyses, no differences were found between the sexes; the data are collapsed across this variable. Table 2 shows the percentage of subjects recalling a critical incident in each of the 12 experimental conditions. Subjects reported more success experiences than failure experiences and outcomes associated with ability and effort attributions were also more likely to be recollected. An analysis of variance of these dichotomous data (see Winer, 1971, pp. 303–305) revealed a main effect of outcome,

Table 3
Percentage of Emotional Recollection as a Function of the Causal Attribution for Success

Affect	Ability	Unstable effort	Stable effort	Personality	Others	Luck
Competence	30*	12	20	19	5	2
Confidence	20	19	18	19	14	4
Contentment	4	4	12*	0	7	2
Excitement	3	9	8	11	16*	6
Gratitude	9	1	4	8	43*	14
Guilt	1	3	0	3	2	18*
Happiness	44	43	43	38	46	48
Pride	39*	28	39	43*	21	8
Relief	4	28*	16	11	13	26*
Satisfaction	19	24*	16	14	9	0
Surprise	7	16	4	14	4	52*
Thankfulness	0	1	0	0	18*	4

* $p < .01$.

$F(1, 858 = 39.39, p < .01$, attribution, $F(5, 858) = 21.44, p < .01$, and a significant Outcome × Attribution interaction, $F(5, 858) = 2.34, p < .05$. Within the given attributions, there was a greater likelihood of reporting success than failure in the ability, stable effort, and others conditions (for all conditions, $p < .01$).

The reported differences could indicate the operation of hedonic or defensive biases in attribution or memory or both. However, in the academic lives of these students, the data reported in Table 2 could readily be interpreted as reflecting the frequency of their actual experiences.

Success. The data analyzed for both success and failure outcomes are the percentage of subjects (based on those recounting a critical incident) who reported a specific emotional experience in response to the various achievement conditions. Only emotions listed by more than 10% of the subjects in response to any one of the causal attributions are included in the final analysis.

The 79 subjects listed 156 different emotions in response to the six success attributions. Of these emotions, 12 met the criterion of being given by more than 10% of the subjects for any given attribution. (See Table 3.) Four of these 12 emotions (competence, gratitude, pride, and surprise) were among the examples of emotions provided to the subjects; there was probably lability or influence on which labels subjects were willing to give to their experiences; perhaps a different set of exemplar affects would have yielded other specific emotional reports.

The most commonly reported affect, happiness, appeared in each condition with relatively equal frequency ($M = 44\%$). This response pattern depicts what we have labeled an outcome dependent – attribution independent emotion and, again, suggests that one source of achievement-related affect is linked with the perception of success and failure regardless of the reason for the outcome. Happiness was also one of the general reactions to success reported by Weiner et al. (1978); as in that investigation, the outcome-linked affect was notably dominant in the

Table 4
Discriminating Affects as a Function of Outcome and Causal Attribution

Attribution	Success	Failure
Ability	Competence Pride	Incompetence Resignation Unhappiness
Unstable effort	Relief Satisfaction	Fear
Stable effort	Contentment	Guilt
Personality	Pride	—
Others	Gratitude Thankfulness Excitement	Anger
Luck	Surprise Guilt Relief	Surprise Sadness Stupidity

Table 5
Percentage of Emotional Recollection as a Function of the Causal Attribution of Failure

Affect	Ability	Unstable effort	Stable effort	Personality	Others	Luck
Anger	16	38	33	25	53*	36
Depression	22	20	21	34	10	29
Disappointment	12	17	17	13	20	12
Disgust	2	10	4	3	8	5
Fear	7	25*	21	13	18	12
Frustration	22	18	4	13	20	19
Guilt	9	15	29*	9	3	5
Incompetence	14*	7	8	0	10	0
Mad	7	3	0	3	10	7
Resignation	16*	0	4	9	2	0
Sadness	9	10	0	6	0	14*
Stupidity	14	10	13	9	0	19*
Surprise	2	3	4	0	7	14*
Unhappiness	14*	3	4	9	2	14
Upset	9	10	8	0	8	10

* $p < .01$.

present study. Confidence also was not discriminably linked with any particular affect, although confidence was not a reported affect, given the attribution of luck.

Table 4 lists the discriminating affects for each causal attribution. A discriminating affect is defined as one that is reported significantly more ($p < .01$) for one attribution relative to a composite of the other attributions. All comparisons were done using arc sine transformations, following the procedure suggested by Langer and Abelson (1972).[1] The linkages corresponding to those reported by Weiner et al. (1978; see Table 1) are ability-competence and ability-pride; stable effort – contentment;[2] personality-pride; others-gratitude; and luck-surprise. In addition, there were some associations that did not appear in the prior research: unstable effort – relief and satisfaction (rather than activation); others-excitment; luck-guilt and luck-relief. It is of interest to note that one of the reactions to success caused by luck is a negative affect.

Failure. In response to the six failure situations, 166 different emotional reactions were listed. Fifteen of these emotions were given by more than 10% of the subjects in response to at least one of the causal attributions (see Table 5). Six of these emotions (depression, fear, guilt, incompetence, resignation, and surprise) were included among the examples that were provided for the subjects and, again, apparently indicate their readiness to appropriate given labels. Of these 15 emotions, six (depression, $M = 23\%$; frustration, $M = 16\%$; disappointment, $M = 15\%$; upset, $M = 8\%$; disgust, $M = 5\%$; and mad, $M = 5\%$) were not discriminably associated with any single attribution. Weiner et al. (1978) also reported that frustration and upset were directly linked with failure outcomes. In the present data, three of these affects (depression, frustration, and upset) were reported to be significantly less experienced in response to a particular attribution—others, stable effort, and personality, respectively.

Table 4 also includes the discriminating affects for each of the causal attributions given failure. These discriminating affects

[1] For example, 30% of the subjects reported the emotion of "competent" given an ability attribution for success. In response to the five other attributions, 11% of the subjects reported competent as one of their three affects. Arc sine transformations were performed on these percentage figures and the difference between the arc sines was evaluated relative to an error term based on the number of subjects representing each percentage figure (Langer & Abelson, 1972).

[2] Although stable effort attributions yield the emotion of pride to as great an extent as do ability attributions, the linkage is not significant because the percentage is based on a smaller n.

Table 6
Dimension-Linked Affects as a Function of Outcome and Locus

Present study	Weiner et al.[a]	Present study	Weiner et al.[a]
Success—internal attribution		Success—external attribution	
pride	pride	gratitude	gratitude
competence	competence	thankfulness	thankfulness
confidence	confidence	surprise	modesty
satisfaction	satisfaction	guilt	
	zest		
Failure—internal attribution		Failure—external attribution	
guilt	guilt	anger	anger[b]
resignation	aimlessness	surprise	surprise[c]
	humbleness		others[d]
	regret		

[a] Weiner, Russell, & Lerman, 1978.
[b] Includes 19 affects related to anger (e.g., hostile, infuriated).
[c] Includes 5 affects related to surprise (e.g., astonished, startled).
[d] Seventeen unclassifiable affects (e.g., concerned, alarmed, hysterical).

were determined using the same procedure as for the success condition. The parallel linkages with those reported by Weiner et al. (1978; see Table 1) are ability-incompetence; stable effort – guilt; others-anger; and luck-surprise. The bonds between ability-resignation and ability-unhappiness, and between luck-sad and luck-stupid, were not exhibited in the prior research. However, there was previous evidence in support of the unstable effort – fear association.

Summary. The finding that certain emotions such as happiness and disappointment are independent of attributions but dependent on outcomes corroborates the results reported by Weiner et al. (1978). In addition, of the 12 attribution conditions, the stated emotions in 9 clearly replicate the prior findings. Perhaps what are least clear are the emotional labels given success or failure attributed to the presence or absence of a short-term burst of effort. In addition, in the present study the attribution of failure to personal shortcomings had no discriminably linked affects. But ascriptions of success or failure to ability, stable effort, others, and luck have idiosyncratic affective associates (see Table 4).

Causal dimensions. Further analyses were conducted to discover the relation between causal dimensions and affective responses. The emotional reactions to the four internal causes used in the present study (ability, unstable effort, stable effort, and personality) were compared with the reactions given the two external attributions (other people and luck). That is, the percentage figures from the four internal attributions were averaged and compared with the average of the emotional responses to the two external attributions, using arc sine transformations of the data. Table 6 shows that for success, pride, competence, confidence, and satisfaction, all were more likely to be experienced given internal rather than external attributions. On the other hand, external attributions were more linked with gratitude, thankfulness, surprise, and guilt than were internal ascriptions. (For all conditions, $p < .01$.)

The findings reported above are suspect because only a subset of the possible internal and external attributions were included among the experimental conditions. In Weiner et al. (1978), 10 causal attributions were examined (the present 6 plus mood, intrinsic motivation, task difficulty, and motivation and personality of others). These 10 account for the majority of attributions in achievement-related contexts (Frieze, 1976) and thus permit a better test of whether affects are tied to locus of causality. Reanalyses of the

Weiner et al. data, not previously reported, show a remarkably close correspondence with the present results (see Table 6).

In sum, the affects for success that are most intimately related to personal esteem (competence, confidence, and pride) are reported to be associated with internal attributions. On the other hand, external attributions are linked with affects, such as gratitude and thankfulness, that require an outside causal agent.

Table 6 also shows the dimension-related responses to failure. Again, there are clear correspondences between the two research studies. Internal attributions exacerbate feelings of guilt and what might be considered a giving up of the goal (resigned and aimless), whereas external ascriptions magnify anger and surprise. Guilt, therefore, emerges as an affective reaction for both success and failure, but given external attributions for success and internal ascriptions for failure. These data reaffirm the intertwining of achievement and moral motivational systems (see Weiner & Peter, 1973).

General Discussion

The data from the present study and the Weiner et al. (1978) investigation are sufficiently similar to warrant speculation about the emotional process in achievement-related contexts. We contend that emotions are a product of three distinct cognitive phases. Prior to elaborating this idea, however, it is important that the reader understand that we are going "beyond the data given." First of all, we have no immediate evidence concerning what people actually experience subsequent to success and failure. The experimental paradigm used by Weiner et al. (1978) might be called a simulated other-perception ("How do you think the person would feel?"), whereas in the present investigation the dependent variable is *recollected* emotions. We also cannot definitively state that emotions follow cognitions, because causal attributions for achievement events were not manipulated. In addition, it is possible that the reported cognition–emotion linkages are due to scripts—in this case, rules that guide which emotions should go with which cognitions, although individuals actually do not progress through cognition–emotion phases. A related approach to the data is that we are studying naive theories of emotion, or what a person thinks are appropriate cognition–emotion bonds.

The lack of methodological sophistication in the field of emotion and the limitations of our studies do not allow us to dispel with certainty these alternative interpretations of the data. Nonetheless, there is an account that we find most compelling for the time being. It is suggested that on experiencing an achievement-related outcome, actors first appraise their performance, assigning it a value on a continuum ranging from subjective success to subjective failure. This evaluation is likely to be based on some internal standard and aspiration level, consensus information, and so on. The success–failure judgment produces a positive or negative feeling of a certain intensity. The intensity no doubt is a function of many factors, such as the importance of the outcome, ego involvement, and long-term implications. The positive or negative reactions to success or failure are given labels, such as *happy* or *disappointed*. This emotional response may be the most short-lived and intense of the affective experiences in achievement contexts.

The actor then immediately and automatically or later and reflectively assigns a reason for the outcome. This could be an impulsive statement ("Was I lucky!") or a more reasoned judgment ("I believe I could not have succeeded without the help I received."). As a consequence of this ascription, a number of attribution-specific emotions follow, such as surprise or gratitude. Of course, there may be multiple perceived causality, in which case a number of distinctive emotions should be elicited.

Finally, the attribution(s) for success and failure is classified into causal dimensions, such as locus of causality. Further affective experiences then follow as a consequence of the fact that some of the ascriptions have implications for how one views oneself. Other attributional dimensions, such as stability, also appear to have affective implications (hopelessness vs. optimism; see Weiner et al., 1978; Weiner, 1979). We suspect that these dimension-tied affects have greater longev-

Table 7
Mean Magnitude of Attributional Inference Elicited in Response to the Indicated Emotional Description

Attribution inference elicited	Success—emotional description					
	confident competent pleased (ability)	uproarious delighted good (unstable effort)	calm relaxed secure (stable effort)	hopeful composed safe (task ease)	surprised astonished thankful (luck)	appreciative grateful modest (others)
Ability	8.06	6.25	6.21	6.10	3.10	4.56
Unstable effort	7.08	6.48	7.04	6.12	3.60	5.60
Stable effort	7.02	6.42	6.77	6.73	3.12	5.08
Task ease	4.67	5.75	5.25	5.08	4.65	5.54
Luck	2.60	2.73	4.73	3.94	7.31	6.08
Others	3.54	3.60	4.92	5.02	5.42	6.69

ity than the outcome- or attribution-linked emotions.

In achievement-related contexts, therefore, the actor might progress through various cognition–emotion scenarios, such as (a) "I just received a D on the exam. That is a very low grade." (This generates intense but relatively fleeting feelings of being frustrated and upset.) "I received this grade because I did not try sufficiently hard." (This is followed by feelings of guilt.) "There really is something lacking in me." (This is ensued by low self-esteem or lack of worth.) "What I lack I probably always will lack." (This produces hopelessness.)

Alternatively, there is (b) "I just received an A on the exam. That is a very high grade." (This generates happiness.) "I received this grade because I worked very hard during the entire school year." (This produces contentment and relaxation.) "I really do have some positive qualities that will persist in the future." (This is followed by high self-esteem, feelings of self-worth, and optimism.)

Experiment 2

Since emotions are coupled with cognitions, an emotional expression can act as a cue to an observer, allowing an inference about the cognitions of the actor. Hence, publicly masking our emotions often serves the function of hiding our thoughts, just as public emotional expression reveals our thoughts. In Experiment 2, we examined whether knowledge about an actor's emotions (here conveyed with verbal labels) enables an observer to infer an actor's causal attributions for an achievement performance.

Method

Six attributions used by Weiner et al. (1978) provided the foundation for the present investigation: ability, unstable effort, stable effort, task difficulty, others, and luck. Task difficulty, therefore, has replaced personality among the causal ascriptions varied in Experiment 1.

Two criteria were used to select the relevant affects associated with each attribution. Highly discriminating affects not intensely experienced might be so subordinate among the affects being expressed that they could not have a cue function. On the other hand, emotions strongly felt (and expressed) for all attributions cannot possibly have signal value for any one ascription. Hence, selection of the affects for each attribution was based on two criteria: the discriminating value of the affect multiplied by its absolute intensity rating (data from Weiner et al., 1978). Three emotions were selected for each of the six attributions; the affects chosen for this study had the highest values using this procedure. (See Tables 7 and 8.)

The subjects were 48 male and female students at the University of California, Los Angeles, participating as part of a course requirement. Twelve scenarios (two levels of outcome × six emotional descriptions) were randomly presented to each subject. A typical story scenario follows.

> A person just received a test back in a course that is very important to him or her. He or she has done very well and feels extremely *surprised*, *astonished*, and *thankful*. Why did this person believe that he or she did so well?

Table 8
Mean Magnitude of Attribution Inference Elicited in Response to the Indicated Emotional Description

	Failure—emotional description					
Attribution inference elicited	incompetent inadequate panicked (ability)	sorry ashamed scared (unstable effort)	humble guilty troubled (stable effort)	stunned sad displeased (task difficulty)	astonished overwhelmed surprised (luck)	bitter furious revengeful (others)
Ability	7.40	5.50	4.56	4.08	3.44	3.35
Unstable effort	4.92	7.48	7.40	3.54	2.73	3.67
Stable effort	5.46	7.56	7.81	3.94	2.90	3.81
Task difficulty	6.35	4.56	4.94	5.56	4.94	6.70
Luck	3.25	3.10	2.85	4.96	6.52	4.85
Others	3.48	3.60	3.58	4.79	3.85	6.75

This was followed by six scales anchored at the extremes with "definitely" and "definitely not," on which the judgments for the six causal ascriptions were made. The scales were divided into nine equal intervals and were randomized both within and between subjects.

Results and Discussion

Tables 7 and 8 show the mean ratings for each emotional description on the six attribution scales in the success and failure conditions, respectively. Inspection of the columns of the tables indicates the extent to which a particular emotional description is associated with the attributional judgments. If the emotion–cognition linkages and the cognition–emotion linkages are symmetrical, then the diagonals in the tables should contain the highest ratings within the columns. For example, the emotions known to be elicited by luck attributions (surprised, astonished, and thankful) should, in turn, lead to an inference of luck as the causal determinant. In the fifth column of Tables 7 and 8, the fifth row should (and does) contain the highest rating. In 9 of the 12 comparisons ($p < .01$) the a priori designated attribution is most highly rated, whereas in two other instances (stable effort emotions for success and unstable effort emotions for failure) the ratings are second highest. The one disconfirmatory finding involves task attributions for success, which primarily elicit the inference that stable effort was the cause of the positive outcome. However, the data in the Weiner et al. (1978) investigation were not very systematic for task attributions; for that reason, the task ascription was excluded from Experiment 1.

Statistical comparisons were performed contrasting the a priori "correct" attributional inference (that is, the attribution that was found to be associated with the emotional description in the Weiner et al., 1978, investigation) to the inferences given the five other attributions. Significant differences ($p < .01$) in the expected direction were found for all the emotional descriptions, except in the case of the emotions associated with task ease.

Causal Dimensions

Three of the attributions used in this study are classified as internal in locus of causality (ability, unstable effort, and stable effort) and three are external (task difficulty, luck, and others). The strength with which internal versus external causal attributions were inferred for each emotional description was then analyzed. In 11 of 12 comparisons, emotional descriptions associated with a particular internal attribution according to Weiner et al. (1978) elicited internal causal inferences significantly more than external causes, with the opposite being true of emotional descriptions associated with external attributions. (For all conditions, $p < .01$.) For example, the emotional description of confident, competent, and pleased, which was predicted to elicit ability

attributions, elicited ability, unstable effort, and stable effort, which are the internal attributions, to a greater extent than inferences that the external attributions of task ease, luck, and others were responsible for success. The lone exception to this pattern was again the emotional description associated with task explanations for success, in which internal attributions were more strongly inferred than external attributions. These data indicate that there is not only a strong inference of the a priori "correct" attribution given an emotional description, but one also infers attributions similar in locus of causality.

These data may not seem so surprising considering that two of the internal causal ascriptions involve effort. However, the external attributions are disparate and the emotional descriptions are extremely varied. There is an intimation that observers might engage in a two-step inference process, in which first the emotions are linked to a particular causal dimension and then there is a search for the particular cause. But at present it is not known whether specific causal judgments follow or precede dimensional placement or whether there is no systematic judgment sequence.

References

Davitz, J. R. *The language of emotion*. New York: Academic Press, 1969.

Elig, T. W., & Frieze, I. H. A multi-dimensional scheme for coding and interpreting perceived causality for success and failure events: The Coding Scheme of Perceived Causality (CSPC). JSAS, *Catalog of Selected Documents in Psychology*, 1975, *5*, 313. (Ms. No. 1069)

Frieze, I. H. Causal attributions and information seeking to explain success and failure. *Journal of Research in Personality*, 1976, *10*, 293–305.

Langer, E. J., & Abelson, R. P. The semantics of asking a favor: How to succeed in getting help without really dying. *Journal of Personality and Social Psychology*, 1972, *24*, 26–32.

Lazarus, R. S. *Psychological stress and the coping process*. New York: McGraw-Hill, 1966.

McClelland, D. C. *Assessing human motivation*. Morristown, N.J.: General Learning Press, 1971.

Schachter, S. The interaction of cognitive and physiological determinants of emotional state. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 1). New York: Academic Press, 1964.

Weiner, B. Attribution and affect: Comments on Sohn's critique. *Journal of Educational Psychology*, 1977, *69*, 506–511.

Weiner, B. A theory of motivation for some classroom experiences. *Journal of Educational Psychology*, 1979, *71*, 3–25.

Weiner, B., et al. *Perceiving the causes of success and failure*. Morristown, N.J.: General Learning Press, 1971.

Weiner, B., & Peter, N. A cognitive-developmental analysis of achievement and moral judgment. *Developmental Psychology*, 1973, *9*, 290–309.

Weiner, B., Russell, D., & Lerman, D. Affective consequences of causal ascriptions. In J. H. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research* (Vol. 2). Hillsdale, N.J.: Erlbaum, 1978.

Winer, B. J. *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill, 1971.

Received August 14, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 7, 1221–1229

Instrumentality Effects in the Assessment of Racial Differences in Self-Esteem

Bernadette Gray-Little and Mark I. Appelbaum
University of North Carolina at Chapel Hill

Previous studies of racial differences in self-esteem have led to highly disparate and confusing results. This study shows that the effects of the measuring instrument as well as those of preexisting individual differences in academic and demographic characteristics are very great. These effects may be sufficient, in many cases, to explain the disparate results of earlier studies.

Attempts to measure racial differences in self-esteem, the evaluative component of the self-concept, have probably been more numerous than social comparisons of any other psychological variables except intelligence and school performance. Much of this research has proceeded on the hypothesis that blacks have degraded self-images and make reduced self-evaluations compared to whites. In many instances, this assumption is explicitly anchored in the self-theory of James (1890) or Mead (1934), both of whom linked self-concept to the evaluative feedback that one receives from important others in the social environment. Although there is substantial empirical support for the relationship between self-evaluations and appraisals from others (Wylie, 1968), such a relationship cannot be offered as conclusive proof of lower self-esteem in blacks. To assume that blacks have low self-esteem, it is also necessary to assume that an individual's self-concept is heavily dependent on society's evaluation of the group with which a person is identified (Cartwright, 1950; McCandless, 1970). Accordingly, since blacks as a group are less positively evaluated than whites, black individuals will have lower self-esteem than whites. There has been some support for this conclusion (Bridgette, 1970; Long & Henderson, 1968), but just as often no support (Gibby and Gabler, 1969; Rosenberg, 1965) or directly opposing results (Baughman & Dahlstrom, 1968; McDonald & Gynther, 1965; Powell, 1973; Rosenberg & Simmons, 1971; Wendland, 1969).

There are probably many reasons for the contradictory findings in this area of research. First, the assumption of correlation between one's self-esteem and society's evaluation of one's racial or ethnic group, although compelling in terms of common sense, has little empirical support. Rosenberg (1965) called this proposition the stratification hypothesis and found no relationship ($r = .04$) between an ethnic group's average self-esteem score and independent ratings of that group's status in society.

Second, the study of self-esteem has been plagued by a number of problems that make racial comparisons even more difficult. For example, self-esteem has been defined and measured in different and, at times, noncomparable ways (Wylie, 1968). Furthermore, subject characteristics such as sex, age, and grade have typically varied from one study to the next, with little attention paid to the relevance of these factors to racial differences found (Christmas, 1973). Moreover, in many instances important mediating factors, such as socioeconomic status and educational achievement, which are probably relevant to

This research was partially supported by grants to the first author from the University Research Council of the University of North Carolina and the Social Science Research Council's Committee on Grants for Research on Racism and other social factors in mental health.

We are grateful to E. Earl Baughman for his helpful critique of an earlier draft of this article.

Requests for reprints should be sent to Bernadette Gray-Little, Department of Psychology, University of North Carolina, Chapel Hill, North Carolina 27514.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3707-1221$00.75

self-concept and which are certainly pertinent to discussions of racial differences, have been neither controlled nor covaried.

The problems attendant on the confounding of socioeconomic status with racial/ethnic membership are illustrated in studies by Long and Henderson (1968) and Soares and Soares (1969). Long and Henderson found economically disadvantaged school beginners in a Southern community to have lower self-esteem scores than more advantaged children on the Self-Social Constructs Test. Soares and Soares asked fourth-grade through eight-grade children to rate themselves on a set of 20 bipolar trait descriptors and found that disadvantaged children had higher self-esteem scores than advantaged children. In the Long and Henderson study, all of the disadvantaged children were black and all of the advantaged children were white. In the Soares and Soares study, $\frac{2}{3}$ of the disadvantaged children were black or Puerto Rican whereas 90% of the advantaged children were white. Neither study made clear whether socioeconomic or racial factors account for the differences in self-esteem scores. Studies that have looked at the relationship between self-esteem and academic achievement have been somewhat more consistent in showing a positive relationship between educational achievement and self-concept scores (Freyberg & Shapiro, 1966; Caplin, 1969). There have, however, been investigations that revealed no differences in the self-concept scores of high-achieving and low-achieving students (Curtis, 1967; Lourenso, Greenberg, & Davidson, 1965).

Two studies that are particularly relevant to the present one and that highlighted the problems encountered in racial comparisons of self-esteem were those of Wendland (1969) and Bridgette (1970). These studies involved children in the same geographical area and, in some instances, in the same school system. Wendland found blacks to have higher self-esteem scores, whereas Bridgette reported higher self-esteem scores for whites. One obvious difference between the studies was that Wendland's subjects were attending segregated schools whereas Bridgette's subjects were in a recently desegregated school in which whites (who were the majority) had vigorously opposed integration. Another potentially relevant difference was the grade level and hence the age of the subjects tested—sixth-grade students in Wendland's study and eleventh-grade students in Bridgette's research. Since there may be different age trends in self-esteem for blacks and whites, it is unclear whether or not the contradictory findings were related to age. Perhaps most important, however, two different measures of self-esteem were used—Fitt's Tennessee Self-Concept Scale (TSCS) in Wendland's research and Coopersmith's Self-Esteem Inventory (SEI) in Bridgette's work. One might reasonably have expected the same results from these two instruments, since they purport to measure the same aspect of self-concept—self-esteem. Coopersmith (1967) described the SEI as a measure of "the evaluation an individual makes and customarily maintains with regard to himself . . . a personal judgment of worthiness" (p. 4). Similarly, Fitts (1965) characterized those who receive high scores on the Total Positive Scale, the most important subscale of the TSCS, as "persons who tend to like themselves, feel they are persons of value and worth" (p. 2).

There are, then, several factors, both theoretical and methodological, that might be important in explaining why some studies have shown blacks to have higher self-esteem scores than whites, whereas others have found whites to have higher scores. Among the methodological issues, one should consider the possibility that seeming racial differences may be attributable to differences in the socioeconomic status or academic achievement of the groups tested or to differences in the racial make-up, segregated or desegregated, of the school. Racial differences may also be artifactual.

The primary purpose of this study is to examine the extent to which the discrepant findings with regard to race are due to the instruments used to measure self-esteem. First, we focus on the race effects obtained on two instruments—the SEI and the TSCS—for junior high and high school students. Then, we examine the correlations of each instrument with important intervening socioeconomic and academic variables and observe whether the pattern of racial differences found with these

measures persists when the correlated variables are covaried.

There were several reasons for selecting these instruments. Both are widely used to measure self-esteem in school-age children. Both are intended to measure the evaluative component of self-concept and both are rating scales. Norms and reliability data are available for both instruments. Finally, because we had access to subjects from the same school systems used by Bridgette (1970) and Wendland (1969) in their investigations, we wished to use the same measures to enhance comparability with their data.

Method

The school system in Central North Carolina, from which our 735 subjects were drawn, is now desegregated. Small farms and textile mills are among the major employers in this area; most of the children live in a small town and its surrounding rural area. Our sample was composed of 146 black males, 145 black females, 238 white males, and 206 white females, with seventh-grade students accounting for roughly 54% of the sample, and the remainder being tenth-graders. The mean age of the seventh-grade students was 12.31 (SD = .68) and of the tenth-grade students, 15.33 (SD = .61). Half of the students at each grade level were tested with the SEI and the other half with the TSCS, yielding 372 SEI scores and 363 TSCS scores. Two female examiners, one black and one white, administered the tests to classroom groups. IQ (Form L of the Otis-Lennon Intelligence Test) and achievement scores (grade-appropriate forms of the Iowa Test of Basic Skills), as well as information regarding the number of children in the family and parents' educational and occupational levels, were obtained from the school records of the tested students. The achievement tests had been administered the previous semester. IQ scores were typically not more than 1 year old. Hollingshead's Two-Factor Index of Social Position (Note 1) was used to classify subjects' socioeconomic status.[1] A score of 1 indicates the highest social class; 5 indicates the lowest.

Two years after the initial testing with the self-esteem instruments, the original seventh-grade students were retested as ninth graders. The purpose of this testing was to allow us to examine changes in self-esteem over time as a function of race and sex and to compare the stability of the SEI scores with those of the TSCS scores. Of the 256 students that were available at the time of the follow-up, 128 had originally been tested with the SEI and 128 with the TSCS. A comparison of each of the two follow-up groups with the total group of subjects who had taken the same test showed them to be very similar. The follow-up SEI group had an average seventh-grade score of 65.44, whereas the total seventh-grade group had an average score of 64.99. Similarly, the follow-up TSCS group had an average seventh-grade score of 329.79, whereas the total seventh-grade group had an average TSCS score of 329.19. The scores for the original and follow-up samples tested with the TSCS are very similar to those found for junior and senior high school students in several other studies (For a review, see Thompson, 1972). The scores obtained by the SEI standardization sample (Coopersmith, 1967) were slightly higher than those obtained by our original and follow-up groups tested with the SEI.

Table 1
Analyses of Variance for Age, Race, and Sex Differences in Self-concept

Variable	SEI[a] (1, 364)	TSCS[b] (1, 355)
Age (A)	1.79	3.68*
Race (B)	11.81**	1.10
Sex (C)	.64	.40
A × B	.23	.54
A × C	.16	.01
B × C	.17	2.33
A × B × C	.02	.28

Note. Tests are those of effects eliminating other effects at the same level (e.g., age eliminating race and sex). See Appelbaum and Cramer (1974). Numbers in parentheses are degrees of freedom.
[a] SEI = Self-Esteem Inventory.
[b] TSCS = Tennessee Self-Concept Scale.
* $p < .056$.
** $p < .001$.

Results

In the following section, we present analyses of the SEI and TSCS in terms of race and age differences as well as the relation of these tests to demographic and academic variables. Initial analyses were on a three-factor (race, age, and sex) design. As indicated in Table 1, however, there were no significant two-way or three-way interactions and no significant sex effects. Therefore, results are presented only for the main effects of race and age. For each test, we examine raw differences between groups as well as associated differences on demographic and academic variables. Given the intact groups used (racial and age groups) we suspected that there would be substantial pre-existing group differences on demographic

[1] Social Position is based on a combination of weighted educational and occupational levels. Father's occupation and education were used to determine social position. When for any reason this information was missing, the mother's occupation and education were used.

Table 2
Racial Differences on the Self-Esteem Inventory (SEI) and the Tennessee Self-Concept Scale (TSCS): Demographic and Academic Variables

Variable	M		F (df)
	Black	White	
SEI data set			
SEI	60.72	66.21	11.81 (1, 364)**
Mother's education	10.57	11.28	6.41 (1, 317)*
Father's education	9.16	10.80	19.14 (1, 288)**
Social position	4.62	4.17	23.69 (1, 364)**
No. children	5.89	3.65	65.35 (1, 345)**
Achievement	5.71	7.34	93.37 (1, 285)**
IQ	84.04	98.79	112.26 (1, 333)**
TSCS data set			
TSCS	323.09	327.38	1.10 (1, 355)
Mother's education	10.60	11.25	4.96 (1, 314)*
Father's education	9.27	10.69	14.85 (1, 297)**
Social position	4.50	4.09	22.02 (1, 355)**
No. children	5.45	3.71	52.08 (1, 344)**
Achievement	6.12	7.53	98.44 (1, 320)**
IQ	86.77	101.19	98.98 (1, 345)**

Note. Varying degrees of freedom associated with the denominator reflect differences in the completeness of data from school records.
* $p < .05$.
** $p < .001$.

variables that might be logically related to SEI and TSCS differences. Although we are aware of the limitations of adjustment procedures with intact groups, we have presented estimates of group differences (and tests of significance) adjusted for covariates. Readers are advised to regard these results with caution, bearing in mind that it is very difficult to judge the quality of the adjusted scores under these circumstances.

Racial Differences

SEI. Table 2 contains group means and *F* values for examining the main effects of race. The upper half of Table 2 contains raw SEI scores and academic and demographic information collected on the SEI groups. As can easily be seen, white subjects received higher self-esteem scores than black subjects. Indeed, black and white subjects differed significantly on all variables, with white students having higher parental education and social position, higher achievement and IQ scores, and fewer siblings. Intercorrelations of these variables with SEI scores are given in Table 3. The intercorrelations of these variables with the TSCS are also presented in Table 3. All of these variables, with the exception of father's occupation, correlate significantly with SEI scores. Given the initial racial differences on these variables and their significant intercorrelations with the SEI, we attempted to obtain estimates of racial differences in the SEI adjusted for these variables through the use of

Table 3
Pearson Product-Moment Correlations of Self-Concept Measures With Demographic and Academic Variables

Variable	SEI (r)	TSCS (r)
Mother's education	.199**	.129
Father's education	.152*	.166*
Father's occupation	.072	−.024
No. children	−.176**	−.082
Achievement	.259**	.202**
IQ	.208**	.191**

* $p < .05$.
** $p < .001$.

Table 4
Racial Differences in Self-Esteem Inventory (SEI) and Tennessee Self-Concept Scale (TSCS), Adjusted for Covariates

Item	Difference (black − white)	*F* (*df*)
SEI data set		
Raw SEI differences	−5.49	11.81 (1, 364)*
Differences adjusted for		
Mother's education	−4.65	7.61 (1, 316)*
Father's education	−4.41	5.85 (1, 287)*
Social position	−4.87	8.86 (1, 363)*
No. children	−3.01	3.01 (1, 344)
Achievement	−1.46	.95 (1, 284)
IQ	− .93	.40 (1, 332)
TSCS data set		
Raw TSCS Differences	−4.29	1.10 (1, 355)
Differences adjusted for		
Mother's education	−3.25	.44 (1, 313)
Father's education	− .63	.03 (1, 296)
Social position	−2.14	.23 (1, 354)
No. children	−3.78	.68 (1, 343)
Achievement	2.52	.56 (1, 319)
IQ	3.79	.91 (1, 344)

* $p < .01$.

an analysis of covariance. The results of these adjustment analyses are given in the upper half of Table 4. Keeping our earlier proviso in mind, we found that each of these adjustments results in some attenuation of the racial difference in SEI scores. White students, nevertheless, continue to have significantly higher SEI scores when social status or the educational level of either parent is covaried. However, the black-white difference on the SEI is virtually eliminated when either achievement or IQ scores are covaried. Likewise, controlling for the number of children in a student's family substantially reduces the racial difference.

TSCS. A somewhat different pattern of relationships emerges when one examines racial differences in self-concept as measured by the Total Positive Score from the TSCS. These results are presented in the lower half of Table 2. Perhaps most important, there is not a significant difference in the TSCS scores obtained by black and white students, even though this group of white students also comes from significantly more advantaged families, has fewer siblings, and higher achievement and IQ scores. A reinspection of Table 3 reveals that both academic variables are significantly related to the TSCS; however, unlike the SEI, the TSCS is significantly correlated with only one of the demographic variables, father's education. The results of the covariance analyses are presented in the lower half of Table 4. Covarying any of the demographic variables results in a decrease in the racial difference; covarying either of the academic variables results in a reversal of the difference so that the scores of black students are nonsignificantly higher than those of white students.

These two measures, then, yield different results when administered to unmatched groups of black and white students. The SEI reveals racial differences; the TSCS does not. One of the reasons for this difference appears to be the fact that the SEI is correlated with a larger number of race-sensitive demographic variables than is the TSCS. Both measures are consistent, however, in showing the major impact that academic variables, such as achievement and IQ scores, have on self-esteem. As suggested in Table 4, when black and white students are matched for either of these variables, their self-esteem scores do not differ. The results of the initial analysis with the

Table 5
Grade Differences in Self-Esteem Inventory (SEI) and Tennessee Self-Concept Scale (TSCS) Demographic and Academic Variables

Variable	*M*		*F* (*df*)
	Grade 7	Grade 10	
	SEI data set		
SEI	64.99	62.84	1.79 (1, 364)
Mother's education	11.04	10.97	.07 (1, 317)
Father's education	10.25	10.17	.05 (1, 288)
Social position	4.45	4.22	6.61 (1, 364)*
No. children	4.70	4.32	2.01 (1, 345)
Achievement	5.69	8.01	228.08 (1, 285)**
IQ	91.04	95.47	10.99 (1, 333)**
	TSCS data set		
TSCS	329.19	321.91	3.68 (1, 355)*
Mother's education	11.17	10.82	1.63 (1, 314)
Father's education	10.48	9.86	3.18 (1, 297)
Social position	4.27	4.23	.19 (1, 355)
No. children	4.69	4.11	5.83 (1, 344)*
Achievement	5.96	8.07	195.61 (1, 320)**
IQ	95.40	95.29	.01 (1, 345)

* $p < .05$.
** $p < .001$.

SEI are consistent with those of Bridgette (1970), who found that white students have significantly higher self-esteem scores than black students. We did not, however, replicate Wendland's (1969) findings of higher TSCS scores for black students. The present results represent a 12-point decrease in TSCS scores for black students and a 4-point increase for white students when compared to Wendland's subjects. An explanation of our success in replicating Bridgette's findings but not Wendland's may be that Wendland tested children while schools were still segregated, whereas Bridgette tested them after desegregation had begun.

Grade Differences

SEI. As indicated in the upper half of Table 5, the mean self-esteem score of seventh-grade subjects was slightly higher than that of the tenth-grade group, but the difference is not significant. The two age groups are largely comparable on demographic variables but, as one might expect, the tenth-grade group had a significantly higher achievement level. The tenth-grade group also had higher IQ scores. When the SEI scores are adjusted for each of the demographic variables, the small age difference in SEI remains basically unchanged. Adjusting SEI scores for achievement level, $F(1, 284) = 12.84$, $p < .001$, or for IQ, $F(1, 332) = 3.99$, $p < .05$, does result in significantly higher SEI scores for the seventh-grade group.

TSCS. The scores received by the seventh-grade subjects on the TSCS are significantly higher than those received by older students (see the lower half of Table 5). These two age groups are comparable on all demographic variables except for the number of children in the family, which is higher for the younger group. The achievement level, but not IQ, of the older subjects is higher. Adjusting TSCS scores for parental education or number of children in the family reduces the age difference so that it is no longer significant. The age difference is increased when TSCS scores are adjusted for social position, $F(1, 354) = 3.92$, $p < .05$. And, as in the case of the SEI, adjusting TSCS scores for achievement, $F(1, 219 = 12.77$, $p < .001$, or IQ, $F(1, 344) = 4.42$, $p < .05$, enhances the age effect.

Table 6
Mean Difference Scores (Grade 9 — Grade 7 Scores)

	SEI			TSCS		
	n	*M* Difference	*SD*	*n*	*M* Difference	*SD*
Black males	27	1.11	12.23	20	−5.90	21.84
Black females	31	−.29	14.50	28	−8.86	33.69
White males	42	−.09	16.82	41	−2.63	33.94
White females	28	1.78	13.70	39	−7.21	35.68

Changes in Self-Esteem

The primary purpose of the following analyses was to assess patterns of change in self-esteem as a function of race and sex. Changes in self-esteem were examined in two ways. First, the difference scores (Grade 9 scores − Grade 7 scores) were examined by means of an analysis of variance in which race and sex were the factors. Second, we performed a similar analysis but used the original seventh-grade score as a covariate to control the possible impact that the level of the initial score might have had on the amount of change.

The mean difference scores for both the SEI and TSCS subjects are presented in Table 6. In examining the scores, one should bear in mind that the possible range of scores on the SEI is relatively small (50 to 100), whereas on the TSCS, scores can range from 100 to 500. Changes in SEI scores were small and inconsistent from one race/sex group to another. Changes in TSCS scores were also small, but were consistent in showing a decline from the seventh grade to the ninth grade. The earlier comparison of these subjects as seventh graders with tenth graders showed the seventh-grade students to have significantly higher scores. The present data also suggest that one can expect a decline in TSCS scores from junior to senior high school.

The analyses of variance revealed no main effects for race or sex nor any interaction effects on either the SEI or TSCS difference scores. Furthermore, the covariance analysis for each test was consistent in showing neither race nor sex differences. Finally, the pooled regression for the four race/sex groups was highly significant ($p < .001$) for each of the tests, indicating that the ninth-grade scores were very predictable from seventh-grade scores in all instances.

Discussion

The major conclusion of this study is that it is possible to ensure or preclude findings of racial or age differences in self-esteem through test selection or control of relevant demographic and academic variables. Although it is true that tests such as the SEI and the TSCS purport to measure the same phenomenon, it is also clear that they are differentially related to a number of factors that affect self-esteem. A casual inspection of these two tests discloses only that the SEI has the advantage of being shorter and the TSCS the advantage of being longer. Further examination of the tests reveals them to be similar in content with both tests sampling from a wide range of self-referent statements that are typically assumed to be important in describing the self-concept. The SEI is, however, more highly related to a number of demographic factors and reveals racial differences. The TSCS shows weaker relationships to these variables and does not reveal racial differences. The problem is to determine which self-esteem measure is to be preferred, especially when one is making racial comparisons.

In considering the relative merits of these two instruments, a decision as to which is better cannot be based on a strict judgment of which has greater construct validity, since there is no uncontested, external criterion of self-esteem. Rather, the choice will have to be made on the basis of the network of relationships that exists between each of these tests and other demographic and experiential values.

One might easily argue that the test that appears most culture-free, that has a weaker correlation with conventional indicators of status and social position, should be preferred. This argument would lead to the selection of the TSCS. One might also argue the inverse—that because the SEI is more highly correlated with other conventional measures of status and position, it will be more predictive of success and should, therefore, be preferred. The latter, culture-bound argument implies that one's self-esteem should be, for example, a reflection of socioeconomic status. Although it is reasonable that factors such as parental status or school achievement might contribute to a child's self-esteem, many self-concept theorists and researchers seem to work under the assumption that self-esteem is an important feature of personality and should, therefore, add to the information derived from demographic or academic data (see Block, 1971).

The many problems that plague this area of research point to the need to incorporate a number of methodological and conceptual refinements in future studies of self-concept. We have already alluded to the necessity for care in test selection and control of subject characteristics; studies such as that of Long and Henderson (1968), which compare disadvantaged black subjects with advantaged white subjects, contribute little to the understanding of racial or social class differences in self-esteem. Furthermore, in reviewing studies relevant to the present one (that is, studies that make direct racial comparisons of self-esteem), it becomes clear that not only test and subject characteristics but also contextual variables need to be considered. Specifically, the racial make-up of the school may be an important determinant of the direction of racial differences in self-esteem. Most of the studies (including the present one) that report higher self-esteem scores for white students or no racial differences were completed in integrated settings (Bridgette, 1970; Rosenberg, 1965; Wylie & Hutchins, 1967). Studies that report more positive self-concept scores for black children seem to have been completed in either segregated schools (Baughman & Dahlstrom, 1968; McDonald & Gynther, 1965; Wendland, 1969) or desegregated schools in which blacks were a majority (Powell, 1973; Rosenberg & Simmons, 1971; St. John, 1975; Soares & Soares, 1969).

There is also a need for greater clarity in the definition of self-esteem. This is a particularly difficult problem. The concept of self-esteem has great appeal and is evoked to explain a host of complex behavioral and personality characteristics. It is unlikely that any one of the instruments typically used to measure self-esteem will adequately encompass this complexity. What promise to be more feasible alternatives are greater articulation of and more specificity in the facet of self-concept being measured. Watkins (1978), for example, suggested an approach to self-esteem that attempted to take into account the value system of the person tested. Subjects were asked to rate themselves in various areas of life—for example, social, family, and general happiness—and the responses were weighted by the relative importance of these areas to the subject. Watkins suggested that this procedure might easily be tailored to the particular social or ethnic group under consideration. Wylie and Hutchins (1967), on the other hand, focused on a specific content area, academic self-concept, in testing groups of black and white adolescents whom they had attempted to equate on socioeconomic status and academic ability.

In addition to delimiting self-concept in the preceding and other ways, we also need to be clear as to whether our intention is to measure the individual's self-concept or the individual's pride in his or her race. Although these two ideas may be interdependent, there is no need to assume that they are equivalent either to one another or to the status one's racial or ethnic group has in society (Rosenberg, 1965).

Finally, the truism that intragroup variations are frequently greater than intergroup differences bears periodic repetition for those studying racial differences in self-esteem. Although this does not mean that racial comparisons should not be made, it does mean that global racial comparison will be of limited utility. If one is primarily interested in studying the plight of blacks and other minorities, then it is eminently more useful to investigate the causes and prevention of racism, poverty, and discrimination. If, on the other hand, one is primarily interested in the antecedents of

self-esteem, then another approach is desirable—one that focuses on variations in self-esteem of subjects who are equated on other characteristics that are known to be relevant. Studies of racial differences in self-esteem are not instructive either with regard to the nature of self-concept or racial differences in personality when the groups initially differ in ways that are confounded with race.

Reference Note

1. Hollingshead, A. B. Two-factor index of Social Position. Unpublished manuscript, Yale University, 1957.

References

Appelbaum, M. I., & Cramer, E. M. Some problems in nonorthogonal analysis of variance. *Psychological Bulletin*, 1974, *81*, 335–343.

Baughman, E. E., & Dahlstrom, W. G. *Negro and white children: A study of the rural South*. New York: Academic Press, 1968.

Block, J. *Lives through time*. Berkeley, California: Bancroft Books, 1971.

Bridgette, R. E. Self-esteem in Negro and white southern adolescents. (Doctoral dissertation, University of North Carolina at Chapel Hill, 1970), *Dissertation Abstracts International*, 1970, *31*, 2977B.

Caplin, M. D. The relationship between self-concept and academic achievement. *The Journal of Experimental Education*, 1969, *37*, 13–16.

Cartwright, D. Emotional dimensions of group life. In N. L. Reymert (Ed.), *Feelings and emotions*. New York: McGraw-Hill, 1950.

Christmas, J. J. Self-concept and attitudes. In K. S. Miller and R. M. Dreger (Eds.), *Comparative studies of blacks and whites in the United States*. New York: Seminar Press, 1973, 249–273.

Coopersmith, S. *The antecedents of self-esteem*. San Francisco: Freeman, 1967.

Curtis, L. Digest of research studies on self-concept (1963–1967). *Graduate Research in Education and Related Disciplines*, 1967, *3*, 82–88.

Fitts, W. H. *Manual for the Tennessee Self Concept Scale*. Nashville: Counselor Recordings and Tests, 1965.

Freyberg, J., & Shapiro, N. Digest of research studies on the relation of socioeconomic status to personality variables (1962–1965). *Graduate Research in Education and Related Disciplines*, 1966, *2*, 102–107.

Gibby, R., & Gabler, R. The self-concept of Negro and white children. *Journal of Clinical Psychology*, 1969, *23*, 144–148.

James, W. *Principles of psychology*. New York: Holt, 1890.

Long, B., & Henderson, E. Self-social concepts of disadvantaged school beginners. *Journal of Genetic Psychology*, 1968, *113*, 41–51.

Lourenso, S. V., Greenberg, J. W., & Davidson, H. H. Personality characteristics revealed in drawings of deprived children who differ in scholastic achievement. *Journal of Educational Research*, 1965, *59*, 63–67.

McCandless, B. *Adolescents: Behavior and development*. Hinsdale, Ill.: Dryden Press, 1970.

McDonald, R. L., & Gynther, M. D. Relationship of self and ideal-self descriptions with sex, race, and class in southern adolescents. *Journal of Personality and Social Psychology*, 1965, *1*, 85–88.

Mead, G. H. *Mind, self, society*. Chicago: University of Chicago Press, 1934.

Powell, G. *Black Monday's children*. New York: Appleton-Century-Crofts, 1973.

Rosenberg, M. *Society and the adolescent self-image*. Princeton, N. J.: Princeton University Press, 1965.

Rosenberg, M., & Simmons, R. G. *Black and white self-esteem: The urban school child*. Washington, D. C.: American Sociological Association, 1971.

Soares, A. T., & Soares, L. M. Self-perceptions of culturally disadvantaged children. *American Educational Research Journal*, 1969, *6*, 31–45.

St. John, N. *School desegregation: Outcomes for children*. New York: Wiley, 1975.

Thompson, W. Correlates of self-concept. *Dede Wallace Center Monograph* (No. 6). Nashville, Tenn., 1972.

Watkins, D. The development and evaluation of self-esteem measuring instruments. *Journal of Personality Assessment*, 1978, *42* (2), 171–182.

Wendland, M. M. Self-concept in southern Negro and white adolescents as related to urban-rural residence. (Doctoral dissertation, University of North Carolina at Chapel Hill, 1969), *Dissertation Abstracts*, 1969, *29*, 2642B.

Wylie, R. C. The present status of self theory. In E. F. Borgatta and W. W. Lambert (Eds.), *Handbook of personality theory and research*. Chicago: Rand McNally, 1968.

Wylie, R. C., & Hutchins, B. Schoolwork ability estimates and aspirations as a function of socioeconomic level, race, and sex. *Psychological Reports*, 1967, *21*, 781–808.

Received September 1, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 7, 1230-1238

Observer Bias in the Attitude Attribution Paradigm: Effect of Time and Information Order

Edward E. Jones, Janet Morgan Riggs, and George Quattrone
Princeton University

Subjects were exposed to an essay taking either a pro or anti stand on the use of minority quotas. They were informed either before or after reading the essay that the writer had or had not been permitted to choose which side of the issue to support. Subjects returned for a second session a week after the first, and their attributions of true attitude to the essayist were again measured. The results replicated previous attitude attribution findings, but the tendency to attribute an essay-consistent attitude when the essay was written under high constraint was much greater when the constraint information appeared *after* exposure to the essay. There was no hint of a sleeper effect and, therefore, no evidence that constraint cues were discounted over time.

The attitude attribution paradigm—diagnosing "true attitudes" from opinions expressed under varying situational pressures—has provided consistent evidence of a tendency to assume correspondence between the expressed opinions and underlying attitudes of a target person acting under high constraint. The data are summarized and discussed in a recent article by Jones (1979). This robust tendency toward observer bias has been referred to by Ross (1977) as the "fundamental attribution error." Ross emphasizes the ubiquity of overattributing personal dispositions to account for behavior and reports the results of several ingenious experiments that carry the phenomenon beyond the realm of attitude attribution per se. Nevertheless, the attitude attribution paradigm stands out because of its historical importance as a basis for the actor–observer divergence proposition (Jones & Nisbett, 1971), and because research within the paradigm has withstood intense scrutiny for artifacts; that is, explanatory alternatives lacking in general theoretical interest.

The present experiment was designed to explore the role of temporal delay in the attitude attribution process, and to determine whether attributions for constrained behavior are affected by the locus of constraint information in the experimental sequence.

The role of temporal delay is of interest because of a clear analogy between the attitude attribution paradigm and the traditional sleeper effect paradigm in attitude change. According to the early formulations of Hovland and his colleagues (Hovland & Weiss, 1951; Kelman & Hovland, 1953), a sleeper effect may occur when a communication is paired with a negative or low credibility source. Over time, their reasoning goes, the discounting information about the source becomes dissociated from the communication, which, therefore, has a delayed effect on attitude change. Gillig and Greenwald (1974) have argued that the evidence for the sleeper effect is suspect. Nevertheless, it seems reasonable that under some conditions, competing pieces of information would have different decay functions in memory. Since the behavior involved in an expressed opinion is presumably more salient than the contextual information about constraining conditions, there is some reason to think that retention of the essay's direction and extremity will be better than retention of the discounting information that it was prepared under highly constraining conditions. The constraint information is viewed as analogous to discounting information about the source in

This research was supported by G-1114X from the National Science Foundation.

Requests for reprints should be sent to Edward E. Jones, Department of Psychology, Princeton, New Jersey 08540.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3707-1230$00.75

sleeper effect studies, and degree of correspondence is analogous to attitude change, since it represents a departure from the neutral or typical attitude that presumably would have been attributed in the absence of any information about the target person.

Order of information has been constant in all previous attitude attribution studies. In the natural world, we usually learn about or make surmises about the context of behavior before or during the action itself. Nevertheless, it might be argued that the tendency to overattribute dispositions from constrained behavior may be adequately explained as the effect of recency on recall, since information about the context in the typical attitude attribution study is temporally more remote than the opinion statement from the dependent attribution rating. In the present study, therefore, statement-context and context-statement orders were directly compared. If the statement-context order were to eliminate the fundamental attribution error, this would place severe qualifications on the generality of the effect.

Method

Subjects

Undergraduate volunteers ($N = 79$) participated for money in a study of impression formation. Subjects were run in groups of 12 or less, and each participated in two experimental sessions that were 1 week apart.

Procedure

Session 1. When subjects arrived for the first experimental session, they were told that the experimenter was interested in "how people form impressions of others depending on the information they receive and how this information is conveyed" (video, audio, written, etc.). Subjects were told to expect to receive information about another person through various means in the experimental sessions in which they would be participating.

A booklet containing the following materials was then given to each subject: an essay supposedly written by a target person (male) concerning the use of a minorities quota system in determining college admissions; instructions supposedly given to the target person at the time the essay was written; and a questionnaire to be completed by the subject. The essay was either in favor of or opposed to the use of a quota system in college admission policy. Each essay was approximately 370 words long and contained four arguments in support of the stand taken.

The instructions included in the booklet either gave the target person the freedom to take whatever position he chose on the issue or assigned him a position, pro or anti, to support in his essay. The essay always endorsed the position assigned. Instructions to the target person were prefaced by the statement: "You might be interested in knowing that the essay was written under the following instructions." The target person was then described as being asked to imagine that he was a member of a debating team and was instructed to prepare a series of persuasive essays or opening statements on various opinion issues. Included in these instructions was either the no-choice manipulation, "Today I would like you to write a persuasive essay favoring [opposing] the quota system as a part of college admission procedures," or the choice manipulation, "Today I would like you to write a persuasive essay either in favor of or opposed to the quota system as a part of college admission procedures." Finally, order of presentation of the essay and instructions was varied; the instructions were either presented immediately before or immediately after the essay in the booklet.

To summarize, the three independent variables manipulated in Session 1 were position taken in an essay toward the use of a quota system in college admissions (pro or anti), instructions under which the essay was prepared (choice or no choice), and order of presentation of essay and instructions in the booklet (essay first or essay last). Each subject was randomly assigned to one of the eight experimental conditions.

After reading the material, subjects completed a questionnaire that included evaluation of the target person's attitude toward the use of minority quotas in college admissions and other related issues. These included evaluation of the target person's attitude toward the busing of school children to facilitate integration, attitude toward the use of minority quotas in the hiring of employees, willingness to support actively the implementation of the quota system in college admissions, frequency of discriminatory behavior toward others based on minority membership, and general liberalism–conservatism. The questionnaire continued with several manipulation checks ("How much freedom did the target person have to choose the position or side he took?" "direction and extremity of the target person's essay..."); a final question asked for the subject's own attitude and his or her estimate of the typical undergraduate attitude toward the use of a minorities quota system. Subjects were then asked to return for another experimental session 1 week later. They were directed to talk with no one concerning the experiment during that week.

Session 2. When subjects returned for the second experimental session, they were first asked to recall the information they had received the week before about the target person and were then asked to complete a questionnaire containing several of the dependent measures administered in the first session. It was explained that some subjects might have changed their feelings or forgotten the first week's material over the time interval; therefore, they were urged to record their current impressions of the target person who wrote the essay. To provide suitable justifications for the second session and for rating changes if the subject's current

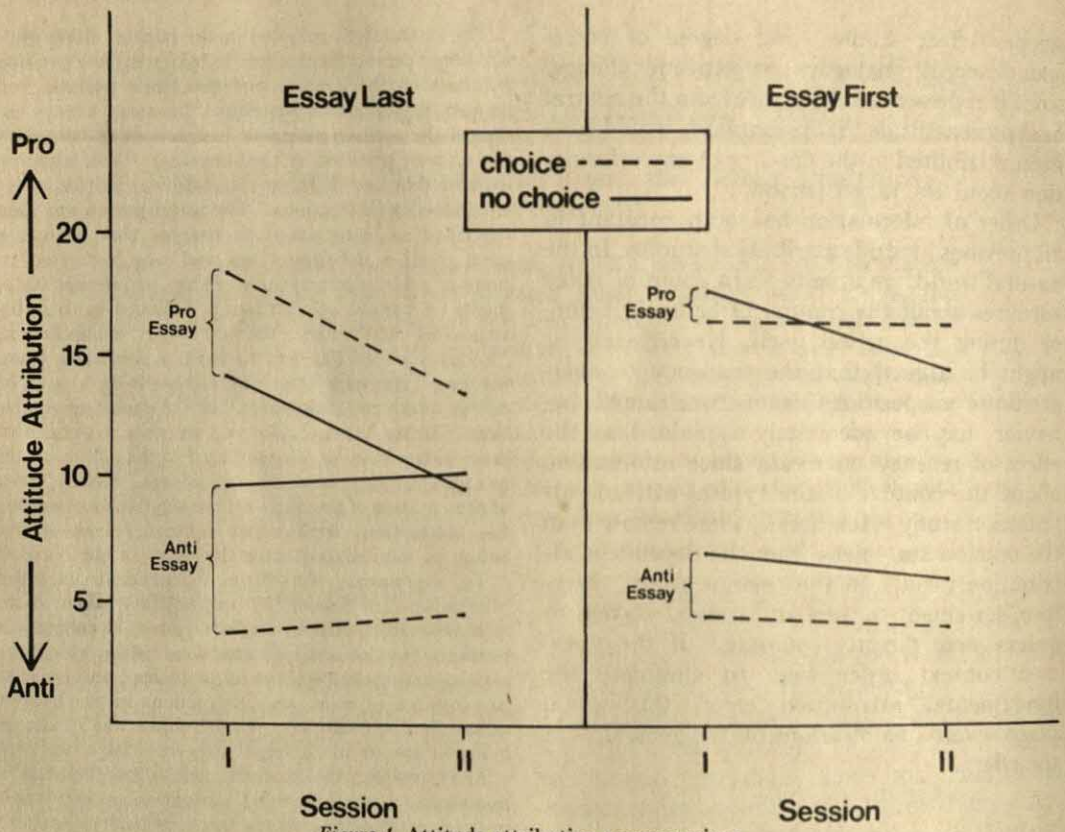


Figure 1. Attitude attribution across sessions.

feelings differed from his or her recalled impression, a 5-minute tape of a supposed interview between the experimenter and target person was then played to provide "additional information." The interview avoided discussion of social issues and described a person who seemed to be a typical college freshman.

Following the tape, subjects were asked to evaluate the target person on a series of interview-relevant dimensions and to reevaluate the target person on the same dependent measures administered previously. This third set of ratings did not differ markedly from the second, although subjects tended to converge toward a more neutral attributed attitude, probably because of the rather bland and noncommittal impression conveyed by the tape. The ratings made prior to the tape in Session 2 are the focus in the subsequent presentation of Session 1–Session 2 differences.

At the close of the second session, subjects were fully debriefed.

Results

The present experiment had two primary goals: (a) to explore the possibility of a sleeper effect or the differential retention of behavioral versus contextual information over time and (b) to investigate the role of information order in leading observers into the "fundamental attribution error." We shall deal with data relevant to both of these goals after evaluating the general evidence on observer bias.

Correspondence and Observer Bias

Previous findings (cf. Jones, 1979) have shown greater correspondence of attitude attribution under choice than no-choice conditions. This was again clearly demonstrated in the present experiment: When the target person was alleged to have a choice in producing the essay read by subjects, he was seen to hold a more extreme attitude toward quota systems in education (in line with the direction of the essay) than when the target person had no apparent choice. The statistical interaction between essay direction and choice was significant in both sessions: Session 1 $F(1, 71) = 20.32$, $p < .001$; Session 2 $F(1, 71) = 8.45$, $p < .01$.

Previous findings of significant correspondence of attribution and essay direction under no-choice conditions were also replicated in both sessions. In Session 1, the no-choice–pro essayist was seen as more in favor of quotas ($\bar{M} = 15.9$) than the no-choice–anti essayist ($\bar{M} = 8.4$), $F(1, 71) = 56.85$, $p < .001$. In Session 2, the tendency was reduced but still significant, $F(1, 71) = 6.14$, $p < .05$. In both sessions target persons with choice were, of course, attributed even more extreme attitudes.

A Sleeper Effect?

Since past evidence suggested that subjects pay more attention to behavior (figure) than its context (ground), it seemed quite possible that behavior in the no-choice conditions would have even more impact as remembered in Session 2 than in Session 1. In other words, there might be a sleeper effect in attitude attribution as a function of forgetting the less salient constraint information more rapidly than the more salient behavior.

The results provide no support for the sleeper effect notion. A repeated-measures analysis of variance showed very little variance ($F < 1.00$) attributable to the crucial interaction between essay direction, choice, and session. In fact, as Figure 1 shows, the observer bias or fundamental attribution error disappears by the second session *if* the context has preceded the essay in the first session. The bias is maintained in the essay-first condition. Clearly, in any event, there is no hint of the kind of sleeper effect when the impact of behavior direction becomes greater over time. If anything, the implications of contextual constraints themselves become stronger over time.

Table 1
Attributed Attitude Toward Quota System in College Admissions (Session 1): Means

Position	Choice		No choice	
	Essay first	Essay last	Essay first	Essay last
Pro	16.3	18.3[a]	17.6	14.1
Anti	4.4	3.8	7.0	9.7

Note. 1 (strongly opposed) to 21 (strongly in favor).
[a] $n = 9$ in this cell; $n = 10$ in all others.

Table 2
Attributed Attitude Toward Quota System in College Admissions (Session 1): Analysis of Variance

Source	*df*	*MS*	*F*
Essay direction (A)	1	2,111.23	159.58**
Constraint (B)	1	38.16	2.88
Information order (C)	1	.49	
A × B	1	160.29	12.15**
A × C	1	15.68	1.19
B × C	1	6.12	
A × B × C	1	95.90	7.25*
Within error	71	13.23	

* $p < .01$. ** $p < .001$.

At least this appears to be true when the no-choice context is initially presented prior to the essay.

Effects of Information Order

It is the case that all previous research on attitude attribution has featured a standardized order of information about contextual constraints followed by an opinion statement. Although this probably parallels mundane reality most of the time, an obvious question is whether context is relatively ignored (and behavior relatively emphasized) because of a recency effect. Since the essay or speech is the last piece of information before the crucial rating in the standard paradigm, perhaps its differential influence is a function of its differential salience in memory.

The present results clearly speak against such a recency effect. In fact, whether one looks at the data from Session 1, Session 2, or combines the data from both in a repeated-measures analysis, the essay has *greater* impact when it comes first. For convenience of exposition we shall concentrate on the Session 1 data, though we have already seen (in Figure 1) that the differences of interest become greater in Session 2.

Tables 1 and 2 present the means and analysis of variance summary, respectively, of ratings on a 21-point scale of attributed attitude toward "the quota system in college admissions." It can be seen that there is a significant triple interaction, $F(1, 71) = 7.25$, $p < .01$, that focuses on the greater spread between no-choice–pro and no-choice–anti

Table 3
Judged Essay Direction, Perceived Freedom, and Confidence: Means

Measure	Essay first				Essay last			
	Pro		Anti		Pro		Anti	
	C	NC	C	NC	C	NC	C	NC
Judged direction and extremity[a]	19.2	19.6	2.9	3.5	18.2	16.4	3.9	5.2
Perceived freedom[b]	3.4	16.2	4.2	14.6	4.6	20.3	2.9	20.5
Confidence[c]	17.3	14.7	16.3	15.6	14.0	10.3	18.5	12.2

Note. C = choice; NC = no choice.
[a] 1 (essay strongly opposed) to 21 (essay strongly favored).
[b] 1 (complete freedom) to 21 (no freedom).
[c] 1 (low confidence) to 21 (high confidence).

ratings when the essay is presented first rather than last. This is also apparent in Figure 1 (see no-choice, Session 1 means). The triple interaction is larger than the simple Direction × Order interaction because order of information has no impact within the choice conditions.

Related traits. The reader will recall that subjects also rated the target person on related opinion issues ranging from busing to general liberalism. These ratings were summed, and an analysis of variance revealed a pattern of F ratios almost identical to those on the most directly relevant admissions quota question. The triple interaction was again significant, $F(1, 71) = 4.43$, $p < .05$, as was the simple Order × Direction interaction, $F(1, 71) = 5.70$, $p < .05$. On both the direct trait rating and the related ratings composite, the simple effect of essay direction was significant regardless of when the context was mentioned, though the difference was significantly greater when the essay preceded the explanation of context.

Supporting perceptions. In approaching the question of why the attribution error is greater when the essay comes first, it is important to note that the essay-first subjects appear to see the behavior and the situation differently than the essay-last subjects. When asked to judge the extremity of the essay, compared to the essay-last subjects the essay-first subjects rated the pro essay as more in favor of and the anti essay as more opposed to the quota system; for the interaction between order and direction, $F(1, 71) = 10.60$, $p < .01$.

Not only was the essay judged more extreme when it preceded the context, but the essayist was judged to have greater freedom when the no-choice context was mentioned after exposure to the essay. The fact that this effect of order on perceived freedom differentiated the two orders only in the no-choice conditions is reflected in a significant Constraint × Order interaction, $F(1, 71) = 5.25$, $p < .05$.

In line with these two findings, subjects in the essay-first condition were significantly more confident of their attitude attributions than were subjects in the essay-last condition. This difference was especially marked in the no-

Table 4
Judged Essay Direction, Perceived Freedom, and Confidence: Analysis of Variance Summary

Source	*F*		
	Extremity	Freedom	Confidence
Essay direction (A)	733.71***	.27	
Constraint (B)	.0	164.21***	10.08**
Information order (C)	.32	5.04*	4.52*
A × B	2.45	.03	
A × C	10.60**	.07	
B × C	.56	5.25*	
A × B × C	1.77	.95	

* $p < .05$. ** $p < .01$. *** $p < .001$.

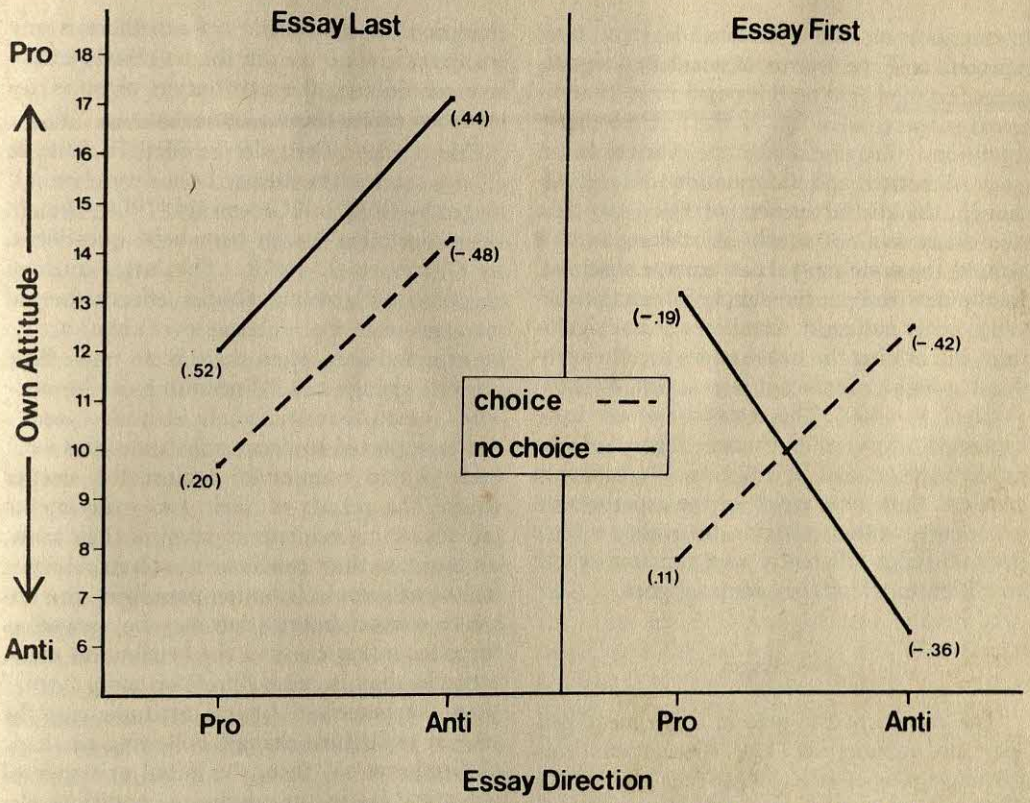


Figure 2. Own attitude by experimental conditions. (Numbers in parentheses refer to within-cell correlations.)

choice conditions, $F(1, 71) = 6.94$, for order within no choice.

The means and analysis of variance summaries on these items are in Tables 3 and 4, respectively.

Own and perceived typical attitude. Subjects were finally asked to rate their own attitude toward the quota system in college admissions as well as that of the average undergraduate. In previous attitude attribution studies, the subjects' own subsequently measured attitude has sometimes, but by no means always, been positively related to their attributions to the target person (cf. Jones, 1979). In the present study a most distinctive interaction was discovered. This is portrayed in Figure 2. In all but those no-choice conditions in which the essay was presented first, subjects showed a contrast effect: Those exposed to a pro essay rate themselves as more against quotas than those exposed to an anti essay. However, this trend was markedly reversed in the no-choice, essay-first conditions. The resulting triple interaction (Essay Direction × Constraint × Order) was highly significant, $F(1, 70) = 15.05$, $p < .001$. The same trends are also present in the data on perceived typical attitude. The triple interaction was again significant, though not quite as strong: $F(1, 70) = 7.75$, $p < .01$.

Figure 2 also indicates the within-cell correlations between attributed attitude and own attitude. None of these correlations was significant, and the pattern of the larger ones is uninterpretable and unrelated to the distribution of means as a function of the independent variables. Thus we are faced not with a general tendency for attributed attitude to follow own attitude or to contrast with own attitude, but a tendency for the various experimental conditions to have systematic effects on attributed attitude, own attitude, and perceived typical attitude.

The most reasonable accounting for the errant findings in the no-choice, essay-first cells, is attribution to chance: The significant

interactions on the own-attitude data may represent a Type I error of accidental significance. To shed light on this possibility, 28 additional subjects were run in the four no-choice conditions (i.e., retaining the variations in essay direction and information order). Although the distinctiveness of the essay-first conditions was not nearly as striking in this sample, the same general pattern was observed. In the new sample the simple interaction between order and essay direction was not significant, but adding the two samples together produced a highly significant interaction, $F(1, 63) = 22.77$, $p < .001$. This seems too robust to be explained away as a chance vagary of the random assignment of subjects. We can only conclude that some aspect of the experimental experience caused subjects to evaluate their own attitudes differently as a function of the condition to which they were assigned.

Discussion

The results of the present experiment provide no support for any speculations concerning a sleeper effect in attitude attribution. Reasoning by analogy with Hovland and Weiss's (1951) finding that in delayed recall subjects remember the content of a message better than discounting characteristics of the source, we speculated that the positions of an author might be better recalled after a week's time than the circumstances under which the author prepared the statement. In the present experiment subjects returned a week after their exposure to a pro or anti essay on minority quotas and were asked to rate the true attitude of the writer without further exposure to his essay or the constraint context in which it was allegedly prepared. Subjects previously run in the high-choice condition showed little or no change in their unequivocal attitude attributions. In contrast, subjects initially exposed to a highly constrained target person showed the customary observer bias in the first session, but this dissipated completely in the second session *if* the contextual information had preceded the essay. When the essay was presented first, there was some regression to a neutral attribution by the second session, but much of the biased attribution remained. There was clearly no evidence for a sleeper effect in any condition; on the contrary, when there is any change over time (i.e., in the no-choice, essay-last conditions), the attribution error is diminished rather thean augmented.

Prior evidence for a sleeper effect in attitude change studies has already been severely challenged by Gillig and Greenwald (1974), though their conclusion has in turn been questioned by Gruder et al. (1978). The latter authors suggest that absolute sleeper effects (actual increases in attitude change over time) are to be expected only when there is no immediate attitude change in a "discounting cue" group. They reason that immediate attitude changes can be expected to decay with time, and such decay would countervail against the sleeper effect. The results of their two experiments provide strong evidence in favor of their view. To translate these concerns into terms relevant for the attitude attribution paradigm, the no-choice context information may be viewed as the discounting cue and the attribution of an attitude that is more "pro" or more "anti" than the perceived typical attitude may be viewed as attitude change. Following the logic of Gruder et al., then, the initial existence of observer bias in the no-choice conditions decreases the likelihood that an absolute sleeper effect would be observed: Such an effect would have to operate against the general tendency for attitude attributions to regress toward the neutral point or the perceived typical attitude.

The present results are not even consistent with the possibility of a *relative* sleeper effect—when the effects of discounted material are relatively more persistent than the effects of undiscounted material. There was, in fact, no discernible decay of effect in the undiscounted (choice) conditions, whereas a clear decay was apparent in one of the discounted conditions (no-choice, essay-last). In contrast to the possibility that the essay, because of its initial salience, would retain its impact better than information about constraint, the direction of the essay seemed to be ignored by the average subject in the no-choice, essay-last condition. This convergence over time (see Figure 1) may reflect regression to the mean, but it should be noted that there was no such convergence in the essay-first conditions. It is still necessary, then, to account for differential regression as a func-

tion of the position of the discounting cue (no-choice instructions) during Session 1.

It should also be noted that the present attitude attribution paradigm differs from the typical sleeper effect paradigm in partially reinstating both the communicator and his "message." Thus in the typical sleeper effect study, the delayed measure of attitude is taken in a separate experimental context, and no mention is made of the initial experimental session. In the present study, however, it was necessary to remind subjects of the prior experimental session and the target person whom they would again be asked to judge. This is *not* analogous to the reinstatement condition of Kelman and Hovland (1953), in which subjects were reexposed to the discounting cue itself (a description of the communicator that manipulated his credibility). Nevertheless, the inescapable partial reinstatement involved in bringing to mind the previous week's session may have somehow muted the sleeper effect.

The most striking results of the present experiment are those associated with the effects of information order. Although the fundamental attribution error is demonstrable whether the essay follows or precedes the information about constraint (no choice), it clearly is strengthened when the essay is read prior to the constraint information. In fact, there was no significant difference between attributions in no-choice and choice conditions when the essays came first (Essay direction $\times$ Order interaction, $F = 2.06$, *ns*). Also, there was little or no decay in the attribution effects over a week's time. It appears that the no-choice information is virtually ignored when the subject has earlier been exposed to a strong pro or anti essay. Why might this be the case?

Our findings are reminiscent of the perseverance effect identified by Ross, Lepper, and Hubbard (1975). These investigators found that both actors' and observers' impressions persisted when their basis was subsequently discounted. The perseverance effect has been explained in terms of a cognitive consolidation process wherein the initial impression is buttressed through a reaching out for supportive data in remembered prior experiences, or through an initial salience of confirmatory associations. In the present experiment, subjects in the essay-first conditions presumably assume at first that the essay is written in the absence of clear-cut constraints and therefore represents the true viewpoint of the writer. Their impression of the writer's attitude toward minority quotas in education is readily buttressed by associations to related attitudes and personality attributes that are automatically inferred. The readers will recall that inferences about related attributes (toward busing, etc.) were significantly more consistent with the essay when it preceded the context than when the essay was last.

When the essay-first subjects then learn that the essay was written under highly constraining conditions, they make an adjustment in their assessment of attitudinal extremity, but the buttressing cognitive work they have already accomplished renders this adjustment insufficient. (Indeed, it is not even statistically significant.) In the essay-last conditions, on the other hand, the adjustment is more adequate and the modest (but still significant) tendency toward overattribution is accompanied by a modest amount of cognitive buttressing. The perseverance explanation, then, implies that the initially attributed attitude remains attributed (or is possibly reattributed) as a necessary implication of the associated attitudes that have already been cognitively constructed by the subject. This would not be the first time that the origins of an inference have been disregarded once the inference is made, and new or different origins have been assigned. Snyder and Frankel (1976) for example, showed that emotional dispositions are inferred to account for emotional behavior that was itself previously inferred from an arousing context.

But what shall we make of the distinctive tendency for the subjects' own attitudes, and their estimates of the perceived typical attitude, to vary positively with the attitude attributed to the essay writer *only* in the no-choice, essay-first conditions? We know from the correlational patterning (and the total lack of any overall correlation between own and attributed attitude—overall $r = -.099$) that this distinctive relationship is clearly produced by the independent variables and their particular order. The one explanation that occurs to us involves a mechanism of consistency-maintaining cognitive work not unlike the well-known strategies of dissonance reduction.

The subjects who read a pro-quota essay, for example, who make a strong attribution and then learn that the writer had no choice, are in a situation of dissonance or inconsistency. They have been "taken in." They have made an inference that is not warranted because of the information they belatedly learn about the context. This may be dissonant or inconsistent with the subjects' desire to see themselves as sophisticated and capable of seeing when someone (in this case the essay writer) does not really mean what he says. Faced with this mildly unpleasant challenge to their self-concept, the subjects may attempt to rationalize their premature commitment by shifting their conception of the typical attitude and of their own attitude. After all, if the essay is at or near the consensus position, it is perfectly reasonable to assume that the essay writer really believed what he said—even though there is a reason to discount the essay. In this way, subjects are able to maintain their attributional commitment for a reason that is not vulnerable to the new, discounting information.

A dissonance account is also consistent with the fact that little attribution decay is observed from Session 1 to Session 2 in the no-choice, essay-first conditions (see Figure 1). If the attribution were buttressed by similar attributions about the self and typical other, then one would expect the attributions to be quite stable over time.

Regardless of the eventual explanation for the strong order effects observed in the present experiment, it is clear that the "fundamental attribution error" is enhanced when discounting information follows the formation of an unequivocal impression. It goes without saying that observer overattribution under conditions of constraint is more than an effect of the subject's most recent exposure to the target person's behavior.

References

Gillig, P. M., & Greenwald, A. G. Is it time to lay the sleeper effect to rest? *Journal of Personality and Social Psychology*, 1974, *29*, 132–139.

Gruder, C. L., et al. Empirical tests of the absolute sleeper effect predicted from the discounting cue hypothesis. *Journal of Personality and Social Psychology*, 1978, *36*, 1061–1074.

Hovland, C. I., & Weiss, W. The influence of source credibility on communication effectiveness. *Public Opinion Quarterly*, 1951, *15*, 635–650.

Jones, E. E. The rocky road from acts to dispositions. *American Psychologist*, 1979, *34*, 107–117.

Jones, E. E., & Nisbett, R. E. *The actor and the observer: Divergent perceptions of the causes of behavior*. New York: General Learning Press, 1971.

Kelman, H. C., & Hovland, C. I. "Reinstatement" of the communicator in delayed measurement of opinion change. *Journal of Abnormal and Social Psychology*, 1953, *48*, 327–335.

Ross, L. The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10). New York: Academic Press, 1977.

Ross, L., Lepper, M. R., & Hubbard M. Perseverance in self-perception and social perception: Biases attributional processes in the debriefing paradigm. *Journal of Personality and Social Psychology*, 1975, *32*, 1004–1013.

Snyder, M. L., & Frankel, A. Observer bias: A stringent test of behavior engulfing the field. *Journal of Personality and Social Psychology*, 1976, *34*, 857–864.

Received September 20, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 7, 1239–1248

Prediction of Behavior From Beliefs: An Extension and Test of a Subjective Probability Model

James Jaccard and Richard Knox
Purdue University

David Brinberg
University of Illinois

A model for the prediction of behavior from beliefs was tested in the context of voting behavior in the presidential election of 1976. A two-wave panel design was used in a field survey in which voting behavior was predicted from beliefs about candidates in accord with a recent subjective probability model. Results supported the model, yielding an average correlation between predicted and obtained voting behavior of .75. The average correlation between predicted and obtained voting intention was .84. The ability of the model to predict behavior and behavioral intentions was relatively unaffected by the educational level of the respondent and the type of beliefs under consideration.

A major concern of social psychology has been the prediction of social behavior. Researchers have used measures of attitude (e.g., Wicker, 1969), norms (e.g , Schwartz, 1973) and personality traits (e.g., Mischel, 1968) in the attempt to achieve accurate behavioral prediction—all with mixed success. Dulany (1968) presented an approach to behavioral prediction based on behavioral intent. According to this perspective, a person's behavior (at least most behaviors of concern to social psychologists) is largely a function of his or her *intention* to perform that behavior. Research by Dulany and others (e.g., Ajzen & Fishbein, 1973) has observed strong relations between behavioral intentions and overt behavior. Conceptual discussions of the relationship between these two variables have been presented by Dulany (1968) and Fishbein and Jaccard (1973).

A number of theorists have attempted to extend this work by specifying factors that determine behavioral intentions. A number of models of the relationship between intentions and psychological variables have been presented in the context of attitude research (Fishbein & Ajzen, 1975), verbal learning (Dulany, 1968), cross-cultural research (Triandis, 1974), and consumer psychology (Sheth, 1973). Each of these approaches is based on the general linear model, in which a set of predictor variables (e.g., attitudes, norms) are weighted and summed to yield an overall prediction of behavioral intent. Generally speaking, the types of predictor variables used have been similar across the various theories.

An alternative approach to understanding the relation between behavioral intentions and psychological variables has been presented by Jaccard and King (1977). These theorists were concerned with the relationship between beliefs and behavioral intentions. A belief was defined as a perceived relation between two "objects." The term *object* was used in its most generic sense, so that it could refer to a person, a behavior, an attribute, a goal, or some concept. Belief strength referred to the subjective probability linking the two objects. For example, in the belief statement, "Smoking cigarettes causes cancer," belief strength refers to the individual's perceived probability that "smoking cigarettes causes cancer." This probability may range from .00 to 1.00. A behavioral intention was defined as a perceived relation between oneself and some behavior. Characterizing this relation is also a subjective probability. For example, the statement, "I intend to vote for Gerald Ford" consists of a perceived relation between "myself" and the

Requests for reprints should be sent to James Jaccard, Department of Psychological Sciences, Purdue University, West Lafayette, Indiana 47907.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3707-1239$00.75

behavior of voting for Gerald Ford. The subjective probability linking these two objects may vary from .00 ("I definitely intend not to vote for Gerald Ford") to 1.00 (I definitely intend to vote for Gerald Ford).

A behavioral intention, according to Jaccard and King, does not differ in its basic structure from any other belief. However, it does possess certain characteristics that distinguish it from other beliefs: (a) it always links a person with some action, (b) it always refers to future behavior, and (c) it is usually correlated with overt behavior (Dulany, 1968; Fishbein & Ajzen, 1975). These characteristics suggest that it would be useful to more fully understand this "special" belief and its relation to other beliefs.

Given that both beliefs and behavioral intentions can be conceptualized as subjective probabilities, Jaccard and King (1977) attempted to extend the work of Wyer (1974) using a model relating these subjective probabilities. Specifically, Wyer (Wyer & Goldberg, 1970; Wyer, 1970, 1974) tested and found support for the following equation as a model of relations among beliefs:

$$P_A = P_B P_{A|B} + (1 - P_B) P_{A|\overline{B}}, \quad (1)$$

in which P_A = the subjective probability that Proposition A is true, P_B = the subjective probability that Proposition B is true, $P_{A|B}$ = the subjective probability that Proposition A is true given that Proposition B is true, and $P_{A|\overline{B}}$ = the subjective probability that Proposition A is true given that Proposition B is not true. Jaccard and King (1977) attempted to extend this model to the analysis of behavioral intentions by rewriting Equation 1, substituting subjective probabilities characterizing a behavioral intention:

$$P_I = P_B P_{I|B} + (1 - P_B) P_{I|\overline{B}} \quad (2)$$

in which P_I = a person's intention to perform a behavior (e.g., vote for Gerald Ford), P_B = a person's perceived probability that Proposition B is true (e.g., Gerald Ford favors reduced government spending), $P_{I|B}$ = a person's perceived probability that he or she would perform the behavior, given that Proposition B is true, and $P_{I|\overline{B}}$ = a person's perceived probability that he or she would perform the behavior, given that Proposition B is not true. A discussion of the conceptual relationship between this model and the models of Fishbein (1972) and others has been presented by Jaccard and King (1977); hence, it will not be considered here. The difference between the two conditional probabilities, $P_{I|B} - P_{I|\overline{B}}$, was termed the psychological relevance of a belief for a behavioral intention. When this difference is large, B may be said to be psychologically relevant to the intention (e.g., a person would vote for Gerald Ford if he favored reduced government spending but would not vote for Gerald Ford if he did not favor reduced government spending). Conversely, when the difference between these two conditional probabilities is low, B may be said to be psychologically irrelevant to the intention (e.g., a person would vote for Gerald Ford regardless of whether or not he favored reduced government spending). In general, a change in a belief will produce a change in a behavioral intention to the extent that the difference between the two conditional probabilities is large.[1]

Jaccard and King (1977) presented data in the context of two experiments that supported the model in Equation 2. Wyer (1970, 1974, 1976) presented data that are consistent with Equation 1. All of these tests, however, dealt with a restricted range of subjects (viz., college students), and the utility of these models has yet to be demonstrated with a more heterogeneous population in a field setting. The purpose of the present investigation is three-fold. (a) To extend previous research by Wyer in an attempt to specify the relationship between beliefs and behavior instead of just the relation of beliefs to beliefs. This involves the incorporation of the concept of behavioral intention into Wyer's model. (b) To test the generality of Wyer's model in the context of an application of the intention model to a noncollege population. Specifically, the predictability of the model is evaluated as a function of the education of respondents. Several theorists (e.g., Bem, 1970) have suggested that models of belief consistency

[1] This assumes that a change in P_B does not affect $P_{I|B}$ or $P_{I|\overline{B}}$. For a discussion of this issue and the model's implication for understanding *changes* in behavioral intentions, see Jaccard and King (1977).

have greater validity for bright college students and intellectuals than for the majority of the population. If such suggestions are correct, then the model should predict better for high versus low education respondents. (c) To investigate psychological factors (e.g., wishful thinking) that may affect the predictability of the intention model.

Method

Respondents

Respondents were 119 males and females living in a Midwestern community of moderate size. Sampling procedures were designed to obtain a heterogeneous sample in terms of education with an approximately equal number of Democrats, Republicans, and Independents. Respondents were contacted by phone and administered a brief phone survey. On completion, they were asked if they would be willing to complete a longer survey in their home. The goal of the sampling procedures was *not* representativeness, but rather heterogeneity, so as to achieve a more stringent test of the generality of the model. The final sample consisted of 54 males and 65 females who ranged in age from 23 to 56 with a mean age of 34.6. Education level varied from less than a high school education to a professional degree, with 34 respondents having not completed high school and 85 respondents with a high school degree or more. A total of 36 Democrats, 48 Republicans, and 34 Independents were interviewed (one individual did not complete the party identification question). Refusals to participate in the final survey were relatively low, being less than 10% of those initially contacted.

Materials

The home interviews took approximately 1 hour to complete. During the first stage of the interview, the interviewer presented an oral introduction to the study and a standardized set of instructions concerning the questionnaire. The respondent then filled out a practice section. This section contained examples of each type of scale and served to anchor the endpoints and eliminate any "warm-up" effects from the data; in addition, the interviewer could detect any misunderstanding of scales prior to actual data collection. The entire interview schedule was self-administered. This probably minimized any effect of "interviewer" on the subject's responses. The interviewer did, however, remain in the house while the respondent was completing the questionnaire to answer any questions the respondent might have.

On the basis of reviews of national magazines, local newspapers, and campaign materials distributed by the Democratic and Republican parties, a set of 17 issues was compiled (e.g., abortion and reduced government spending). A list of these issues is presented in the results section of this article. Data were collected to test Equation 2 for the presidential election. Specifically, for each candidate, respondents provided estimates of P_B, $P_{I|B}$ and $P_{I|\bar{B}}$ for each of the 17 issues. Examples of the statements used to measure these parameters are as follows. P_B = Jimmy Carter favors reduced government spending, $P_{I|B}$: Suppose Jimmy Carter did, in fact, favor reduced government spending. How likely is it you would vote for Jimmy Carter, $P_{I|B}$: Suppose Jimmy Carter did *not* favor reduced government spending. How likely is it you would vote for Jimmy Carter. Each statement was rated on a 21-point unlikely–likely scale ranging from 0 to 100 with a midpoint of 50. Scores were converted to probability units by multiplying each score by .01. These three measures were obtained for each candidate for each of the 17 issues, yielding 102 questions. Behavioral intention (P_I) was ascertained by asking each respondent the probability that he or she would vote for Jimmy Carter and Gerald Ford on a 21-point unlikely–likely scale. A behavioral intention measure was independently obtained for both candidates.

In addition to the above measure, evaluations of each of the 17 issues were obtained on a 7-point bipolar semantic differential scale, with good and bad as the endpoints. A number of demographic variables were also ascertained, including party identification (using the traditional measure of Berelson, Lazarsfeld, & McPhee, 1954), education, religion, self-reported liberal–conservativeness, age, and marital status. In addition, a number of general beliefs about the Democratic and Republican parties were measured. The order of all items was (a) evaluations of each of the 17 issues, (b) beliefs about the Democratic Party, (c) beliefs about the Republican Party, (d) belief strength (P_{Bi}) that Jimmy Carter favors each of 17 issues, (e) belief strength that Gerald Ford favors each of 17 issues, (f) demographics, (g) behavioral intentions (P_I), (h) additional attitudinal and background variables, and (i) the conditional probabilities, $P_{I|Bi}$ and $P_{I|\bar{B}i}$. Note that measures pertaining to each of the parameters of Equation 2 were separated so as to reduce the possibility of respondents perceiving consistencies among them.

Questionnaires were administered approximately 1 week prior to the election. One day after the election, respondents were recontacted and asked if they had voted and, if so, for whom. These self-reports served as estimates of voting behavior.[2]

Results

Reliability of Measures

To ascertain test–retest reliabilities of the probability scales used in the present investigation, a number of items were repeated at the end of the questionnaire. The average test–

[2] Research in political science has tended to indicate that such self-reports of voting behavior are veridical (cf. Campbell, Converse, Miller & Stokes, 1960; Fishbein & Coombs, 1974).

Table 1
Correlations and Mean Absolute Discrepancies for Predicting Voting Behavior and Voting Intention from the Subjective Probability Model: Total Sample

Issue	Carter			Ford		
	B with $\hat{P}_I$	P_I with $\hat{P}_I$	$\lvert P_I - \hat{P}_I \rvert$	B with $\hat{P}_I$	P_I with $\hat{P}_I$	$\lvert P_I - \hat{P}_I \rvert$
Amendment to ban abortion	.77	.85	.12	.68	.81	.16
Reduce defense budget	.80	.87	.13	.78	.86	.13
Busing	.70	.79	.16	.72	.83	.15
Amnesty	.76	.86	.12	.76	.87	.13
Registration of guns	.77	.86	.13	.71	.77	.17
National health system	.76	.85	.14	.77	.80	.16
Inflation by unemployment	.77	.87	.14	.74	.83	.16
Labor unions	.75	.84	.14	.72	.81	.16
Nuclear energy	.73	.81	.16	.75	.84	.15
Guaranteed annual income	.74	.83	.15	.72	.83	.15
Local vs. federal control	.75	.85	.14	.75	.79	.17
Cooperation with Russia	.75	.84	.14	.71	.76	.18
Closing tax loopholes	.76	.86	.14	.78	.86	.15
Honesty	.70	.81	.16	.78	.84	.14
Experience	.78	.88	.12	.72	.87	.12
Clarity on positions	.79	.88	.12	.76	.85	.13
Reduced government spending	.79	.87	.12	.78	.86	.13
M	.758	.850	—	.744	.831	—

retest correlation of these items was .93, indicating that the scales were relatively reliable. In addition, a number of checks were made to ensure that subjects' responses were not artificially correlated by means of method variance (Campbell & Fiske, 1959). For example, it was assumed a priori that belief strength in two statements, (a) Jimmy Carter is in favor of a constitutional amendment to ban abortion and (b) Jimmy Carter favors reducing the defense budget, should be theoretically uncorrelated (i.e., knowing an individual's score on the first belief should not be predictive of his score on the second belief). Any correlation between these two beliefs would, therefore, reflect common method variance. The correlation between these beliefs was .04 (*ns*). Similar comparisons on a total of 10 belief statements (5 comparisons per candidate), assumed to be uncorrelated a priori, yielded an average absolute correlation of .08.[3] Other data, which indicate that responses are not artificial, are presented in a later part of the Results section.

Prediction of Behavior from Behavioral Intention

To test the relation between behavioral intention (P_I) and voting behavior, a point biserial correlation was computed between P_I and behavioral score (1 = voted for the candidate, 0 = did not vote for the candidate) for each candidate separately. The correlation between stated intention to vote for Carter and observed voting behavior was .87, $p < .01$. The corresponding correlation for Ford was .85 ($p < .01$). These correlations did not differ across a number of individual difference variables (e.g., education and religion) and indicate a relatively strong relation between intentions and behavior.

Prediction of Behavioral Intention from Beliefs

To test the model of behavioral intention proposed by Jaccard and King (1977), a predicted P_I score was derived in accord with Equation 2 for each subject and each of the 17 issues. Table 1 presents the correlation of each of these predicted intentions ($\hat{P}_I$) with behavior and with the observed intention to vote for Carter and Ford. In terms of predicting

[3] Obviously, such a test for method variance assumes the underlying assumption of uncorrelated variables is veridical. Details of the additional comparisons for method variance can be obtained by writing the authors.

voting behavior (VB), the correlation between $\hat{P}_I$ and VB ranged from .68 to .80 with a mean r to Z transformation of .75. All correlations were significant ($p < .01$) and did not vary appreciably across issues or candidates. In terms of predicting behavioral intention, the correlation between predicted and observed intention ranged from .76 to .88, with a mean of .84. Strictly speaking, if Equation 2 is valid, then no curve-fitting parameters should be required to predict observed intention from the predicted intention score. In other words, a regression analysis should yield a slope of 1.0 and an intercept of 0. The appropriate slopes and regression weights were computed for each belief. In no case did the observed slope differ significantly from 1.0 nor did the intercept differ significantly from 0. Finally, in no case did the correlation with P_I of any individual parameter of the model (i.e., P_B, $P_{I|B}$, $P_{I|\bar{B}}$, $P_{I|B} - P_{I|\bar{B}}$) exceed the correlation between P_I and $\hat{P}_I$. The correlation of P_I with each P_B ranged from $-.45$ to .67 with an average absolute correlation of .19; for P_I with $P_{I|B}$ the correlations ranged from .49 to .86 with an average correlation of .79; for P_I with $P_{I|\bar{B}}$ the correlations ranged from .50 to .86 with an average correlation of .75; for P_I with $P_{I|B} - P_{I|\bar{B}}$ the correlations ranged from $-.24$ to .45 with an average correlation of .16; for P_I with $P_B P_{I|B}$ the correlation ranged from .29 to .81 with an average correlation of .64; for P_I with $(1 - P_B)P_{I|\bar{B}}$ the correlations ranged from .05 to .70 with an average correlation of .45.

In terms of prediction, the full model presented in Equation 1 was a consistently better predictor of the P_I than any component of the equation considered in isolation. This is reflected not only in the correlational analyses above but in other indices of goodness of fit. The true superiority of the full model is probably underestimated due to the fact that the increased number of parameters and combinatorial rules result in greater degrees of measurement error than the correlations based on single components.

In addition to the correlational analyses above, the absolute difference between P_I and $\hat{P}_I$ was computed for each subject on each belief. The mean discrepancy score for each belief is presented in Table 1. Discrepancy scores ranged from .12 to .18, indicating little variability in predictability across beliefs. However, these discrepancy scores did differ significantly from perfect predictability (i.e., a hypothetical discrepancy mean of zero). This is not surprising, considering measurement and experimental error in the parameters of the model.

To test the generality of the model across individuals, the sample was divided into two groups: (a) high education (high school graduate or more) and (b) low education (did not complete high school). For the high education group, the correlation between $\hat{P}_I$ and VB ranged from .63 to .80, with a mean of .74. Similarly, the range of correlations between P_I and $\hat{P}_I$ was .77 to .91, with a mean of .85. In no case did the slope differ significantly from 1.0, and in only one case did the intercept differ significantly from 0. For the low education group, the results were comparable. The correlation between $\hat{P}_I$ and VB ranged from .65 to .84, with a mean of .75. The correlation between P_I and $\hat{P}_I$ ranged from .60 to .88, with a mean of .80. In no case did the slope differ significantly from 1.0 nor the intercept from 0. Although the difference in correlations between the two groups is not large, there is a consistent tendency for the model to predict better for high versus low education subjects. Out of 34 correlations (17 for each candidate), the correlation coefficient between P_I and $\hat{P}_I$ was higher (albeit, nonsignificantly so) for the high education group on 27 of them. This was significant ($p < .05$) via a sign test. More important, the standard error of estimate was lower for the high education group on all 34 relations. Thus, there was a tendency for the model to predict consistently better for high versus low education respondents.

In addition to the above correlation analyses, an analysis of mean $|P_I - \hat{P}_I|$ scores was performed. Specifically, a Hotelling T^2 test was performed comparing high versus low education respondents on the vector of mean discrepancy scores for each belief. For the 17 beliefs concerning Carter, a T^2 of 13.95, $F(17, 101) = .71$, *ns*, was observed, whereas for the 17 beliefs concerning Ford, a T^2 of 27.54, $F(17, 101) = 1.40$, *ns*, was observed. Thus, the analysis of mean absolute dis-

Table 2
Mean Scores of Selected Parameters of the Subjective Probability Model

Issue	Carter					Ford				
	$\hat{P}_I$	P_B	$P_{I\|B}$	$P_{I\|\bar{B}}$	Relevance	$\hat{P}_I$	P_B	$P_{I\|B}$	$P_{I\|\bar{B}}$	Relevance
Amendment to ban abortion	.44	.41	.39	.44	−.05	.59	.53	.56	.58	−.02
Reduce defense budget	.45	.77	.43	.44	−.01	.60	.33	.60	.59	.01
Busing	.44	.56	.39	.48	−.09	.58	.48	.53	.62	−.09
Amnesty	.44	.76	.44	.42	.02	.57	.33	.56	.58	−.01
Registration of guns	.46	.66	.45	.43	.02	.58	.47	.58	.57	.01
National health system	.47	.77	.47	.41	.06	.56	.37	.55	.57	−.02
Inflation by unemployment	.46	.38	.31	.49	−.18	.54	.54	.48	.61	−.13
Labor unions	.41	.73	.39	.45	−.06	.56	.45	.53	.51	−.08
Nuclear energy	.46	.55	.46	.40	.06	.60	.72	.63	.53	.10
Guaranteed annual income	.41	.65	.39	.44	−.05	.59	.34	.49	.62	−.13
Local vs. federal control	.45	.58	.47	.39	.08	.60	.59	.63	.52	.09
Cooperation with Russia	.44	.57	.46	.41	.05	.60	.76	.61	.53	.08
Closing tax loopholes	.50	.80	.52	.34	.18	.55	.48	.69	.47	.22
Honesty	.42	.51	.61	.20	.41	.61	.71	.71	.30	.41
Experience	.43	.43	.53	.34	.19	.64	.81	.68	.46	.22
Clarity on positions	.40	.36	.56	.32	.24	.59	.72	.65	.41	.24
Reduced government spending	.46	.59	.52	.35	.17	.61	.75	.67	.46	.21

Note. Observed P_I Carter = .41; observed P_I Ford = .61; N = 119.

crepancy scores yielded no differences in predictability as a function of education. In general, mean discrepancy scores for each separate educational group were similar to those presented in Table 1 for the total sample.

Table 2 presents mean scores for the components of the behavioral intention model as well as for $\hat{P}_I$. The observed mean P_I for Carter was .41, whereas for Ford it was .61. It can be seen that the model also tended to accurately predict mean behavioral intention scores. The range of differences in mean $\hat{P}_I$ versus mean P_I was −.09 to .07, with a mean of −.01. In no case did the observed mean differ significantly from the predicted means.

From an applied perspective, Table 2 has some interesting implications. First, one can examine the mean belief strength scores (P_B) for each candidate. For each of the 17 issues, the candidates took specific positions during the course of their campaigns. It is interesting to note that in every case the direction of the difference in means is consistent with the stands taken by the candidates. For example, relative to Ford, Carter was more in favor of busing, amnesty, registration of guns, a national health system, labor unions, a guaranteed annual income, reducing the defense budget, and closing tax loopholes for big business. These differences were reflected in the mean P_B scores. This suggests, contrary to many perspectives (e.g., Campbell, Converse, Miller, & Stokes, 1960; Sears 1969), that voters may have considerable information about candidates and that they may accurately perceive differences between them.

Examination of mean psychological relevance scores (i.e., $P_{I|B} - P_{I|\bar{B}}$) indicates the following beliefs as being the most important for influencing behavioral intentions: (a) reducing inflation by policies encouragning unemployment, (b) closing tax loopholes for big business, (c) being trustworthy, (d) being experienced, (e) clearly stating one's positions, and (f) reduction of government spending. However, it should be noted that in general, the difference between the conditional probabilities for each of these beliefs was relatively small (approximately .20). This probably reflects the fact that change in any one belief about a candidate would not have a great impact on voting intention due to the importance of several other beliefs in the determination of voting intentions. However, changing several beliefs may have a dramatic impact on voting intention and behavior. For further

discussion of this issue, see Jaccard and King (1977).

Additional Factors Affecting Model Predictability

In addition to individual differences in model predictability, a set of analyses was undertaken to determine whether the type of belief under consideration influenced the ability of the subjective probability model to predict intentions. For each belief statement, a predictability index was defined as $|P_I - \hat{P}_I|$, for each individual. McGuire (1960a, 1960b, 1960c) has suggested that such variables as (a) the favorability of the beliefs in question (i.e., wishful thinking) and (b) how "affectively charged" beliefs are, may affect the logical relations between beliefs. Similarly, these variables may affect the predictability of the intention model.

To test the extent to which model predictability is related to how "desirable" the belief in question was, the predictability index was correlated with the evaluative ratings of each issue (on a 7-point semantic differential good–bad scale scored +3 to −3). In general, the correlations were low and nonsignificant. The range was −.19 to .11 with an average correlation of −.02.

In addition, the correlations between the predictability index and the absolute value of the evaluative ratings was also computed for each belief statement. This analysis would indicate the extent to which model predictability is influenced by how "affectively charged" the issue is, irrespective of the direction of that affect (positive or negative). Again, the correlations were typically low, with a range of −.20 to .15 and an average correlation of −.03.

These correlations must be interpreted cautiously, since they involve the correlation of one variable with a difference score. It is a well-known fact that difference scores can be highly unreliable and would tend to mask a significant true correlation via measurement error. However, "test" measures used in the study generally indicated high reliability, suggesting this may not be a problem. Nevertheless, no direct data are available to estimate the reliability of the difference score.

Discussion

The present data support the model of behavior and behavioral intention based on Equation 2. The average correlation between voting behavior and voting intention was .86. The strength of this relationship did not vary as a function of numerous individual difference variables. In addition, the average correlation between predicted versus obtained voting intention was .84. In general, regression analyses indicated that ad hoc curve fitting procedures were unnecessary for prediction of voting intention (i.e., intercepts did not differ significantly from 0.0 and slopes did not differ significantly from 1.0). Furthermore, the strength of the relationship between predicted and observed voting intention did not vary appreciably as a function of the respondent's education. However, there was a tendency for the model to predict consistently better for high versus low education respondents. Predictability of the model was found to be unaffected by the desirability of the belief on which it was based or how "affectively charged" the belief was viewed. These data indicate that Wyer's (1970) probability model can be applied to the understanding and prediction of behavior as well as beliefs.

The present model of behavioral intention has a number of interesting implications for theoretical and applied research. First, it represents an interesting alternative to models of behavioral intention presented by Fishbein (1972), Triandis (1974), Sheth (1973), and others. As noted earlier, each of these approaches is based on the general linear model in which a set of predictor variables (e.g., beliefs) are weighted and summed to yield an overall prediction of behavioral intent. Each of these models relies on a regression equation for its basic format. In contrast, the present model does not rely on such an approach and circumvents a number of problems raised in the context of correlational approaches (see Jaccard & King, 1977, for an elaboration).

Second, the present model has a number of interesting implications for changing behavioral intentions. If $P_{I|B}$ and $P_{I|\bar{B}}$ are held constant, then the following relation should hold:

$$\Delta P_I = \Delta P_B(P_{I|B} - P_{I|\bar{B}}), \qquad (3)$$

in which ΔP_I = the change in the subjective probability characterizing the behavioral intention, ΔP_B = the change in the subjective probability characterizing the belief, and all other terms are as previously defined. If a persuasive message effectively changes P_B, its corresponding influence on P_I will be proportional to the difference $P_{I|B} - P_{I|\overline{B}}$. More important, Equation 3 suggests that the logical inconsistency created by the change in P_B can be resolved in three ways: (a) by changing the intention to perform the behavior (P_I), (b) by changing the psychological relevance of the belief for the intention (i.e., by changing the difference $P_{I|B} - P_{I|\overline{B}}$), or (c) by a combination of the above. Which mode of resolution individuals use and the factors that influence this choice remain to be investigated.

Third, the model possesses obvious relevance to researchers in applied areas. On the basis of literature reviews or pretests, a set of beliefs may be generated for which measures in Equation 2 could be obtained. To understand variations in intentions, it would be possible to compare those who intend to perform the behavior with those who do not intend to on the strength of belief in each belief statement (P_{Bi}). A second comparison could be made between the two groups for differences in the psychological relevance of the belief for the intention (i.e., $P_{I|Bi} - P_{I|\overline{B}i}$). It is entirely possible that a given belief (P_{Bi}) will *not* covary with intention (i.e., that intenders and nonintenders will not differ on P_{Bi}), but that the belief will still be psychologically relevant to the intention (i.e., $P_{I|Bi} - P_{I|\overline{B}i}$ will be nonzero), and furthermore, that the magnitude of this relevance may differ as a function of whether or not the individuals intend or do not intend to perform the behavior. The present model allows the researcher to examine these factors for any number of groups defined on the basis of any "external" variable (such as sex or religion).

The present model of behavioral intention is based closely on Wyer's (1970) model of cognitive organization. It is not clear, however, whether research generated by Wyer and others in this area (e.g., McGuire, 1960a, 1960b, 1960c) is directly applicable to the present model. For example, it is not clear whether such phenomena as the Socratic effect and cognitive inertia would be observed when the criterion in question is a behavioral intention rather than a belief. The present data suggest that wishful thinking, for example, may *not* affect the behavioral intention model, although it has been shown to influence the cognitive organization model. Future research should help to identify the influence of these phenomena on beliefs versus behavioral intentions.

The present data do not allow a meaningful comparative test of the model of intention based on Equation 2 with other more prominent models (e.g., Fishbein, 1972: Triandis, 1974). Due to differing numbers of parameters, differing combinatorial rules, and differing reliability of measures, a prediction strategy such as the present one is simply not conducive to such comparative tests. However, the results of the present investigation are encouraging and suggest that comparative tests are worth pursuing. These would probably take the form of manipulative or change studies whereby differential predictions of the competing models are derived (e.g., one model predicts that a change in a belief should lead to a change in an intention, whereas another model does not).

Some additional advantages of the present model should be noted relative to other strategies for understanding the relationship between beliefs and behavior. One common approach used in relating beliefs to behavior is to obtain individuals' agreement or disagreement with a set of belief statements and then to correlate each one of these with behavior (in the present study, this would be analogous to correlating each P_B with the behavioral measure). These beliefs may be entered into a multiple regression equation and the overall multiple R as well as standardized regression coefficients of each belief interpreted. Alternatively, the belief ratings may be factor analyzed and the resulting factor scores entered into a regression equation. These correlational approaches suffer from two distinct shortcomings, however: (a) a belief may be highly correlated with a behavior yet still not have causal implications for the behavior (i.e., the correlation may be spurious) and (b) a belief may be uncorrelated with a behavior yet still have

causal implications for the behavior.[4] Any form of correlational analysis is ambiguous with respect to causality. In contrast, these problems are circumvented in the context of the conditional probabilities of the present model. If a belief is spuriously correlated with an intention (and hence, behavior), this should be reflected in a small difference between the conditionals $P_{I|B} - P_{I|\overline{B}}$. Similarly, if a belief is uncorrelated with an intention, yet has causal implications for it, this should be reflected in a large difference between the conditionals. Thus, the present model, if valid, circumvents some of the problems associated with these traditional correlational strategies.

A major difference between the present model and alternative approaches is the fact that most of these approaches explicitly consider multiple beliefs as determinants of intentions. Multiple belief measures are obtained and then combined in some fashion and related to the corresponding intention. This view is not incompatible with the present approach in that multiple beliefs are also considered. However, Equation 2 is applicable to only one belief at a time (i.e., measures of P_B, $P_{I|B}$, and $P_{I|\overline{B}}$ are obtained for each belief separately and analyzed accordingly). It would be possible to modify Equation 2 in accord with probability theory to incorporate multiple beliefs simultaneously. However, the equation would become cumbersome and would probably be of little practical utility when the number of beliefs is three or greater. The approach outlined in the present article is probably a more appropriate strategy. This approach should provide insights into the extent to which changes in any *one* belief will produce changes in behavior. It is possible that changes in one belief will not produce changes in behavior but that changes in multiple beliefs would. This state of affairs would be suggested by a set of beliefs whose psychological relevance $(P_{I|B} - P_{I|\overline{B}})$ was uniformly small.[5] In this situation, a model of multiple beliefs should be applied as a supplement to the present one. None of the current theories consider this problem, and such a model is now being developed and tested.

References

Ajzen, I., & Fishbein, M. Attitudinal and normative variables as predictors of specific behaviors. *Journal of Personality and Social Psychology*, 1973, *27*, 41–57.

Bem, D. *Beliefs, attitudes, and human affairs*. Belmont, Calif.: Brooks-Cok, 1970.

Berelson, B., Lazarsfeld, P., & McPhee, W. N. *Voting: A study of opinion formation in a presidential campaign*. Chicago, Ill.: University of Chicago Press, 1954.

Campbell, A., Converse, P. E., Miller, W. E., & Stokes, D. E. *The American voter*. New York: Wiley, 1960.

Campbell, D. T., & Fiske, D. W. Convergent and discrimination validation by the multitrait–multimethod matrix. *Psychological Bulletin*, 1959, *56*, 81–105.

Dulany, D. E. Awareness, rules and propositional control: A confrontation with S-R behavior theory. In D. Horton & T. Dixon (Eds.), *Verbal behavior and S–R behavior theory*. Englewood Cliffs, N.J.: Prentice-Hall, 1968.

Fishbein, M. Toward an understanding of family planning behavior. *Journal of Applied Social Psychology*, 1972, *2*, 214–227.

Fishbein, M., & Ajzen, I. *Belief, attitude, intention and behavior: An introduction to theory and research*. Reading, Mass.: Addison-Wesley, 1975.

Fishbein, M., & Coombs, F. Basis for decision: An attitudinal analysis of voting behavior. *Journal of Applied Social Psychology*, 1974, *4*, 95–124.

Fishbein, M., & Jaccard, J., Theoretical and methodological issues in the prediction of family planning intentions and behavior. *Representative Research in Social Psychology*, 1973, *4*, 37–52.

Jaccard, J., & King, G. W. A probabilistic model of the relationship between beliefs and behavioral intentions. *Human Communications Research*, 1977, *3*, 332–342.

McGuire, W. J. Cognitive consistency and attitude change. *Journal of Abnormal and Social Psychology*, 1960, *60*, 345–353. (a)

McGuire, W. J. Direct and indirect persuasive effects of dissonance producing messages. *Journal of Abnormal and Social Psychology*, 1960, *60*, 354–358. (b)

McGuire, W. J. A syllogistic analysis of cognitive relationships. In C. Hovland & M. Rosenberg (Eds.), *Attitude organization and change*. New Haven, Conn.: Yale University Press, 1960. (c)

Mischel, W. *Personality and assessment*. New York: Wiley, 1968.

---

[4] An example will make this latter point explicit. A group of individuals may believe very strongly that a given airline is safe to fly on. Due to a restriction in range, a measure of this belief would not correlate with intention to fly on the airline. Nevertheless, if individuals learned that the airline was unsafe, this would probably have a decided impact on intentions.

[5] Alternatively, this might also suggest that each of the beliefs has nothing to do with the behavior.

Schwartz, S: Normative explanations of helping behavior: A critique, proposal, and empirical test. *Journal of Experimental Social Psychology*, 1973, *9*, 349–364.

Sears, D. Political behavior. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology*. Reading, Mass.: Addison-Wesley, 1969.

Sheth, J. N. A field study of attitude structure and attitude-behavior relationship. In J. N. Sheth (Ed.), *Models of buying behavior: Conceptual, quantitative, empirical*. New York: Harper & Row, 1973.

Triandis, H. C. Culture training, cognitive complexity, and interpersonal attitudes. In R. Brislin, S. Bochner, & W. Lonner (Eds.), *Cross-cultural perspectives on learning*. Beverly Hills, Calif.: Sage Publications, 1974.

Wicker, A. W. Attitudes vs. actions: The relationship of verbal and overt behavioral responses to attitude objects. *Journal of Social Issues*, 1969, *25*, 41–78.

Wyer, R. S. Quantitative prediction of belief and opinion change: A further test of a subjective probability model. *Journal of Personality and Social Psychology*, 1970, *16*, 559–570.

Wyer, R. S. *Cognitive organization and change*. Potomac, Md.: Erlbaum, 1974.

Wyer, R. S. An investigation of the relations among probability estimates. *Organizational Behavior and Human Performance*, 1976, *15*, 1–18.

Wyer, R. S., Goldberg, L. A probabilistic analysis of relationships among beliefs and attitudes. *Psychological Review*, 1970, *77*, 100–120.

Received September 25, 1978 ■

Editors Named for New *Journal of Personality and Social Psychology*

The Publications and Communications Board has named Melvin Manis, Ivan D. Steiner, and Robert Hogan to edit the *Journal of Personality and Social Psychology* (JPSP), which will appear in a new format in January 1980. The affiliations of the new editors are, respectively, University of Michigan and Veteran's Hospital, Ann Arbor; University of Massachusetts; and Johns Hopkins University. Manis will edit the section on ATTITUDES AND SOCIAL COGNITION, Steiner the section on INTERPERSONAL RELATIONS AND GROUP PROCESSES, and Hogan the section on PERSONALITY PROCESSES AND INDIVIDUAL DIFFERENCES.

An announcement on the restructuring of JPSP appeared on p. 668 of the June 1978 issue of JPSP. As indicated at that time, the three editors will work independently with separate groups of advisory editors. Their acceptances will appear as distinct sections in monthly issues of JPSP. Detailed descriptions of each section's coverage appear on pages 468–469 of the March issue of JPSP.

Authors should submit manuscripts to the appropriate section editor as follows:

ATTITUDES AND SOCIAL COGNITION

Melvin Manis
Research Center for Group Dynamics
Institute for Social Research
University of Michigan
Ann Arbor, Michigan 48106

INTERPERSONAL RELATIONS AND GROUP PROCESSES

Ivan D. Steiner
Department of Psychology
University of Massachusetts
Amherst, Massachusetts 01003

PERSONALITY PROCESSES AND INDIVIDUAL DIFFERENCES

Robert Hogan
Johns Hopkins University
Department of Psychology
33rd and Charles Streets
Baltimore, Maryland 21218

NEW FOR 1980

ANNOUNCING!

PsycSCAN™ CLINICAL PSYCHOLOGY

PsycSCAN™ DEVELOPMENTAL PSYCHOLOGY

Two new abstract journals each with over 1,500 full abstracts per year from relevant, selected journals—APA and non-APA—in the fields of clinical and developmental psychology. These new publications are the results of more than three years of study and analysis of the information needs of APA members.

The unique scanning format of PsycSCAN allows the reader to quickly review the specially set title, index terms, abstract, and citation for each entry.

Keep current with the ongoing literature in these important fields in psychology. Plan to order one or both PsycSCAN journals when you complete your 1980 dues statement.

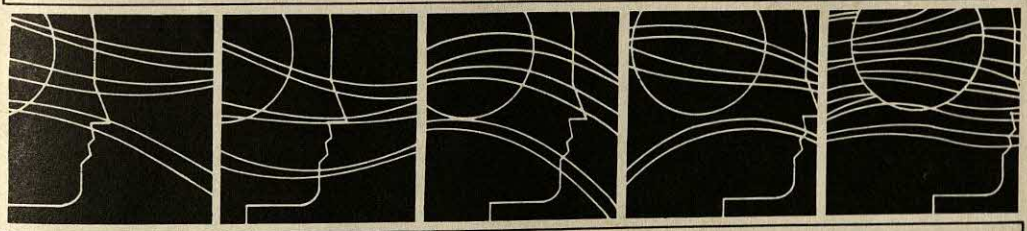
Each PsycSCAN journal is:

- $8 per year for APA members ($10 for nonmembers)
- Quarterly, starting in early 1980
- Available for easy check-off ordering on the 1980 dues statement

PsycSCAN™ is produced by PsycINFO™: a service of the American Psychological Association 1200 17th Street, NW Washington, DC 20036

Willo P. White, editor

Resources in Environment and Behavior



NEW FROM APA IN 1979

An invaluable sourcebook for students, instructors, and researchers in the new field of environment and behavior. This solid reference includes:

- Overview and history of this emerging field
- Graduate programs—both formal and informal
- Teaching innovations introduced in the United States, Canada, and Great Britain
- Funding sources
- Career opportunities
- Directory of key individuals currently working in the field
- Annotated bibliography
- Listing of relevant journals

Resources in Environment and Behavior is available from the American Psychological Association for $10 (soft cover only). To order, make your check payable to APA.

American Psychological Association
Order Department
1200 17th Street, NW
Washington, DC 20036

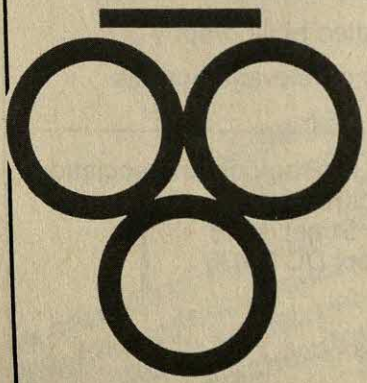
Please include full remittance for orders of $25 or less.

APA EMPLOYMENT BULLETIN

Possibly your key to greater career opportunities.

Each month 125 to 150 position openings are listed by geographical area and employment field for psychologists with a minimum of a Master's degree. Members may also publish availability notices. Coded identification numbers assure confidentiality. Ideal for employers, students, and Master's and Ph.D graduates.

A one year subscription is as little as $9.00 for members, $16.00 for non-members. If you're in a hurry, First Class delivery is available for an additional $4.00.



For more information, contact:
The Editor, Employment Bulletin,
or send your order to:
American Psychological Association
Subscription Dept.-Employment Bulletin
1200 Seventeenth St., N.W.
Washington, D.C. 20036

(Continued from front cover)

Observer Bias in the Attitude Attribution Paradigm: Effect of Time and Information Order 1230
Edward E. Jones, Janet Morgan Riggs, and George Quattrone

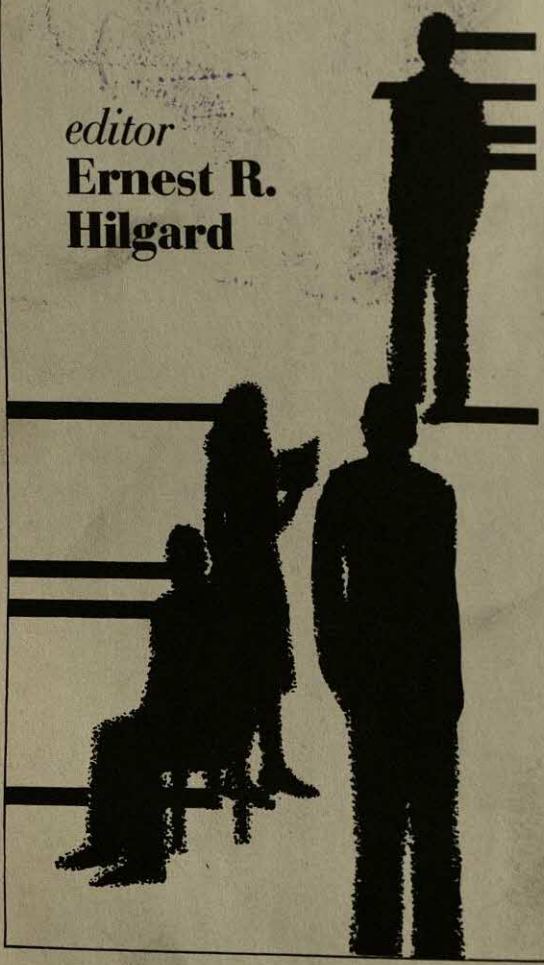
Prediction of Behavior From Beliefs: An Extension and Test of a Subjective Probability Model 1239
James Jaccard, Richard Knox, and David Brinberg

Editors Named for New *Journal of Personality and Social Psychology* 1249

American Psychology in Historical Perspective

1892-1977

editor
Ernest R. Hilgard



Now available in one book — 21 APA presidential addresses in their entirety from some of the most important names in American psychology. Included are classic pieces from James, Cattell, Dewey, Thorndike, Woodworth, and Watson. You'll also find Harlow's "The Nature of Love" and Miller's "Analytical Studies of Drive and Reward."

This book is truly a milestone for historical psychology. It provides both a fascinating chronology of the presidents of the American Psychological Association from 1892 to 1977 and a valuable collection of significant essays.

This new APA publication traces the development of American psychology over four broad periods in psychology's history: the first 25 years (1892-1916), the years of the two world wars (1917-1945), the 20 years after World War II (1946-1967), and the recent past (1968-1977).

For each of these periods, the editor, Ernest R. Hilgard, provides a brief summary of the thinking in psychology at the time, biographies of all the APA presidents with abstracts of their presidential addresses, and the selected presidential addresses in full.

American Psychology in Historical Perspective *may be ordered in hard cover for $18 or in soft cover for $15 by writing to: American Psychological Association, Order Dept., 1200 Seventeenth Street, NW, Washington, D.C. 20036*

Please include full remittance for orders of $25 or less.

JOURNAL OF

Personality AND Social Psychology

VOLUME 37 NUMBER 8 · AUGUST 1979

A Cybernetic Model of Self-Attention Processes 1251
Charles S. Carver

Some Effects of Deception and Level of Self-Monitoring on Planning and Reacting to a Self-Presentation 1282
Gregory C. Elliott

Effects of Speech Style and Sex of Speaker on Person Perception 1293
Nora Newcombe and Diane B. Arnkoff

Effects of Resource Availability and Importance of Behavior on the Experience of Crowding 1304
Richard McCallum, Caryl E. Rusbult, George K. Hong, Tedra A. Walden, and John Schopler

Identifying Victims of Groupthink From Public Statements of Decision Makers 1314
Philip E. Tetlock

The Birth Order Puzzle 1325
R. B. Zajonc, Hazel Markus, and Gregory B. Markus

Cultural Difference in Attribution of Performance: An Integration-Theoretical Analysis 1342
Ramadhar Singh, Meenakshi Gupta, and Ajit K. Dalal

The Effects of Reward Contingency and Performance Feedback on Intrinsic Motivation 1352
Judith M. Harackiewicz

Variables That Moderate The Attitude–Behavior Relation: Results of a Longitudinal Survey 1364
Andrew R. Davidson and James J. Jaccard

Mind Over Matter: Perceived Success at Psychokinesis 1377
Victor A. Benassi, Paul D. Sweeney, and Gregg E. Drevno

Communicator Physical Attractiveness and Persuasion 1387
Shelly Chaiken

(Continued on inside back cover)

Published monthly by the

AMERICAN PSYCHOLOGICAL ASSOCIATION

ISSN 0022-3514

Acting Editor

Clyde Hendrick, *University of Miami*

Action Editors

Irwin Altman, *University of Utah*
Charles S. Carver, *University of Miami*
Richard J. Harris, *University of New Mexico*
John H. Harvey, *Vanderbilt University*
William Ickes, *University of Missouri–St. Louis*
George Levinger, *University of Massachusetts—Amherst*
Brendan Gail Rule, *University of Alberta, Edmonton, Canada*
David R. Shaffer, *University of Georgia*
Cookie White Stephan, *New Mexico State University*
Barbara Strudler Wallston, *George Peabody College for Teachers*

Consulting Editors

Robert Steven Baron, *University of Iowa*
Paul A. Bell, *Colorado State University*
Ellen Berscheid, *University of Minnesota (Minneapolis)*
Michael H. Birnbaum, *University of Illinois at Urbana-Champaign*
Jack Brehm, *University of Kansas*
Philip Brickman, *Northwestern University*
J. Merrill Carlsmith, *Stanford University*
Paul Christopher Cozby, *California State University, Fullerton*
Rick Crandall, *Texas Christian University*
James M. Dabbs, Jr., *Georgia State University*
John M. Darley, *Princeton University*
Andrew R. Davidson, *University of Washington*
James H. Davis, *University of Illinois at Urbana-Champaign*
Kay Deaux, *Purdue University*
Marshall Dermer, *University of Wisconsin—Milwaukee*
Edward Donnerstein, *Iowa State University*
Ebbe B. Ebbesen, *University of California, San Diego*
Yakov M. Epstein, *Rutgers—The State University*
Irene Hanson Frieze, *University of Pittsburgh*
J. Martin Giesen, *Mississippi State University*
Lewis R. Goldberg, *University of Oregon*
William G. Graziano, *University of Georgia*
Alan E. Gross, *University of Maryland (College Park)*
Nancy Hirschberg, *University of Illinois at Chicago Circle*
Chester A. Insko, *University of North Carolina at Chapel Hill*
Alice M. Isen, *University of Maryland, Baltimore County*
Martin F. Kaplan, *Northern Illinois University*
Samuel S. Komorita, *University of Illinois at Urbana-Champaign*
Vladimir J. Konečni, *University of California, San Diego*
Ellen Langer, *Harvard University*
Mark R. Lepper, *Stanford University*
Darwyn E. Linder, *Arizona State University*
Dale T. Miller, *University of British Columbia, Vancouver, Canada*
Judson Mills, *University of Maryland (College Park)*
Warren T. Norman, *University of Michigan (Ann Arbor)*
Michael Ross, *University of Waterloo, Waterloo, Canada*
Barry R. Schlenker, *University of Florida*
Kelly G. Shaver, *College of William & Mary*
Phillip Shaver, *New York University*
Marvin E. Shaw, *University of Florida*
Steven J. Sherman, *Indiana University*
Harold Sigall, *University of Maryland (College Park)*
Mark Snyder, *University of Minnesota (Minneapolis)*
Melvin L. Snyder, *University of Texas at Austin*
Shelley E. Taylor, *Harvard University*
Stephen G. West, *Florida State University*
Steve Worchel, *University of Virginia*
Mark P. Zanna, *University of Waterloo, Waterloo, Canada*

Assistant to the Editor

Sally Miller

APA Journal Staff

Anita DeVivo, *Executive Editor*
Ann I. Mahoney, Manager, *Journal Production*
Marjorie M. Henry, *Production Supervisor*
Robert J. Hayward, *Advertising Representative*
Juanita Brodie, *Subscription Manager*

The Journal of Personality and Social Psychology publishes original reports of research, methodology, theory, criticism, and literature review in the fields of personality and social psychology of normal processes and populations. Submitted manuscripts should delineate an important question and provide a clear answer to it. In the case of research reports, this will often require demonstration that the reported results are generalizable beyond a single research setting. Consult the editorial in the January 1976 issue (Vol. 33, pp. 1–7) for a detailed statement of policies and instructions to authors.

Manuscripts: Submit manuscripts in quadruplicate, prepared according to the *Publication Manual of the American Psychological Association* (2nd ed.). Copies should be highly legible and will not be returned to authors unless they are original typewritten copies. Mail manuscripts according to instructions on p. 1439.

Required supplementary information: Manuscripts cannot be reviewed for publication unless the author provides supplementary information in the following categories: (a) a cover letter that indicates the relation (or the lack of relation) of the submitted manuscript to other papers by the same author(s) that are currently submitted, in press, or recently published; (b) one copy of each such related paper; (c) one copy of a statement of the procedures of the research reported in the submitted manuscript in sufficient detail to serve as a guide for conduct of an exact replication; and (d) one copy of the summary tables (e.g., correlation tables, analysis of variance summaries) for the *major* analyses reported. If it appears that the supplementary materials in categories (c) or (d) might be bulky, the author should exercise judgment and restraint in selecting what to send. It is not necessary to send supplementary material that repeats verbatim manuscript content.

After a manuscript is accepted, authors will be expected to provide assurance that their data are available in unanalyzed form for reanalyses by other investigators and that their research procedures were consistent with the American Psychological Association's published principles of research ethics, except insofar as may be detailed otherwise in the manuscript.

Copyright and Permission: All rights reserved. Written permission must be obtained from the American Psychological Association for copying or reprinting text of more than 500 words, tables, or figures. Permission is normally granted contingent upon like permission of the author, inclusion of the APA copyright notice on the first page of reproduced material, and payment of a fee of $10 per page, table, or figure. Abstracting is permitted with credit to the source. Permissions and fees are waived for authors who wish to reproduce their own material for personal use; fees only are waived for authors who wish to use their own material commercially. Permission and fees are waived for the photocopying of isolated articles for nonprofit classroom or library reserve use by instructors and educational institutions. Libraries are permitted to photocopy beyond the limits of U.S. copyright law: (1) those post-1977 articles with a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through the Copyright Clearance Center, P.O. Box 765, Schenectady, NY 12301; (2) pre-1978 articles, provided that the per-copy fee stated in the Publishers' Fee List is paid through the Copyright Clearance Center, P.O. Box 765, Schenectady, NY 12301. Address requests for reprint permission to the Permissions Office, APA, 1200 Seventeenth Street, N.W., Washington, D.C. 20036.

Subscriptions: Subscriptions are available on a calendar year basis only (January through December). Nonmember rates for 1979: $60 domestic, $64 foreign, $6 single issue. APA member rate: $20. Write to Subscription Section, APA.

Change of Address: Send change of address notice and a recent mailing label to the attention of the Subscription Section, APA, 30 days prior to the actual change of address. APA will not replace undelivered copies resulting from address changes; journals will be forwarded only if subscribers notify the local post office in writing that they will guarantee second-class forwarding postage.

Reprints: Authors may order reprints of their articles from the printer when they receive proofs.

Back Issues and Back Volumes: For information regarding back issues or back volumes write to Order Department, APA.

Microform Editions: For information regarding microform editions write to any of the following: Johnson Associates, Inc., P.O. Box 1017, Greenwich, Connecticut 06830; University Microfilms, Ann Arbor, Michigan 48106; or Princeton Microfilms, Princeton, New Jersey 08540.

Published monthly in one volume per year by the American Psychological Association, Inc., 1400 North Uhle Street, Arlington, Virginia 22201. Printed in the U.S.A. Second-class postage paid at Arlington, Va., and at additional mailing offices.

Diary No. 192
Date 17. 4. 80
Librarian.
Bureau Edn. Psy. Resear

JOURNAL OF
Personality
AND
Social Psychology

Volume 37 · August 1979 Number 8

A Cybernetic Model of Self-Attention Processes

Charles S. Carver
University of Miami

An analysis of the experiential and behavioral consequences of self-directed attention is proposed, based on cybernetic or information-processing ideas. The proposed model includes the following assumptions: When attention is directed to environmental stimuli, those stimuli are analyzed and categorized according to the person's preexisting recognitory schemas. Self-directed attention often leads to a similar analysis of *self*-information; experientially, such a state of self-focus comprises an enhanced awareness of one's salient self-aspects. In some cases categorization—either of one's context or of some self-element—elicits a response schema, which constitutes a behavioral standard. If a prior categorization has evoked such a behavioral standard, subsequent self-attention engages an automatic sequence in which behavior is altered to conform more closely with the standard. This matching-to-standard is construed as the occurrence within a psychological system of a test-operate-test-exit unit or a negative feedback loop—a cybernetic construct applicable to many different phenomena both within and outside of psychology. More specifically, self-focus when a standard is salient is seen as the "test" or comparison phase of that feedback loop. If the matching-to-standard process is interrupted by one of several specified classes of events, subsequent behavior depends on an outcome expectancy judgment. A favorable outcome expectancy causes a return to the matching-to-standard attempt; an unfavorable expectancy results in behavioral withdrawal. Research support for the model is discussed, and comparisons are made with self-awareness theory, Bandura's theory of fear-based behavior, helplessness theory, and social comparison theory.

Psychological theorizing in this century has been characterized by the occasional emergence of quite general theories of behavior (such as psychoanalytic theory and learning theory), with intervening periods in which no general theory has seemed universally applicable and problem-specific theories instead have predominated. In recent years yet another general approach to understanding behavior has begun to coalesce, an approach usually identified with the terms *information-processing*, *cybernetic*, or *control* theory (e.g., Apter, 1970; Buckley, 1968; Powers, 1973; Singh, 1966; Wiener, 1948). In the minds of many theorists and researchers, information-processing ideas provide an ideal framework within which to confront a wide variety of problems. Indeed, it is a framework that, by virtue of its abstractness, many believe may provide the best basis yet attainable for a

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3708-1251$00.75

comprehensive model of behavior. This article represents an attempt to apply such ideas to an area of interest to social psychologists that has not been treated before in these terms. That area is the experiential and behavioral consequences of self-directed attention.

The article is divided into four major sections. The first section is intended to provide some background to the theory that is presented later. This background material includes a brief summary of Duval and Wicklund's (1972) theory of objective self-awareness and a basic introduction to the uses of the terms *cybernetic, control,* and *information processing*. In the second section a theoretical model is presented, along with some evidence in support of its propositions. This section represents the central portion of the article. Because of the unique relationship between this theory and Duval and Wicklund's (1972) self-awareness theory, a third section is included, which addresses specific points of conflict between the two models. The present theory also has some broader implications, however. These implications are considered in the final major section. That section includes comparisons between the present model and each of the following: Bandura's recent analysis of fear-based behavior (Bandura, 1977), learned helplessness theory (e.g., Abramson, Seligman, & Teasdale, 1978), and social comparison theory (Festinger, 1950, 1954).

Background

Self-Awareness Theory

In 1972 Duval and Wicklund proposed a theory of self-awareness that consisted of two central assumptions. First, they suggested that the objects of conscious attention could be dichotomized: Attention can be directed outward to the environment, or attention can be directed inward to the self (self-awareness).[1] Self-awareness is increased by stimuli that remind a person of himself or herself (e.g., a camera or a mirror); self-awareness is decreased by stimuli that distract attention from the self (e.g., perceptual-motor tasks, such as a pursuit rotor).

Duval and Wicklund's (1972) second assumption was that self-focus leads to a self-critical evaluation process such that the self-attentive person compares himself or herself with some standard on whatever behavioral dimension happens to be salient (see also Wicklund, 1975a).[2] Those authors further argued that self-directed attention is therefore usually aversive, because in most cases the person's actual behavior or state is worse

My most sincere gratitude is expressed to Clyde Hendrick, Hazel Markus, Paul Neumann, John Rosen, Michael Scheier, Neil Schneiderman, Shelley Taylor, Robert Wicklund, and particularly Paul Blaney, all of whom generously commented on one or more drafts of this article; to Mark Bickhard, who gave me my introduction to the concepts of cybernetics; and to my wife, Nora de la Garza, who remained unfailingly supportive throughout the long and arduous process of shaping the manuscript.

Requests for reprints should be addressed to Charles S. Carver, Department of Psychology, P.O. Box 248185, University of Miami, Coral Gables, Florida 33124.

[1] Duval and Wicklund (1972) originally termed self-focus "objective" self-awareness and environment focus "subjective" self-awareness. In recent years this rather cumbersome convention has been widely abandoned in favor of such alternatives as high versus low self-awareness or self-focus versus environment focus.

[2] It will be noted that many *potential* standards of comparison exist at any given moment, even with respect to a single behavioral dimension. However, in Duval and Wicklund's theory it was assumed that at any given time the following two conditions obtain: One standard is more salient than others, and the person *adopts* that standard as his or her comparison value, at least temporarily. The same assumption will be made throughout the present discussion. Cases in which the individual's attention is drawn to a potential standard that is not adopted as a comparison value (because it is seen as irrelevant to one's behavior or for some other reason) are explicitly excluded from the analysis. This issue will be taken up in somewhat greater detail below.

Perhaps the distinction between standards and dimensions should also be clarified at this point. The distinction is similar to that commonly made between the terms *value* and *variable*. A behavioral dimension (or a variable) such as aggressiveness or politeness is defined by some quality that may be present in one's behavior to a greater or lesser degree. A standard is a point on that behavioral dimension, for example, nonaggressiveness or moderate aggressiveness. The existence of a behavioral standard, of course, always presupposes the existence of a behavioral dimension.

than the standard of comparison. This aversiveness presumably leads the person to attempt to escape the self-aware state (Wicklund, 1975a, 1975b). If it is not possible to avoid self-awareness-inducing stimuli (and thus the aversive state), the self-focused person may attempt to alter his or her behavior so that it conforms more closely to the standard, as a way of reducing the aversiveness of self-awareness. That is, the aversiveness of self-attention was seen as the motivator for behavioral alterations. Thus self-awareness theory, as proposed by Duval and Wicklund (1972), was one of a larger class of drive theories—one in which the presence or absence of drive was determined by the direction of one's focus of attention.

This theory has been provocative, and it has led to a good deal of imaginative research. However, evidence has begun to accumulate that its assumptions are incorrect. The remainder of this article will be devoted to the assertion that the elements of self-awareness theory that have been supported by research findings can be most usefully construed as components of a more elaborate model of social cognition, and further, that such a model is most usefully framed in information-processing or cybernetic terms.

Information Processing, Cybernetics, and Control

Definitions. Some amplification on the uses of the terms *information processing, cybernetic,* and *control* is in order at the outset. The term *information processing* is used in two related but distinguishable ways by psychologists. At a micro level, the term refers to analyses of the exact bases by which information is encoded in and retrieved from memory. At the macro level, *information processing* applies to descriptions of control systems and is more synonymous with the term *cybernetic.* Cybernetics is the science of control and communications systems (Apter, 1970; Singh, 1966; Wiener, 1948, 1954). The term *control,* as used here, refers to the sequencing that is implicit in a set of instructions, each of which awaits and depends on the execution of the previous instruction (as in a computer program).

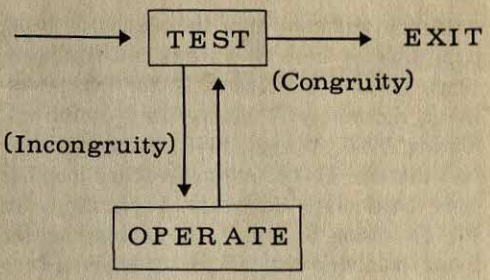


Figure 1. The TOTE unit, as presented and discussed by Miller, Galanter, and Pribram (1960).

A specific illustration may clarify the concept of control. Miller, Galanter, and Pribram (1960) introduced to psychologists a hypothetical construct called a TOTE unit (Figure 1); the word TOTE is an acronym for test-operate-test-exit. This construct is known more widely among cyberneticists and engineers as a negative feedback loop (negative because it reduces discrepancies between an existing state and a standard).

The diagram of a TOTE unit (Figure 1) portrays a self-regulatory system. The arrows in the diagram convey three things simultaneously. They represent energy, that is, the activation of a physical mechanism to do some sort of work; they represent information transfer, that is, the information that congruity between two values has or has not been attained; and they represent control, the fact that one process logically depends on an outcome of the previous process and then leads to a subsequent process. Any given process is said to control another when the execution of the first triggers the execution of the second. In effect, any decision-making flow chart describes a control sequence.

In general, the present use of the term *information processing* is meant to convey the latter (macro-level) meaning of the term. Though there are differences in implication between the terms *information-processing, cybernetic,* and *control* theory, the differences are relatively subtle, and the terms will be used interchangeably here.

Applications. Many people tend to think of the word *cybernetic* as applying solely to electronic computers. However, cybernetics as

a science was seen from its infancy as being applicable to both living and nonliving systems. To most cyberneticists, the living–nonliving dichotomy is not terribly useful (cf. Turing, 1950; Wiener, 1948). Because of the fact that the TOTE unit or feedback loop has some particularly important implications for the discussion to follow, let us consider the living–nonliving distinction as it applies to the TOTE unit (Figure 1).

In the control system of the TOTE unit, three mechanisms are activated in an orderly sequence. The first of these—*test*—is a mechanism by which a comparison is made between some input and a standard. If the input is not congruous with the standard, the second mechanism—*operate*—is activated. *Operate* alters the existing state of affairs in some way. Following *operate, test* recurs. If the input now is congruous with the standard, the final mechanism—*exit*—occurs, deactivating the system or freeing it for other applications. If, on the other hand, the second *test* reveals that incongruity still exists, *operate* is reactivated and continues to sequence alternately with *test* until congruity between input and standard is attained [3]

The most commonly used illustration of the operation of a TOTE unit is the behavior of a nonliving system: the room thermostat. The thermostat senses existing temperature, compares that existing temperature with a preset standard, and activates a furnace when the temperature falls discriminably below the standard or activates an air conditioner when the temperature rises above the standard. However, this control structure is certainly not limited to electrical systems. Indeed, in the view of cyberneticists, the feedback loop may occur in virtually any type of physical system (cf. Buckley, 1968; Kuhn, 1974). The same control structure that is realized electrically in the thermostat is realized biologically in innumerable forms: for example, the homeostatic mechanisms that maintain normal body temperature, maintain appropriate levels of oxygen and nutrients in the blood and tissues, and regulate hormonal balances (cf. Wiener, 1948, especially p. 114). Moreover, the same control structure can be realized in psychological systems. For example, it is implicit in a child's self-correcting attempts t reach out to touch a nearby object, and in th adult's checking and adjusting his or he clothing in a mirror.

Until recently, cybernetic constructs hav not been widely applied to subjects of interes to social psychologists. The theorizing pre sented below is intended, in part, as a ste in that direction.[4]

The Model

The control-theory model of self-attention processes proposed here subsumes many findings generated by researchers in self-awareness theory. It does not, however, incorporate all of the same assumptions as does that theory. The difference in assumptions between the two allows the present model to accommodate existing data with greater internal consistency and to make predictions that seem not to be easily derivable from self-awareness theory as framed by Duval and Wicklund (1972). The model consists of the following propositions, which will be presented along with some amplification and supporting evidence. (Figures 2 and 3 provide a flow-chart illustration of the interrelationship among propositions.)

1. *A dichotomy may be imposed on the potential objects of conscious attention, such*

---

[3] This alternation may be quite rapid, of course, depending on the characteristics of the physical system in which it occurs. (Indeed, some theorists, e.g., Powers, 1973, hold that this information flow is better conceptualized as being continuous at all stages of the feedback loop.) Thus, although we can easily separate the component elements logically, for other purposes it is more convenient to think of the entire feedback loop as the unit of analysis.

[4] Control theorists seem to have been more aware of the potential applications of such constructs to social behavior than have social psychologists. For example, MacKay (1963) has noted that "an artifact capable of receiving and acting on information about the state of its own body can begin to parallel many of the modes of activity we associate with self-consciousness" (p. 227). An important exception to this general rule is the work of Powers (1973, 1978), whose background includes training in both psychology and physics. Even Powers's writing, however, has been concerned predominantly with psychological processes other than those involved in social behavior.

that attention may be said to be directed either outward toward the environment or inward toward the self.

This seemingly simple statement (which is identical to the first assumption made by Duval and Wicklund, 1972) requires several kinds of qualification and elaboration. First, it should be clear that the dichotomy between self-focus and environment focus is an imperfect one (as in some respects is the distinction between what is self and what is nonself). When one's attention is focused on the environment, one is attending to immediate perceptual input from the distance receptors: the eyes and ears. Self-focus, in comparison, includes a wider variety of possibilities. When attention is self-directed, it sometimes takes the form of focus on internal perceptual events, that is, information from those sensory receptors that react to changes in bodily activity (for example, the autonomic and proprioceptive activity that contributes to the subjective experience of emotion). Self-focus may also take the form of an enhanced awareness of one's present or past physical behavior, that is, a heightened cognizance of what one is doing or what one is like. Alternatively, self-attention can be an awareness of the more or less permanently encoded bits of information that comprise, for example, one's attitudes. It can even be an enhanced awareness of temporarily encoded bits of information that have been gleaned from previous focus on the environment; subjectively, this would be experienced as a recollection or impression of that previous event.

It is sometimes difficult to maintain the clarity of this dichotomy, as is illustrated by the following example. To look at a part of one's own body is, initially, to focus on the environment. The visual perception is a perception of a thing out in the world, until that percept is categorized as an aspect of self. Because the visual stimulus in question is a part of one's body, however, it is likely that it will ultimately be categorized as a component of self. At that point the experience may become one of self-attention, as one mentally considers that self-aspect. It is this mechanism by which experimental manipulations such as mirrors presumably act to heighten self-focus: by inducing visual perceptions that remind subjects of themselves in a very general sense, that is, remind them of self-dimensions other than their visual representations.

It should also be clear that the term *heightened self-awareness* and its synonyms as they are used in this article do not necessarily connote a long, deep examination of the self. Rather, such terms imply an increase in the probability or the frequency with which attention is momentarily directed to some aspect of the self. In research, such increases in self-focus are caused by stimuli that remind subjects of themselves. It is assumed, however, that attention normally shifts back and forth between self and environment to some degree, whether such reminders are present or not.

2. *When attention is directed to the environment, incoming stimulus information is processed, leading to classification of that input.* The process of classifying incoming stimuli has been studied extensively by cognitive psychologists. Posner (1969; see also Posner & Rogers, 1978), Franks and Bransford (1971), Reitman and Bower (1973), and Neumann (1974) are among the theorists who have developed models to account for the classificatory process in various ways. Though these models vary in other respects, they have as a unifying idea the notion that, over a succession of perceptual experiences, a single representation in memory is abstracted from a series of relatively similar percepts. This single representation, then, stands for a class of perceptual events. These abstracted representations—sometimes termed *recognitory schemas* or *recognitory prototypes*—are stored in memory and are subsequently used to classify and interpret new inputs.

Evidence for the existence and use of recognitory prototypes in classifying stimuli in the environment comes from a vast accumulation of research in cognitive psychology. In cognitive research, the prototypes are typically generated experimentally by presenting subjects with a series of discrete training stimuli that vary along several dimensions. After training, subjects are found to classify new stimuli according to the organization implicit in the training series. Presumably the catego-

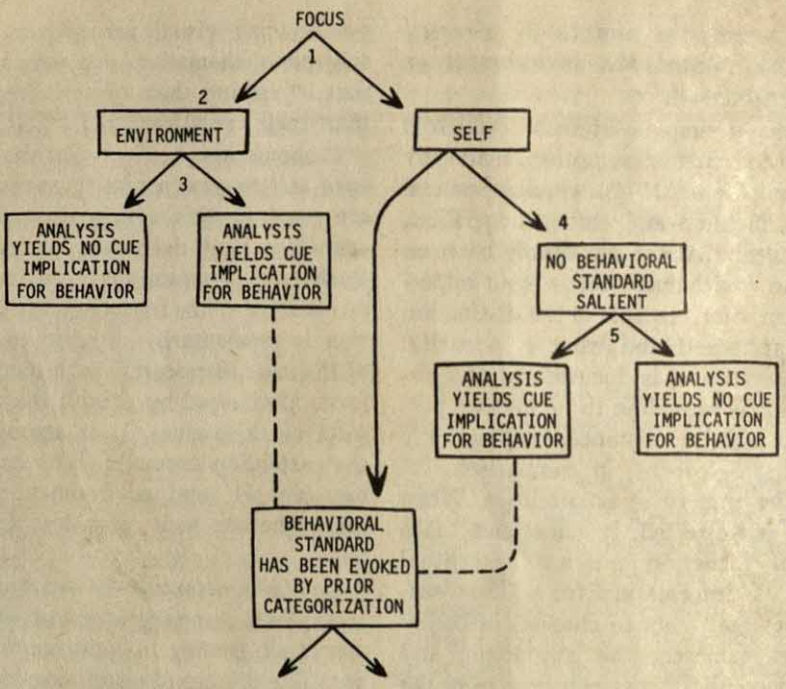


Figure 2. Flow-chart description of the cybernetic model of the consequences of self-directed attention. (Numbers in the figure refer to comparable propositions in the text.)

rizational structures that we carry with us in our day-to-day lives similarly reflect our experiential histories, but histories that are much longer and more complex.[5]

In the restricted settings of cognitive research, the subject categorizes only one stimulus at a time. However, in the broader context of social behavior one doubtless categorizes the nature of the interaction setting as a whole, as well as categorizing single stimuli. The kinds of information contributing to this overall identification undoubtedly include gestural, postural, and other nonverbal cues emitted by persons in the visual field; significant physical elements in the situation (e.g., presence of objects with identifiable uses); and verbally transmitted information, including inferences that the listener makes from the verbalization. Each cue that is encoded presumably contributes something to the resultant categorization of the context.

The work of a number of environmental psychologists (e.g., Barker, 1968; Fredericksen, 1974; Gump, 1971) would seem to fit very easily into this line of thought. Those researchers have emphasized that behavior settings vary along many discriminable dimensions, and they have worked toward the development of taxonomies of the situations in which behaviors take place. Presumably their efforts in categorization are similar to the processes that we all use continuously, albeit less consciously. In a similar vein, Cantor and Mischel (1977) have recently found evidence that personality-trait terms act very much as recognitory prototypes in the classification of other persons.

The classification of environmental inputs, which is assumed in the present model to

[5] The specific processes by which such schemas are formed and altered, although obviously very important, are outside the scope of this article. Thus, the experimental literature of that area is treated quite peripherally here. Inasmuch as this article is intended to deal primarily with consequences of self-directed attention, discussion will emphasize research with direct relevance to the propositions that deal with self-focus.

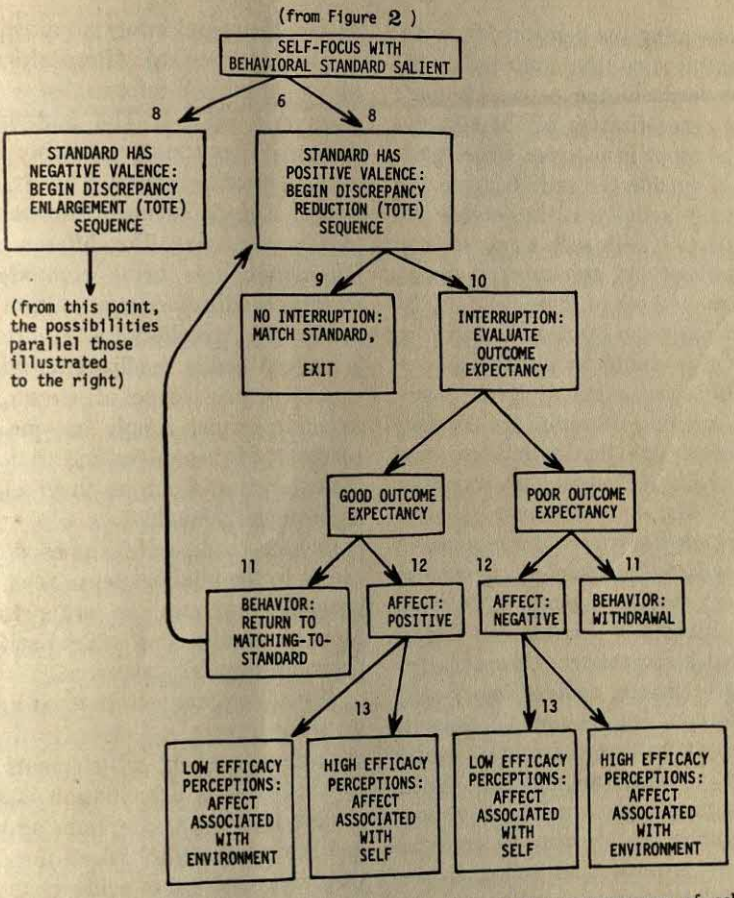


Figure 3. Flow-chart description of the cybernetic model of the consequences of self-directed attention. (Numbers in the figure refer to comparable propositions in the text.)

occur whenever attention is directed outward to the environment, is presumed to be an interactive process. That is, it involves a comparison of the incoming information with preexisting recognitory schemas.[6] Indeed, many believe that the resulting percept is as much a function of the previously encoded information as it is a function of the input information itself (cf. Neisser, 1976). The fact that this categorization involves such an interaction between input and schema suggests another qualification that must be made regarding the proposed dichotomy between self-focus and environment focus. Specifically, if processing and classification of even a very simple sensory input occurs by reference to preexisting recognitory prototypes, it would appear to be an oversimplification to say that attention is ever directed "at" the environment alone. The prototypes are, in a very basic sense, an aspect of oneself. The processing of environmental information thus seems to involve the simultaneous utilization of an aspect of the self.

At our level of analysis, however, the dichotomy still remains useful and meaningful. When internally stored schemas are accessed for comparison with environmental input during normal perception, this accessing occurs preattentively. It is the resulting percept that is represented in consciousness, not the evoking of the schemas. With regard to conscious experience, then, it seems quite reasonable for us simply to ignore the role played by the

[6] The perceptive reader may have noted that this activity in itself represents a TOTE unit at a lower level of analysis than is being emphasized in this paper.

schemas in generating the percept. Moreover, it bears repeating that the function of the sequence that leads to the percept is quite obviously the categorization of the stimulus input—and the input is an event whose point of initiation is outside the self. Because this is the goal of the activity, it thus seems reasonable to refer to it with such terms as *focus on the environment* and *processing* of *incoming information*. In both of these respects, the proposed dichotomy can be viewed as a verbal shorthand that is useful in classifying conscious experience at a macro level (where most of our present concerns are located) even while recognizing that at a micro level much more complicated events are occurring.

3. *In some cases, response-prototype information either constitutes part of a recognitory prototype or is directly implied by the recognitory prototype; in other cases, this information is not contained or implied* (see Figure 2). Saying that a recognitory schema used to identify one's behavioral context does specify response-prototype information is exactly equivalent to saying that the context evokes in the person a behavioral standard. As an example, most people in our culture have been taught to dampen the intensity of their behavior—that is, to speak and walk quietly—in places that are classified as religious in nature. When a person with this behavioral schema classifies a behavioral setting as being religious, quietness is immediately evoked as a behavioral standard.

Recognitory schemas do not always contain this kind of response-specifying information, however. In social psychological terms, many situations exist in which no particular standard of appropriate behavior is salient, where guidelines for action are not implicit in one's recognition of the nature of the context. Whether a situation has or has not evoked a response prototype as a behavioral standard has important implications for what occurs when attention is self-directed.

4. *If a behavioral standard has not already been evoked by classification of one's context, attention directed to the self leads to increased awareness of salient self-elements* (Figure 2). Stated another way, the analysis and categorization that were postulated above to occur for environmental information when attention is directed outward (Proposition 2) occur with regard to self-information when attention is directed inward. The underlying components of this process presumably are the same as those discussed above. That is, comparisons are being made between information and recognitory schemas. The difference is that the information now being compared with the schemas is self-information—information from within (e.g., recollections of prior behaviors or present bodily feedback activity). Experientially, in this type of circumstance, self-attentive persons simply become increasingly cognizant of themselves and their own salient characteristics, much as they might examine any external stimulus.

In such cases, self-focus is not evaluative except in the minimal sense that preattentive analysis "evaluates" in order to categorize. More specifically, this state of self-focus is not phenomenologically aversive.

The notion that self-directed attention may in some cases lead simply to heightened awareness of salient self-elements rather than to a self-*critical* examination is a major departure from the assumptions on which Duval and Wicklund (1972) based their theory. Yet there is a good deal of evidence to support this assertion. This evidence will be reviewed in the following paragraphs.

Duval and Wicklund (1973) themselves demonstrated that more self-attentive subjects make greater self-attributions for hypothetically experienced outcomes than do less self-attentive subjects. This outcome has since been replicated (Buss & Scheier, 1976), although other research suggests some limitations on the effect (Hull & Levy, 1979). It may be that the mere heightened salience of self as an entity led subjects in the Duval and Wicklund (1973) and Buss and Scheier (1976) studies to weight the self more heavily in the judgments they made. Alternatively, perhaps the salient aspect of self in this attribution-making context is one's potential as a causal agent (cf. Duval & Wicklund, 1973).

The reasoning that salient self-aspects may be more fully represented in consciousness when self-focus is high than when it is low has also been extended to the internal experi-

ence of emotional states (Scheier & Carver, 1977). In that research, it was proposed that when one's attention is self-focused in a context in which there is affect, one may attend to that affect as being the salient component of self. Thus, in two of Scheier and Carver's studies, subjects who had experienced positive and negative mood inductions subsequently reported feeling more elation and depression, respectively, when more self-attentive than when less so. In two additional studies, subjects reported experiencing greater attraction and repulsion toward pleasant and unpleasant slides, respectively, when self-attentive than when not. Others (Borkovec & O'Brien, 1977) have found similarly that directing a subject's attention to his or her bodily responses leads to an increase in the self-reported intensity of the emotion being experienced.

The notion that heightened self-focus leads to enhanced awareness of one's internal state suggests another interesting possibility: that self-focus can lead to heightened awareness of the *absence* of anticipated internal activity. This hypothesis was recently tested (Gibbons, Carver, Scheier, & Hormuth, 1979), in a study in which some subjects were led to expect arousal symptoms from a pill, whereas others were not. In all cases the pill was a placebo. Among subjects who expected arousal, fewer symptoms subsequently were reported by self-focused subjects than by those with less self-focus.

The Gibbons et al. (1979) finding appears to indicate that when led to expect a different internal state than is actually present, the self-attentive person has greater access to the veridical internal state than does the less self-attentive person. Two additional studies (Scheier, Carver, & Gibbons, in press-a) have investigated further this heightened awareness of internal experience among self-attentive persons by assessing whether it can serve to reduce other kinds of suggestibility phenomena. In the first of these studies, subjects were exposed to stimuli of moderate sexual attractiveness, which they were instructed to evaluate on the basis of their own bodily reactions. Subjects were led, however, to anticipate either highly arousing or very nonarousing stimuli. Subjects with experimentally heightened self-attention were less misled by these cues than were subjects in whom self-focus had not been increased. The second study extended the reasoning to the experience of taste. Subjects were led to expect either an increase or a decrease in a flavor intensity (relative to a previous sample); they then received either a slightly stronger or a slightly weaker solution. The intensity judgments subsequently made by highly self-attentive subjects were less in line with manipulated anticipations and more in line with actual flavor intensities than were judgments made by less self-attentive subjects. Thus, again, self-focus minimized suggestibility, apparently by increasing awareness of an actual internal experience.

Nor is the increased accuracy of self-report that occurs under high self-focus limited to reports of internal perceptual experiences. Other studies (Pryor, Gibbons, Wicklund, Fazio, & Hood, 1977; Scheier, Buss, & Buss, 1978) indicate that self-focus enhances people's accuracy in reporting on their habitual behavioral tendencies. Thus, for example, self-reports of sociability that were made under conditions of heightened self-attention were found to be highly correlated with unobtrusive measures of actual social behavior; reports made by less self-attentive subjects, on the other hand, poorly represented their actual behavior (Pryor et al., 1977). Presumably, self-attention provides more veridical or more thorough access to the memories that are relevant to such self-reports. Additional evidence that self-focus increases the activation of self-relevant memory areas has been provided by Geller and Shaver (1976).[7]

---

[7] In addition to determining ease of access to previously stored information, the direction of one's attention probably also has implications for the access one has to schemas for use in encoding new stimuli (whether they be self-stimuli or environmental stimuli). Thus, for example, when one is self-focused, one may not have good access to the schemas that are typically used for encoding information about the environment (cf. Vallacher, 1978). Indeed, easy access to a schema may increase its use, even in situations where such use would not otherwise be expected. This may be the mechanism behind the "top of the head" phenomena discussed by Taylor and Fiske (1978), in which some particularly salient

A final illustration of the fact that self-attention can lead simply to greater cognizance of the self comes from research conducted by Carver and Scheier (1978) as validation of self-awareness manipulations. Subjects in that research completed a self-focus sentence completion blank (Exner, 1973) developed and validated earlier as a measure of egocentrism. This instrument was completed either in an empty room or in the presence of a self-awareness-inducing stimulus. As was expected, proportionally more self-focus sentence completions were emitted in conditions of heightened self-awareness than under control conditions.

Salience. One final issue, implicit in much of the above discussion, probably should be addressed at this point. What determines which aspect of the self is "salient" when attention is self-directed and which aspect of the environment is salient when attention is directed to the environment? A full and complete answer to this question cannot yet be given. Doubtlessly the ultimate answer will be complex. However, one determinant of salience probably derives directly from the interactive nature of the categorization process described above. As was noted, classification of the environment takes place by comparison of input against recognitory prototypes. If the prototypes to which the environmental inputs correspond have cue implications for some specific self-information, the self-information is probably accessed (in a preliminary preattentive fashion) by the very process of classifying the environmental input. The aspect of self that is represented by that self-information presumably will be more salient than other self-aspects when attention is subsequently directed inward. Another way of saying this is that when we attend to the self, we are often examining an aspect of self that has been suggested by some cue in our environmental context. For example, an environmental stimulus that is categorized as *typically emotion-inducing* may cue a search for bodily arousal as a salient self-dimension when attention is self-directed. Indeed, such a cue implication may serve in vivo as an impetus to shift focus from the environment to the self.

Similarly, the self-aspect that has just been under examination may be a major determinant of the aspect of the environment that is salient when attention is directed outward. For example, consider a case in which self-attention has had as its object an existing state of bodily arousal for which there is no recognized antecedent. In such a case, it seems likely that attention subsequently directed outward would be oriented specifically to searching for aspects of the environment that are categorizable as potentially arousing. Thus a self-aspect can have a cue implication for some aspect of the environment, leading to heightened salience of that aspect when one's focus turns outward. More simply, when we attend to the environment, we are often focusing on aspects of the environment that have been suggested by some cue within ourselves.

5. *Just as categorization of environmental stimuli sometimes evokes a behavioral prototype as a standard, a categorization of a salient self-aspect may in some cases evoke a behavioral standard* (see Figure 2). As an example, a salient emotion might evoke a prototypic response *to* that emotion. The research presented above regarding awareness of affect was limited to reports of subjective emotional experiences. Other research has demonstrated, however, that heightened awareness of an affect may also lead to increased behavioral responsivity to the affect. For example, Scheier (1976) found that provoked aggression was more intense among more self-focused than among less self-focused subjects.[8] Other research (Scheier, Carver,

stimulus exerts an inordinate influence on subjects' behavior. This notion might also have some relevance to the recent finding that imagining an event increases the degree to which a person expects the event to occur (Carroll, 1978).

[8] Scheier (1976) found that subjects' personal attitudes toward aggression did not correspond to the aggression they actually emitted. However, an attitude is only one type of standard that is potentially applicable in that context. The issue of varieties of standards will be addressed more fully below. The position taken here with regard to the Scheier (1976) data is that awareness of anger evoked a stereotypic retaliatory response prototype and that the prototype was used as a comparison value for subsequent behavior.

Schulz, Glass, & Katz, 1978) has demonstrated that feelings of sympathy are expressed to a greater degree among more self-attentive than among less self-attentive persons. Finally, yet other studies (Scheier, Carver, & Gibbons, in press-b) have shown that strong fear leads to greater avoidance among more self-attentive than among less self-attentive persons.

6. *A standard can have either a positive or a negative valence.* A positive valence implies that the standard is taken as a "desired" goal state. A negative valence exists when a standard of comparison is taken as an "undesired" goal state. Valences presumably are based on previous learning and generalization and on temporary encoding through imitation and verbal transmission. Multiple or component valences can easily exist with respect to any specific standard, of course. Consider, for example, the teacher's behavioral standard of going from the office to teach a class. On a given day the teacher may feel that he or she should go to the class because teaching is his or her vocation (positive valence); the teacher may also feel like not going because the teacher does not enjoy the class (negative valence). One important component valence, of course, occurs as a function of an obligation to do the behavior in question (cf. Vickers, 1973).

It seems likely that any given component valence can also vary in salience over time. Its salience is likely to depend on the salience of the environmental cues to which the valence is attached. For example, if bright, interested students have just been in the teacher's office, a positive valence may be salient. Presumably the accumulation of these component valences (weighted by salience) determines the overall valence of the standard as discussed here.

7. *Standards and behavioral dimensions can also vary in importance.* There are at least two ways in which importance may be construed in informational terms. The first depends on association of positive or negative events with the stimulus class during the person's prior experiences. A *priority label* may become part of the categorization as a function of the intensities of such events. High intensity events (e.g., strong punishment for poor behavior with respect to a standard) would lead to a high priority tag or "importance"; low intensity events would lead to a tag of less priority.

A second possibility concerns the centrality or interrelatedness of a behavioral dimension with respect to other dimensions. Importance in this case would be defined according to the number of cross-reference tags attached to a given behavioral dimension that refer to other behavioral dimensions. For example, the amount of work that one person does may be weighed in his or her mind as important, but its importance does not extend to other areas of the person's life. The work, in effect, is left at the office. In contrast, another person may not only see his or her work output as important but may also view it as having implications for other parts of his or her life, for example, interpersonal relationships. The work output dimension thus would be effectively more important for the second person than for the first, by virtue of its interconnectedness with other dimensions.[9]

8. *If a behavioral standard has been evoked either by environmental input or by a salient self-element, subsequent self-focus leads to a comparison between self (one's present behavior or characteristics) and the previously evoked standard. Comparison between self and a positively valenced standard leads to discrepancy reduction; comparison between self and a negatively valenced standard leads to discrepancy enlargement* (see Figure 3). The comparison between self and standard is regarded as representing the test phase of a psychological TOTE unit. It thus leads to the engagement of one or the other of two possible adjustment and feedback sequences: either a negative feedback loop (discrepancy reduction) or a positive feedback loop (discrepancy enlargement).

The notion that heightened self-focus leads

[9] Standards can also vary in their degree of acceptedness, which is logically distinguishable from importance, even though the two in reality are often confounded. A standard that one accepts half-heartedly will be less reflected in behavior than one accepted enthusiastically.

to increased conformity of one's behavior to a salient standard was, of course, a main tenet of self-awareness theory, although for quite a different reason than is assumed here. Self-awareness theory held that discrepancy reduction was one of several potential responses to the postulated aversiveness of self-attention. The present model assumes that the discrepancy-reduction impulse is affect free and is the prepotent response to the recognition of a discrepancy. These distinctions are important ones and will be elaborated more fully in a later section of the paper.

Discrepancy reduction. Whatever its basis, the prediction that self-directed attention leads to heightened conformity to the salient behavioral standard has received ample research support. For example, a series of studies of instrumental aggression showed that subjects matched their behavior to salient standards (which varied from study to study) to a greater degree when made more self-aware than when less self-aware. In the first of these studies (Scheier, Fenigstein, & Buss, 1974), self-focus caused male subjects to conform to a "chivalry" norm when shocking a female victim who had not provoked them. In a second study (Carver, 1974), female subjects shocked at higher levels when self-focus was increased, in line with an experimentally induced standard in which high shock was portrayed as desirable. In yet other studies (Carver, 1975), subjects' aggressive behavior was more consistent with their previously verbalized attitudes when the subjects were more self-attentive than when less so.

Moreover, demonstrations of the fact that self-focus leads to increased conformity to standards are by no means limited to studies of aggression. For example, in an experiment in which speed of copying prose was made salient as a standard of comparison (Wicklund & Duval, 1971), more prose was copied within a given time span by subjects in whom self-focus was heightened than by those in whom it was not. In addition, compared to less self-aware controls, self-aware subjects have been found to behave more consistently with their previously assessed stage of moral reasoning (Froming, in press) and more consistently with their previously assessed levels of sex guilt (Gibbons, 1978).

Discrepancy enlargement. The notion that standards can have negative valuation and that focus on those standards leads to discrepancy enlargement has received much less attention than has the matching-to-standard process. However, this notion is logically consistent with an older idea in social psychology. Newcomb (1958) wrote of the "negative reference group," a group with which one does not identify, but with which one actively compares oneself. Such a comparison allows one to emphasize, and indeed to further increase, the existing differences. Newcomb coined the term negative reference group in the context of social comparison theory (Festinger, 1950, 1954). Yet the nature of the process that Newcomb proposed is strikingly similar to the process proposed above for cases in which some negatively valenced behavioral standard is salient. The difference is that in the present model the comparison value need not be the attitudes or behavior of a particular group of people. It can be any negatively valenced value that is taken up for purposes of comparison. (The relationship between the present theorizing and social comparison theory will be examined in more detail in a later section of the article.)

Another class of behavior that can be viewed as discrepancy enlargement is a phenomenon that is usually analyzed in terms of psychological reactance (Brehm, 1966; Wicklund, 1974). Researchers have demonstrated repeatedly that coercive attempts to persuade lead to resistance to persuasion or even to attitude change in a direction opposite to the position being advocated (e.g., Brehm & Brehm, 1966; Snyder & Wicklund, 1976). Such effects may be interpreted as attempts to distance oneself from the standard of comparison (i.e., the position that is advocated in the communication). In the present model that standard would acquire a negative valence in very much the same manner as was postulated by Brehm (1966)—that is, by virtue of the threatening elements contained in the associated communication. Consistent with the reasoning that distancing-from-standard might occur under these circumstances

in response to self-focus, Carver (1977) has found that attitude reversal in response to a coercive communication was greater among self-focused subjects than among less self-focused subjects. Similar results have occurred in two additional studies (Carver & Scheier, in press).

Standards. Standards for self-comparison can vary widely, both in terms of their specific nature and in terms of their source. One reasonable distinction in this regard would be between the *type* of a standard and its *content*. For example, smoking in response to stress is different in content from eating, or pacing in response to stress. Yet they are all similar in one respect: They are of the type called *habit*. A personal-preference attitude is a different type of standard. A third type is an attitude based on more complex reasoning, for example, moral reasoning. Similarly, an instruction from another person that is adopted and conformed to is different in type from all these, as is the temporarily assumed socially desirable behavior that others nearby are exhibiting. With these multiple possibilities of source, type, and content, what determines which is evoked in a given situation and is thus taken as the comparison value when one is self-focused?

Specific applications of this general question abound in personality and social psychology. For example, the literature of moral development is replete with illustrations of the fact that moral reasoning is a judgmental *capability* rather than a typical way of behaving. Under which conditions will a person respond to a moral choice by exerting his or her psychological capabilities to their fullest, and under which conditions will the person's behavior be dictated by other considerations? The answer to this question is by no means clear, for the following reason. Although categorization processes and the use of categorizational schemas to classify and interpret information have recently begun to receive attention from social psychologists (e.g., Cantor & Mischel, 1977; Kuiper & Rogers, 1979; Markus, 1977; Rogers, Kuiper, & Kirker, 1977; Rogers, Rogers, & Kuiper, 1979), thus far virtually no work has been aimed explicitly at understanding the process of extracting behavioral prototypes in response to environmental cues. Which standard becomes salient in a given experimental situation has usually been more a matter of good intuitions on the part of the researcher who is preparing the situation than a matter of precise analysis of the determinants of salience. Quite obviously, this entire question is an important area for future work. For the present, let us simply note that matching-to-standard effects appear to have been demonstrated with regard to all of the following: habits[10] (Liebling, Seiler, & Shaver, 1974), personal-preference attitudes (Carver, 1975), moral reasoning capabilities (Froming, in press), and experimentally provided instructions (Carver, 1974).

9. *Matching-to-standard is the normal consequence of self-attention when a behavioral standard is salient, until and unless the process is impeded in some way.* (Although this and the following propositions will be framed in terms of matching-to-standard, it should be clear that the same logic is applicable to distancing-from-standard phenomena; see also Figure 3). Throughout a discrepancy-reduction sequence, self-focus represents a recurrence of the *test* mechanism of the psychological TOTE unit. *Test* will continue to alternate with one or more *operate* mechanisms until behavior or present state is brought into line with the standard. As in any TOTE se-

---

[10] The fact that matching-to-standard apparently can occur with regard to habits as standards raises the additional question of how "conscious" the process is in such cases. Matching to a habit standard may feel very unconscious, phenomenologically, for the following reason. The degree to which execution of a behavioral sequence seems conscious depends on the salience of its component acts and their consequences; this in turn depends on the degree to which those acts are monitored while being carried out. In habitual activity, the behavioral components have previously been encoded as a well-connected chain (cf. Schank & Abelson, 1977, discussion of cognitive "scripts"). To use TOTE terminology, a single *operate* consists of a behavioral sequence rather than a small bit of behavior change. Therefore, fewer tests are required over the course of the behavior than is the case for other, less well-known behavioral sequences. The relative paucity of attention devoted to monitoring this behavior suggests that it may be experienced less completely in consciousness.

quence, if successful completion of matching-to-standard occurs, it is followed by disengagement.

This proposition stands in direct contradiction to the assumption in recent statements of self-awareness theory (Wicklund, 1975a, 1975b) that the prepotent response to self-focus in a standard-salient context is avoidance of self-awareness rather than discrepancy reduction. As was noted above, this contradiction has important ramifications that will be considered further in a later section of the paper.

10. *If something impedes the matching-to-standard process, behavior is interrupted and an assessment process is evoked* (see Figure 3). This assessment entails the further processing of relevant information, yielding an outcome expectancy: an estimate of the likelihood of being able to more closely approximate the standard, based on the nature of the situation and on the behaviors available to the person. This assessment process can occur prior to the initiation of behavior if, for example, the person has foreknowledge that the behavior is going to be difficult to execute successfully. Alternatively, it can occur during the matching-to-standard sequence if the person encounters difficulty in either selecting or executing the appropriate operation to move toward the standard.

The assessment of outcome expectancy may be an evaluation of the likelihood of attaining the overall goal, given the context and one's resources. Alternatively, if the matching-to-standard sequence involves separable steps, it may be an assessment of the likelihood of attaining the next step. In either case, the assessment process has two products: One is behavioral, the other affective. These will be considered in turn.[11]

11(a). *If the assessment process leads to a favorable outcome expectancy, the behavioral response is a return to the matching-to-standard sequence* (Figure 3).

11(b). *If examination of the context and one's resources reveals a low subjective probability of being able to alter behavior appropriately, the behavioral consequence is withdrawal.* The latter would appear to represent a disengagement from the dimension in question, a kind of going-out-of-the-field phenomenon (cf. Lewin, 1935). In practice, this could take the form of a mental dissociation from the dimension (refusal to consider further) or could occur as a behavioral withdrawal from contextual elements that either heighten self-attention or make the standard salient.

There are several types of events that can potentially prevent or disrupt the matching-to-standard process. One possibility is that the behavior in question is fixed in the past. If there is no opportunity available to repeat, reattempt, or undo the behavior, it is thus rendered unchangeable. In such a case, discrepancy reduction cannot even begin. An example of this state of affairs occurred in a study conducted by Duval, Wicklund, and Fine (1972). Some subjects in that study were informed that their scores on a previously administered test of intelligence and creativity were very discrepant from the levels that they had anticipated and desired. Moreover, it was implicit that the test was not going to be repeated, nor was any other opportunity to be provided for reducing the discrepancy (thus all subjects would have had low outcome expectancies with regard to matching the behavior to standard). In this study, subjects who were to wait for another experimenter after receiving the negative feedback withdrew sooner if the room contained a self-attention-heightening stimulus than if it did not.

Results of a more recent study (Steenbarger & Aderman, in press) confirm that this withdrawal tendency depended on the fact that the behavioral discrepancy was portrayed as unalterable. This study made use of a paradigm derived from that of Duval et al. (1972), but one in which there was ostensibly to be a second task. Steenbarger and Aderman (in press) found that when subjects experi-

[11] This proposition and the two following propositions, though derived independently, have a good deal in common with Stotland's (1969) analysis of *hope* as a psychological construct. Stotland used the term *hope* to refer to the perception of a high probability of attaining a goal and the term *anxiety* to refer to the perception of low probability of attaining a goal.

enced an inflexible discrepancy—that is, when the impression was created that subjects could not expect to alter their performances appreciably—self-awareness hastened withdrawal. When the impression had been created that improvement was a real possibility, on the other hand, no increase in withdrawal occurred as a function of self-attention.

Intrapersonal deficit. A second set of conditions that may lead to behavioral interruption stems from the possibility that the execution of the behavior in question requires specific skills, abilities, or characteristics that the person is not certain he or she possesses. The implication that these skills are required may cause the discrepancy-reduction attempt to be suspended and cue the assessment process. An example of the importance of this possibility is provided by recent studies of fear-based behavior conducted by Carver and Blaney (1977a, Experiment 3; 1977b). In those studies, some subjects were chosen explicitly on the basis of their doubts about their ability to cope with fear, whereas others were chosen as believing that they could cope with their fear. All of these subjects, however, reported having identical (moderate) degrees of fear. It seems reasonable that signs of fear while attempting to approach and pick up the feared stimulus (in this case, a snake) should cue a self-assessment process in all such subjects. In the Carver and Blaney (1977b) study, the subjects who doubted their abilities to overcome their fear (i.e., who had chronically low outcome expectancies) reacted to the perception of rising fear, induced via false heartbeat feedback, by avoiding attending to the behavior–goal comparison (according to self-reports) and by withdrawing behaviorally. The latter effect replicated a previous finding (Carver & Blaney, 1977a, Experiment 3). In contrast, those subjects who had expressed more confidence about being able to go through with the behavior (i.e., who had chronically higher outcome expectancies) responded to the perception of anxiety by focusing *more* on the behavior–goal comparison and attempting—successfully—to match the one with the other.

The Carver and Blaney (1977a, 1977b) studies investigated the behavioral consequences of variations in expectancy, but those studies did not explicitly incorporate experimental manipulations of self-focus. More recently, however, a similar study has been conducted using a conventional self-awareness manipulation (Carver, Blaney, & Scheier, 1979). In that experiment, confident subjects were found to be undeterred by heightened self-awareness during the approach task, but doubtful subjects withdrew earlier in the approach sequence when self-focus was high than when it was lower. Postexperimental self-reports formed a pattern that was consistent with the present model, that is, that heightened self-attention had led to increased awareness of fear among *all* subjects and thus to self-assessment during the approach sequence; that self-assessment, in turn, had led to continued approach attempts among confident subjects and to early withdrawal among doubtful subjects.

This result has recently been conceptually replicated (Carver, Blaney, & Scheier, in press) in quite a different behavioral domain: i.e., responses to failure. All subjects in that research were confronted with a failure experience on an intellectual task, in order to create a large self-versus-standard discrepancy. Subjects then undertook a second task, ostensibly bearing on the same intellectual skill. Rather than assessing subjects' chronic expectancies, however, as had been the case in the Carver et al. (1979) study, an experimental manipulation of outcome expectancy was introduced. Subjects were led to believe either that they could potentially do quite well on the second task, or that they would probably do quite poorly. In reality, the second task was a measure of persistence. As predicted, subjects in whom unfavorable expectancies had been induced were less persistent when self-focus was high than when it was low. Also as predicted, subjects with favorable expectancies were more persistent when self-focus was high than when it was low.

The tasks in all of the above research provided subjects with escapable situations. That is, it was possible for subjects engaged in task attempts in those studies to withdraw physically from the behavioral context. This pos-

sibility does not always exist, however, when a perceived self-deficiency threatens one's behavioral adequacy. For example, consider the person with severe test anxiety who must take an unavoidable test. While in the testing situation, such persons probably are very susceptible to being interrupted from their performance attempts by stimuli that cue outcome assessment. In many test situations, this assessment yields unfavorable expectancy judgments. Because they cannot withdraw behaviorally, however, these persons may be frozen in the self-assessment phase of the sequence, where they repeatedly reconfront the evidence of their own inadequacy. Attempts at task-relevant activity are sporadic because outcome assessment indicates that such attempts will be unsuccessful, and the attempts are short-lived because the assessment process is likely to be re-evoked. Thus the test-anxious person is caught in a cognitive loop in which self-assessment and cognitive disengagement from the task predominate, rather than renewed effort.

This portrayal is consistent with Wine's (1971) characterization of the test-anxious person as being chronically self-focused in situations where the adaptive response is to be task focused. The present model seems more adequate than is Wine's, however, in two respects. Though Wine's analysis would not necessarily predict that self-focus can facilitate performance among persons low in test anxiety, the present model clearly does. Such performance facilitation has been demonstrated elsewhere (Wicklund & Duval, 1971). The present model also implies that self-focus can facilitate performance even among highly test-anxious persons, if circumstances lead those persons to have positive outcome expectancies. This possibility does not seem derivable from Wine's (1971) theory. Recent research (Slapion & Carver, Note 1) has yielded evidence that this can occur as well.

Environmental constraints. Yet a third set of conditions that would lead to behavioral interruption is the possibility that some feature of the environment may prevent the successful execution of the appropriate behavior (a circumstance that fits the usual operational definition of frustration). To use a somewhat trivial example, if one goes shopping at a time when stores are closed, one typically withdraws from the attempt. This withdrawal does not occur, however, until one's attempt has been interrupted and the realization has occurred that the desired outcome is unlikely (i.e., low outcome probability). The notion that environmental constraints lead either to withdrawal or to attempts to circumvent them seems a very obvious point. This case is discussed here largely in order to point out that the psychological processes implicit in this class of events are the same as those that occur in the other event classes discussed above. This indicates the high degree of generality of the processes under consideration.

Novelty. The analysis presented here of processes leading to withdrawal also bears some resemblance to Berlyne's (1963) discussion of the somewhat perplexing fact that novelty sometimes leads to approach, sometimes to withdrawal. He proposed that the presence of some degree of "collative" (i.e., comparison-requiring) properties in a stimulus (novelty, for example) engages exploration, movement toward the stimulus in order to assimilate it. Too much of the same properties, however, leads to withdrawal. According to Berlyne, the reason for this withdrawal is that the organism assesses the alterations required to assimilate the stimulus as being beyond its present capabilities. In effect, outcome expectancy with regard to assimilation is negative, and withdrawal is the consequence.

12(a). *The judgment that one cannot alter one's behavior in the direction appropriate to the standard may also lead to negative affect in proportion to the importance of the behavioral dimension or the standard and the perceived magnitude of the discrepancy.*[12] This negative affect remains present as long as the person focuses on the judgment that he

[12] It seems worth reemphasizing that importance here refers to importance as defined by oneself. A standard that others view as important but that seems irrelevant to oneself is for present purposes *not* an important standard.

or she cannot reduce the discrepancy. Evidence for this proposition comes from a study (Steenbarger & Aderman, in press) that was discussed above in connection with the assumption that self-focus leads to withdrawal only when discrepancy reduction is prevented in some way. That same study also showed that subjects who had been led to believe that they could not reduce the discrepancy experienced more negative affect when self-focus was high than when it was lower. This effect did not occur among subjects who expected to be able to reduce the discrepancy.

12(b). *Engagement of the assessment process can also lead to positive affect, if assessment leads to the perception of a favorable outcome probability.* This is comparable to Simonov's (1970) argument that emotion becomes positive when "there is a surplus of information available as compared with the information necessary to satisfy a sufficiently strong need" (p. 147). This is the elation experienced, for example, when one knows that one knows *how* to arrive at the solution of a problem, even though the problem has not yet been solved.[13]

The assumption of two potential responses (behavioral and affective) to the assessment of outcome expectancy raises a question as to their relationship. The two may be parallel responses. For example, both the cessation of behavioral attempts and the experience of negative affect may be cued by a recognition that nothing can be done to improve one's present state. This would be similar in some respects to Leventhal's (1970) parallel response model of reactions to fear communications. An alternative possibility is that behavioral withdrawal may occur as a response to the existence of negative affect. It is not possible to specify the relationship unequivocally at this time. It seems clear, however, that environmental constraints on potential behavior will have a large role in determining which of the two response modes predominates, particularly in cases of unfavorable outcome expectancy. That is, when the environment permits withdrawal, withdrawal should predominate. When the environment does not permit withdrawal, negative affect should predominate.

13. *Affect that results from assessment of outcome expectancy can take one of two forms, depending on the cognitions associated with the affect* (see Figure 3). What cognitions are linked to negative affect will be determined by the person's perception of the locus of the barrier to behavior alteration. If the person has a low *efficacy* expectancy—that is, if he or she perceives that behavior is being impeded by a self-deficit—the affect will be associated with the self. It may be reflected in such ways as reduced self-esteem. If the person has high efficacy expectancy and perceives that his or her behavior is impeded by some environmental constraint, the affect will be associated with those constraining elements. It may be reflected by such things as expressions of resentment.[14] Similarly, what cognitions are associated with positive affect will be determined by the perceived locus of the cause of favorable outcome expectancy. The affect will be associated with the self if personal efficacy is seen as having caused the good outcome to be likely; it will be associated with the environment if the positive outcome expectancy is ascribed to a benign environment. Although the logic behind this proposition is relatively straightforward, these predictions remain to be tested.

Comparisons With Self-Awareness Theory

At least three questions deserve further consideration with regard to differences between the present model and self-awareness theory as proposed by Duval and Wicklund (1972)

---

[13] This analysis may help to account for the observation that the anticipation of a pleasant event often seems more pleasurable than the event itself, as if knowing that the outcome will occur is better than its occurrence. Indeed, this would be consistent with the frequently expressed observation among psychologists that "the fun is over" once an idea has been conceived and a study to test it is designed.

[14] It may be this variable that distinguishes, for example, between depression attributed by a client to environmental contingencies and depression attributed to inadequacies of the self. That is, insufficiently positive outcome expectancies lead to depressed affect, but whether cognitions of self or environment are associated with the negative affect may vary with efficacy expectancies.

and amended by Wicklund (1975a). They are the following: (a) Under what conditions is self-attention aversive? (b) What instigates or energizes the matching-to-standard process? (c) Finally, what is the prepotent response to self-focus when a behavioral standard is salient—discrepancy reduction or withdrawal? These issues will be discussed in the following three sections.

Aversiveness of Self-Focus

One important area of conflict between the present model and that of Duval and Wicklund concerns whether negative affect is an inevitable consequence of self-attention, and if not, what other conditions are necessary for it to occur. Duval and Wicklund (1972) and Wicklund (1975a, 1975b) have argued that any time a discrepancy exists such that one's present state or behavior is worse than the standard of comparison, self-attention is aversive.[15] "It is assumed that the objectively self-aware person will not simply react to himself impartially and in a neutral manner, but that he will come to evaluate himself as soon as the objective state occurs" (Duval & Wicklund, 1972, p. 3). "Certainly objective self-awareness is postulated to be an aversive, motivational state" (Wicklund, 1975b, p. 80). Indeed, as was stated at the beginning of the present article, the postulated aversiveness of self-attention was that theory's crux. In contrast, it has been proposed here that self-directed attention leads to negative affect only when the person perceives that he or she cannot alter his or her present state in the direction of the standard. (This discussion will be framed in terms of discrepancy reduction, but it is of course applicable to discrepancy enlargement as well.)

It is instructive to note that the studies cited as evidence of the unpleasantness of self-focus were invariably cases in which a within-subjects discrepancy existed *and in which subjects were prevented from altering their present state or behavior*. For example, in the first demonstration that subjects would attempt to escape from self-focus (Duval et al., 1972), the subject's behavior on the relevant dimension was anchored in the past and thus could not be altered in the direction of the standard. Similarly, though self-esteem has been found to be reduced when subjects were made self-aware (Ickes, Wicklund, & Ferris, 1973), it is unclear to what degree that finding depended on the nature of the study's dependent measure. The dependent measure was a comparison between subjects' self-rated real selves and ideal selves along several dimensions. Subjects in that research had no way to *alter* their real selves. They were simply reminded of the real–ideal discrepancy to a greater degree when self-awareness was heightened than when it was not.

Aside from these examples in which subjects' behavior or characteristics were unchangeable, there appears to be no evidence that self-attention itself is phenomenologically aversive. To the contrary, Steenbarger and Aderman (in press) found an increase in negative affect only when self-focus was combined with a nonreducible discrepancy; when the discrepancy was flexible, the opposite tendency occurred. Similarly, Carver and Scheier (1978), in demonstrating that the presence of either a mirror or an audience increases self-focus, were unable to find any evidence that self-attention led to negative affect. A similar conclusion has also been reached by other researchers (Davis & Brock, 1975; Hull & Levy, 1979).

Affect and valuation. There is only one restricted sense in which the present model implies the existence of something resembling negative affect whenever self-attention is directed to a behavior–standard comparison. It may be useful in this regard to draw a distinction between the phenomenological experience of affect and the function of valuation (cf. Diggory, 1966, p. 70). The event that initiates the matching-to-standard sequence according to the present model is the fact that the *test* of the TOTE unit assesses

[15] Wicklund (1975a) has also pointed out that the existence of a positive discrepancy (i.e., self better than the standard) is a different case, leading in his view to positive affect. The present discussion, however, will be limited to instances of negative discrepancy, as has implicitly been the case throughout the article.

the existence of a discrepancy between a present state and an accepted goal state. Inasmuch as the function of the behavior change (*operate*) is to reduce the discrepancy, it is possible to construe the recognition of the discrepancy as being (on some level) a negatively evaluated event. However, there seems to be no reason to assume a concomitant phenomenological experience of negative emotion. As was noted in the preceding paragraphs, however, there is substantial reason to believe that negative affect occurs in cases when one is prevented from reducing those discrepancies (Duval et al., 1972; Steenbarger & Aderman, in press). Thus the data seem consistent with the assumption in the present model that self-focus has phenomenologically affective consequences only following outcome assessment.

To the degree that negative affect is experienced in contexts in which behavior change is (objectively) not difficult, it may be seen as a variation of the inability-to-alter experience. For example, a workman who repeatedly reaches to the wrong spot for tools may feel negative affect, not because of failure to reach the tool, but because the inability to correct the initial reaching impulse is beginning to alter his perceptions of self-efficacy with regard to tool handling. Similarly, a man at a formal dinner who suddenly realizes that he has been eating his salad with the wrong fork can easily alter his behavior. Any negative affect he experiences is caused by concern that his behavior may be regarded by others as uncultured and the possibility that he will not be able to alter the first impression that he has thus created. Finally, the perception of the need for a discrepancy reduction that will take an extraordinarily long time can lead to negative affect. However, this should occur only if there are salient doubts as to the ultimate outcome. If there is a positive outcome expectancy—*hope,* in Stotland's (1969) terms —positive affect should ensue, even if many months of behavior remain between the person and his or her goal.

Instigation of Matching-to-Standard

There is a second important distinction between the present model and Duval and Wicklund's (1972) theory that is intimately bound to the question of whether self-focus is aversive. This second distinction concerns the question of how behavior change is instigated and energized when self-focus occurs and a behavioral standard is salient. As was mentioned above, Duval and Wicklund (1972) devised their model from a drive-theory perspective. By postulating that self-focus is aversive any time a within-self discrepancy exists, those authors could assume that subsequent behavior change was energized by a drive state and was instigated by that drive state's aversiveness.

In the present model, in contrast, the matching-to-standard sequence that is engaged when self-focus occurs in the presence of a salient behavioral standard is seen as the realization within a psychological system of a negative feedback loop. The fact that self-attention is required in order to engage the matching-to-standard sequence is interpreted in control-theory terms in the following way. Self-focus in a standard-salient context represents the *test* phase of the TOTE unit: an assessment of whether a discrepancy exists. The behavioral response to self-focus (the alteration of behavior in the direction of the standard) is the *operate* phase. The ordering of the control sequence of the feedback loop dictates that *operate* cannot take place before *test* reveals a discrepancy. Thus, focus on a discrepancy is required to engage the TOTE sequence. In phenomenological terms, this awareness simply corresponds to the realization that one is *in* a context in which there is a standard to match with one's behavior. A change in behavior depends on that realization.

Inasmuch as both models assume a comparison process between self and standard, the critical difference between them in this context seems to be the presence or absence of a drive postulate. This issue introduces considerable theoretical complexity, stemming from two facts: first, that a great many variants of drive as a construct have been proposed over the years, and second, that there is far from universal agreement on what measurable state, if any, corresponds to a theoretical state of heightened drive. A completely

adequate treatment of this complexity is well beyond the scope of the present article. The interested reader is referred to Appley's (1970) review article for a more detailed critique of drive and arousal constructs. Consideration of the issues here will be much more limited.

Does self-attention heighten drive, or does it not? One might at first view this as an empirical question, although it will be argued below that this appearance may be illusory. However, let us consider the existing evidence on the relationship between self-focus and drive. Although agreement on the nature of drive is often difficult to obtain, as was noted above, many researchers have assumed that drive is either equivalent to, or reflected by, autonomic arousal.[16] To the limited degree that evidence on this point is available, self-attention appears not to lead to an increase in physiological arousal. If one were willing to equate such arousal increases with heightened drive, it would thus appear that self-focus is not drive inducing. Evidence on this issue comes from two sources.

Salience of nonarousal. One set of data comes from a study (Gibbons et al., 1979) that was discussed earlier as evidence that self-attention when no standard is salient merely makes one more conscious of one's salient characteristics. Subjects in that study were led to anticipate that arousal symptoms would occur as a function of a pill (actually a placebo). Subjects later reported the extent of their symptoms, with self-attention heightened or not. If self-directed attention per se were arousal inducing, greater arousal should have been experienced, and thus reported, by subjects in the high self-awareness condition than by control subjects. Instead, the opposite occurred. Self-attentive subjects apparently became more conscious of the *absence* of arousal (consistent with the present informational analysis) and consequently reported fewer symptoms.

One difficulty with applying this finding to the drive question, however, is the fact that in the early statement of their theory, Duval and Wicklund (1972) largely ignored the possibility that a standard of comparison might not always be present. Wicklund (1975a) later concluded that self-focus was aversive only if one's present state was worse than a salient standard of comparison. However, he argued at the same time that one was almost always worse than some standard and that one would ultimately become conscious of that discrepancy. Nevertheless, if self-attention should be drive inducing only when a discrepancy exists between self and standard, and if it were possible to assume a case within Duval and Wicklund's framework in which no standard is salient, then those authors might thus anticipate that self-focus would not increase drive in the Gibbons et al. (1979) case.

Physiological data. A second set of data on the arousal question does exist, however (Paulus, Annis, & Risner, 1978). Moreover, this finding is not liable to the caveat that was applied above to the Gibbons et al. finding. Paulus et al. gave subjects a task (copying prose) with a clear standard (copying quickly). As subjects prepared to attempt the task, either before a self-focusing stimulus or with no such stimulus present, the experimenter took a physiological measure called the palmar sweat index (Johnson & Dabbs, 1967). Increased palmar sweat values have commonly been believed to reflect increased arousal (cf. Geen & Gange, 1977; Martens, 1969a, 1969b). However, in the Paulus et al. study, *lower* values on this index were found among more self-aware than among less self-aware subjects. The meaning of the palmar sweat index is open to considerable question on other grounds (see the following paragraphs). But if that measure were acceptable as an index of drive, the finding of Paulus et al. would seem to contradict Duval and Wicklund's assumption that self-attention is drive increasing.[17]

---

[16] Though it certainly oversimplifies the case to imply that all theories utilizing a drive construct contain this assumption, it is probably fair to say that it is implicit in the large majority of drive models in social psychology.

[17] The data of Paulus et al. were derived in the context of a social facilitation paradigm. The present theoretical model has rather extensive implications for social facilitation phenomena, although consideration of those implications is beyond the scope of this article. These implications are, how-

It was noted above that the arousal or drive issue is not entirely an empirical one. There are also two important logical considerations that render the entire arousal controversy somewhat less meaningful.

The meaning of physiological indices. First, contemporary researchers in physiological psychology tend to recoil at the mere suggestion that autonomic arousal is a unitary phenomenon. Such psychologists are inclined to look at autonomic indicants in terms of what specific functions they imply rather than whether the organism is aroused. For example, Sokolov (1963) has proposed that one pattern of physiological responses comprises an "orienting" response, whereas another pattern comprises a "defensive" response. The orienting response is held to facilitate greater knowledge of an external stimulus by making the organism more receptive to input information from that stimulus. The defensive response, in contrast, involves an inhibition of sensitivity to input information (cf. also Hare & Blevings, 1975). This latter function presumably protects the organism from overstimulation.

This argument is easily taken one step farther, by assuming that such function-specific physiological response patterning is a normal concomitant of different kinds of *consciously controlled* behavioral involvement. Williams, Bittker, Buchsbaum, and Wynne (1975) have discussed this possibility in some detail (see also Lacey, 1967). They reasoned that experimental tasks vary in the types of attentional demands they make (as opposed to the degree of demand). Some tasks (e.g., a visual decoding task) require sensory intake; others (e.g., mental arithmetic) require sensory rejection. Conceptually, the sensory intake pattern seems similar to Sokolov's (1963) orienting response; the sensory rejection pattern, though not particularly "defensive," seems to embody many functional components of Sokolov's defensive reaction, in that there is a suppression of sensory input. In the Williams et al. analysis, however, this pattern simply reflects the attempt to concentrate focus inward and prevent distraction. In a test of this reasoning, Williams et al. (1975) were able to show that subjects differed as predicted in cardiovascular response patterns as a function of the type of task being performed.

It is instructive to note, in this regard, that initial theorizing about the palmar sweat index (Dabbs, Johnson, & Leventhal, 1968) concerned attentional issues fully as much as it concerned arousal. Dabbs et al. (1968) argued that increased palmar sweat was associated with readiness to engage the environment, and that decreased values reflected self-directed attention and an attempt to concentrate. These descriptions seem a good deal like the sensory intake and sensory rejection patterns of Williams et al. (1975). Taken together, these two descriptions converge on two notions, one of them general and the other more specific. The general one is that at least some physiological indices may be more usefully construed as providing information about attentional and information-processing phenomena than as providing information about arousal. The more specific one is this: that the finding of decreased palmar sweat in the presence of a self-attention-inducing stimulus (Paulus et al., 1978) may most reasonably be interpreted in attentional terms, as reflecting inward focus of attention and the simultaneous suppression of environmental input.

Cybernetics and the energizing of behavior. In addition to constraints imposed by the complexity of the meaning of autonomic responses, another issue of logic should be addressed. This issue concerns the nature of cybernetic theory.

Among persons accustomed to think of behavior as being "energized" and thus impelled by some kind of drive, the idea that a control-theory model of motivation is truly adequate to explain variations in the intensity of behavior will seem difficult to accept. This is true even though a number of motivational psychologists (e.g., Guilford, 1965; Hunt, 1965; Taylor, 1960; Vickers, 1973) have for a number of years characterized the TOTE

---

ever, taken up elsewhere (Carver & Scheier, Note 2). That paper also presents further data on the role of drive in social facilitation and self-awareness phenomena.

unit as the most promising model of motives leading to intentional activity.

The sticking point for many people seems to be the obvious fact that it takes energy to do work. Thus a drive model would seem to be preferable to a control-theory or a cybernetic one. But this view fails to recognize the fact that physiological changes that occur in the course of executing a behavior are in no way incompatible with cybernetic assumptions. Cybernetic theory would simply relegate the observed physiological change to a secondary causal role. As Guilford (1965) has noted, "In automated devices, as in living organisms, there is a clear distinction between two functions of energy. There is the energy that carries out the various operations, where work must be done, and there is the energy that has control functions, where triggering and release of much greater quantities of energy are involved" (p. 320). Cyberneticists would tend to emphasize the importance of the energy that has control functions, even though it is the work-doing energy that is most readily observed in terms of changes in physiological state. Thus Wiener argued long ago (1948) that the then-current focus on arousal was misguided, that the body is in fact very far from being a conservative system with limited energy available. He argued that the appropriate type of "bookkeeping" for events in the nervous system is not one that is energy based. Instead, the critical notions were those of message, quantity of information, coding technique, and so on.

It should be repeated, however, that the fact that physical work systems are subsidiary components of the body's control system has this important implication: Covariation of physiological change with behavioral change in no way constitutes a threat to a control-theory analysis of behavior regulation. In such an analysis, a change in physiological activity leads merely to the question of what specific functions are being performed, rather than the more misleading question of whether arousal exists. Indeed, for this reason it is arguable (see also Carver & Scheier, Note 2) that control theory provides a more subtle understanding of physiological events than does drive theory.

Prepotent Response Tendency

There is a third distinction between the Duval and Wicklund (1972) model and the present model that should also be addressed. It bears on the following question. In a context in which a behavioral standard is salient, which response is prepotent when self-focus occurs: matching-to-standard or withdrawal? In the model proposed here, the matching-to-standard sequence is prepotent, as part of the human being's normal self-regulatory system. Withdrawal occurs only after an outcome assessment and only if that assessment yields a negative expectancy. In contrast, although Duval and Wicklund (1972) originally felt that it was not possible to know which of these two modes would be preferred, Wicklund later argued for an ordering opposite to that assumed in the present model. "Certainly a successful averting of self-focused attention would eliminate the negative affect, however temporarily; thus an individual's immediate reaction to objective self-awareness should be an avoidance of self-focusing stimuli and/or efforts to find distractions" (Wicklund, 1975a, p. 236). Thus, according to Wicklund, discrepancy reduction occurs only when self-focus cannot be avoided.

Results of two recent studies (Carver, Blaney, & Scheier, in press; McDonald, in press) suggest that the present analysis is the more accurate of the two. In McDonald's research, subjects in whom a large negative within-self discrepancy had been made salient were more persistent on a subsequent task (for which persistence had been defined as appropriate) when self-focus was increased than when it was not. Moreover, there was also evidence that this effect was exaggerated by instructions that the second task was relevant to the dimension on which the discrepancy had been created. The possibility that subjects had simply been distracting themselves from self-focus by burying themselves in their task was considered and rejected by McDonald, because of the fact that a much simpler and quicker means of avoiding self-focus was readily available but was not used, that is, announcing "I'm finished" and leaving the context. Thus McDonald con-

cluded that discrepancy reduction is prepotent under self-focus when a means of reducing the discrepancy is available.

The Carver et al. (in press) research was discussed earlier in this article, in the section on outcome expectancy. Some subjects in that research were led to have favorable expectancies of being able to reduce a large within-self discrepancy, others were led to have unfavorable expectancies. Subjects with favorable expectancies were more persistent when self-attention was high than when it was low, consistent with the McDonald (in press) results. Carver et al. rejected the possibility that this persistence represented attempted distraction for the same reason as had McDonald, that is, that there was an easier way to avoid self-focus readily available. Indeed, this point is further emphasized in the Carver et al. research by the fact that one group of subjects *did* utilize that way of avoiding self-focus—specifically, subjects with negative outcome expectancies. Thus those results appear to be unequivocal in supporting the present theory.

Other Theoretical Comparisons

The model proposed in this article also bears certain similarities to models in social psychology other than self-awareness theory. Three of the similarities are striking enough that they deserve some separate attention. The first of these concerns Bandura's recent (1977) analysis of fear-based avoidance behavior; the second concerns recent analyses of human helplessness (Abramson et al., 1978; Wortman & Brehm, 1975); the third concerns social comparison theory (Festinger, 1950, 1954). These similarities will be discussed in the following sections.

Self-Efficacy and Avoidance

The fact that the approach–withdrawal decision process postulated here has been framed in terms of outcome and efficacy expectancies suggests that some commonality may exist between the present model and Bandura's recent analysis of the role of self-efficacy expectancies in avoidance behavior (Bandura, 1977; Bandura, Adams, & Beyer, 1977). Bandura's analysis and the present one do share some common ground, but they also differ in important respects.

Terminology. The first difference is one of terminology (see Figure 4). Bandura used the term *outcome expectancy* to refer to "a person's estimate that a given behavior will lead to certain outcomes" (Bandura, 1977, p. 193). His primary concern was in distinguishing that belief from the expectancy that one can *execute* the behavior. This he labeled *efficacy expectancy*. Clinically, such a distinction is quite an important one. In many cases people know what behavior a situation requires but do not believe themselves capable of doing the behavior. Because efficacy expectancy was most clinically relevant and thus most important to Bandura, he devoted relatively little attention to defining and discussing outcome expectancy. Implicitly, however, outcome expectancy in Bandura's terms is based only on the presence or absence of knowledge about the normal consequences of a behavior. In the present model, in contrast, the term outcome expectancy is used to denote expectancy about an event's likelihood of occurrence. This usage quite explicitly includes a variety of inputs: knowledge of the normal consequences of task-appropriate behavior, constraints imposed by external forces or by the passage of time, and efficacy expectancy—one's judgment about whether or not one can execute the requisite behavior (Figure 4).

In the present model, therefore, outcome expectancy is the direct determinant of the person's subsequent behavior and affect (Propositions 11 and 12). Efficacy expectancy is merely one input into outcome expectancy, but one determining what cognitions are associated with affect experienced as a function of that outcome expectancy. In Bandura's model, in contrast, the person's efficacy judgment is the direct determinant of his or her subsequent behavior. Outcome expectancy can be seen as a partial—though often trivial—determinant of efficacy expectancy in Bandura's model (Figure 4) in that one cannot feel capable of doing a behavior unless one knows the behavior and what its ordinary consequences are.

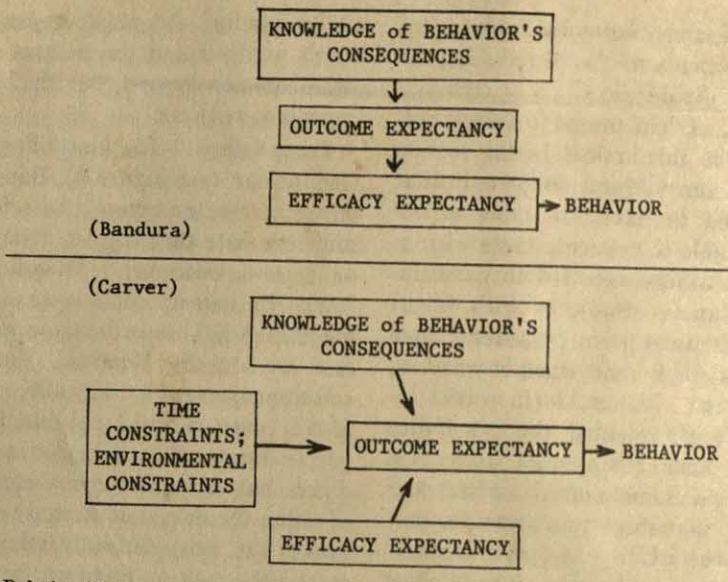


Figure 4. Relationships among outcome expectancy, efficacy expectancy, and behavior, as conceptualized by Bandura (1977) and Carver.

The models are in somewhat closer agreement with regard to the nature and determinants of efficacy expectancy. Bandura conceived of efficacy expectancy as being a product of four classes of information: prior performance accomplishments, vicarious experiences, verbal instructions, and emotional arousal. An additional set of relevant information would seem to be the degree to which one perceives that one can separate a behavior into components, each of which may be manageable by itself. These various sources of information (involving both previously encoded stimuli and presently experienced stimuli) are processed together, yielding a judgment of efficacy expectancy.

Role of emotional arousal. Even with regard to the determinants of efficacy expectancy, however, there is a point of conflict between the models: specifically, the role that is assumed to be played by the perception of emotional arousal. As was noted in the preceding paragraph, Bandura (1977) held that fear arousal was one kind of input into efficacy expectancy. Although this position has been moderated to take into account attributional contingencies (Bandura et al., 1977), there still appears to be a difference between Bandura's position and the position taken in the model proposed here. The present model holds that the role of perceptions of fear arousal in an approach attempt (initially, at least) is to cue the self-assessment process (see Figure 5). This assessment results in an outcome expectancy judgment, which for fear-based behavior is largely determined by efficacy expectancy (which derives, in turn, from the other informational sources enumerated by Bandura, 1977). This expectancy determines whether the person subsequently withdraws or returns to the approach attempt.

This analysis has received empirical support in three studies. Carver and Blaney (1977a, Experiment 3; 1977b) showed that persons who differed only in their chronic efficacy expectancies responded to false feedback of arousal in two quite different ways. Confident subjects focused on their approach attempt and were not debilitated behaviorally; doubtful subjects avoided focusing on the approach attempt and withdrew more quickly, compared to doubtful subjects given feedback of nonarousal. This interactive influence of fear perceptions on focus of attention and behavior—a set of findings that has subsequently been conceptually replicated using a self-awareness manipulation to heighten awareness of veridical fear (Carver et al., 1979)—does not seem easy to predict from Bandura's assumptions. The findings are, how-

ever, clearly predicted from the present model of self-attention processes.

Generality. A final useful comparison between the theories concerns their relative generality. Bandura's (1977) analysis was aimed quite explicitly at fear-related behavior and the therapeutic change of such behavior. In applying the two models to situations in which the likelihood of a positive outcome is entirely dependent on *intra*personal factors, which is the case in fear-based behavior, the two are functionally equivalent. This is true because outcome and efficacy expectancies as defined in the present model are perfectly confounded with each other in this particular class of situations. A consideration of broader contexts reveals, however, that the present model is more general in two respects than is Bandura's. First, because outcome expectancy is not held to be determined solely by knowledge of the task-appropriate behavior, the present model can generate predictions for two classes of behavioral interruptions not considered by Bandura: incomplete or inadequate behavior that occurred in the past and behavior constrained by environmental factors (see Figure 3). Secondly, since efficacy expectancy is held to exert an influence on behavior that is independent of outcome expectancy, the present model explicitly predicts that different affective experiences will be associated with similar behavioral events, based on differences in efficacy expectancy. That is, Proposition 13 above holds that if there is negative outcome expectancy, the presence of high efficacy perceptions predicts that affect will occur as resentment toward the environment; presence of low efficacy perceptions predicts negative affect associated with the self. If outcome expectancy is favorable, low efficacy perceptions predict that positive affect will be associated with the environment; high efficacy perceptions predict that the affect will be associated with the self. These

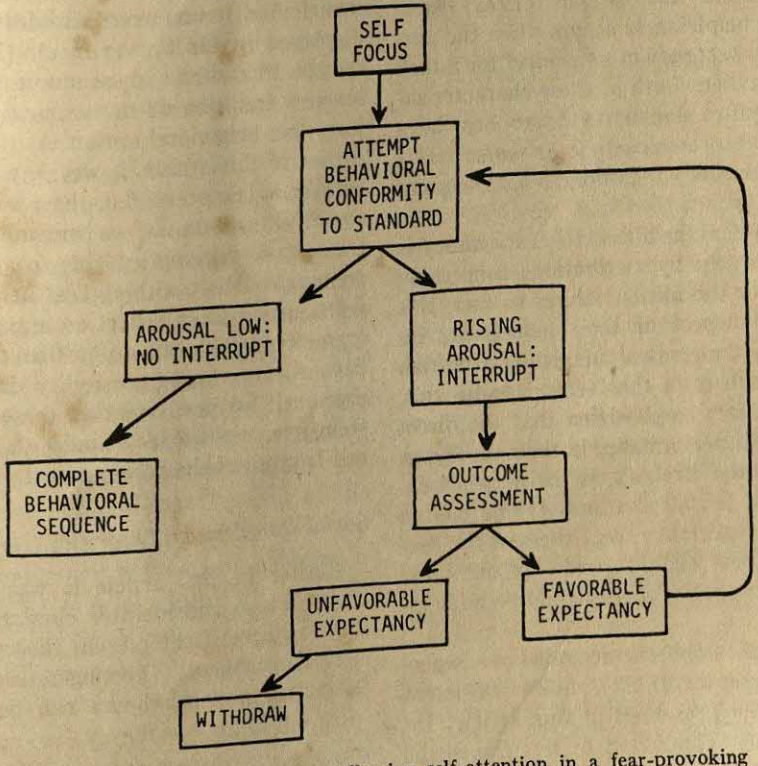


Figure 5. Information-processing sequence following self-attention in a fear-provoking behavioral context. (This sequence represents a specialized application of the model of behavioral regulation that was presented in Figure 3.)

predictions, which are now being tested, do not seem to be derivable from Bandura's (1977) theory.

Helplessness Theory

The present theoretical model also appears to have some implications for the area of research and theory known as "learned helplessness." Helplessness is a performance deficit that reflects presumed motivational, cognitive, or learning deficits resulting from fairly extended exposure to uncontrollable outcomes. There is at least one important similarity between aspects of the present model and the kind of cognitive analysis developed in recent years to account for helplessness effects among humans (e.g., Abramson et al., 1978; Wortman & Brehm, 1975). Abramson et al. (1978), for example, have proposed that the impact of uncontrollability on subsequent performance depends upon the development of an *expectation of future noncontingency*. Similarly, Wortman and Brehm (1975) have argued that helplessness occurs when the person has an *expectancy of no control* for a task being undertaken. Both of these characterizations are quite similar to what has been termed in the present article an *unfavorable outcome expectancy* regarding the subsequent task.

Abramson et al. held that this expectancy is influenced largely by attributions concerning the reason for the initial failure. In fact, this attributional aspect of their model has received by far the greatest amount of attention among researchers in that area. Despite this, however, it bears emphasizing that the direct determinant of performance in their theory, as in Wortman and Brehm's, is the person's expectancy, not the attributions. If one has an unfavorable expectancy, regardless of the reason, helplessness should result. If one's expectancy is favorable, there should be no helplessness.

The central emphasis accorded to expectancy by these theorists is quite consistent with the model presented in this article. Indeed, one study conducted to test the model (Carver et al., in press, discussed in detail earlier) could be regarded quite easily as a helplessness experiment. Subjects underwent a failure pretreatment, and some of them later displayed reduced persistence, which is often taken as a sign of helplessness. In providing support for the present model, that research thus also appears to offer support for other models of helplessness that rely on postulates about expectancies.

However, that research also suggests two respects in which the present model adds to such theories of helplessness. The more obvious contribution of this model stems from the fact that no other analysis of helplessness effects includes any consideration of the role of self-directed attention in promoting such effects. As is demonstrated in the Carver et al. (in press) research, that role may be quite an important one.

The second contribution of the present theory is its suggestion that the impulse to withdraw is basic to a wide variety of helplessness effects. The assumption of a withdrawal impulse is inherent in the present model. And it was overt withdrawal that was displayed in the Carver et al. (in press) research. In contrast, most studies in the helplessness tradition do not explicitly allow subjects this behavioral option. As was suggested earlier in this article, however, when physical withdrawal is prevented, the result may be a *cognitive* withdrawal—a mental dissociation from task attempts. This characterization seems consistent with typical helplessness effects, which often reflect an apparent unwillingness or inability to utilize task-relevant cues. An interesting possibility that should be examined further is that all these effects may stem from a thwarted impulse to remove oneself from the behavioral context.

Social Comparison Theory

Earlier in this article it was noted that there is also considerable similarity between some aspects of the present theory and social comparison theory (Festinger, 1950, 1954).[18] Social comparison theory can be viewed as

[18] Although this similarity exists for Duval and Wicklund's theory, as well as for the present model, this discussion will focus on the present model.

having three central assumptions. The first is that we tend to define much of reality—especially our social reality—by comparing our own opinions, reactions, characteristics, and capabilities with those of other people, usually termed *reference groups*. The second assumption is that the tendency to do this is exaggerated by conditions of ambiguity. The third assumption is that once consensus has been reached and some normative value implicitly established, then there is pressure toward conformity to this value among the members of the group.

Holding in abeyance the role of perceived ambiguity in the instigation of these processes, the elements of social comparison theory thus seem to fulfill two functions: first, to establish some value as a standard of comparison (by means of an implicit social consensus), and second, to increase behavioral conformity to that standard. These two functions are, of course, precisely the same as those assumed in a control-theory model of behavioral self-regulation such as the one proposed here. That is, a control-theory approach to motivation assumes two kinds of information-processing systems. The first of these is the system that analyzes and categorizes perceptual input, yielding a behavioral standard. The second system—a TOTE unit—regulates behavior with regard to that standard.

One way in which social comparison theory is similar to the model proposed here concerns the first stage of social comparison, in which a standard is established by implicit social consensus. This aspect of that model would seem to represent one important way in which a person can extract a behavioral standard through categorization of his or her environmental context (Propositions 2 and 3 of the present model). This may indeed be one of the most commonly used methods of choosing a behavior among human adults. In principle, however, it represents only a subset of a larger class of potential ways to determine behavior. Thus, it would seem that this facet of social comparison theory could be subsumed under the more general process of categorizing the nature of one's context (a process that certainly deserves much more attention than it has received in the present article).[19]

The second stage of the social comparison process, in which conformity pressure occurs, is also functionally similar to a major facet of the present model. The difference between the two is in the locus assumed for the impetus to conform. In the present model, the conformity tendency (via the normal matching-to-standard sequence) is construed as internally based. Discussions of social comparison theory, in contrast, often seem to assume the existence of external pressure to conform.

Finally, it seems worthy of brief note that a subsidiary aspect of social comparison theory holds that if a person cannot conform to the group's standard, there is a tendency for either the group or the person to withdraw from the other. This seems not unlike the present proposition that when one cannot match one's behavior to the standard—that is, when outcome expectancy is unfavorable—there is an impulse to withdraw from the attempt.

Summary and Concluding Comment

In the preceding pages, a cybernetic model of self-attention processes was presented, along with support for its propositions. Summarized briefly, self-focus in a context where no behavioral standard is salient leads to heightened cognizance of salient self-elements. Self-focus when a behavioral standard does exist leads to an automatic matching-to-standard sequence. This sequence is regarded as the realization within a psychological system of a negative feedback loop. If matching-to-standard is interrupted, an outcome assessment ensues. A favorable outcome perception

---

[19] The assumption in social comparison theory that ambiguity increases the tendency to undertake social comparison suggests another possibility not previously addressed in this article. That is, it may be a more general rule that ambiguity leads to increased information search, even when the context is nonsocial, in order to determine better what behavior is appropriate. This possibility certainly should receive more attention. If confirmed, it would further increase the degree to which these two analyses overlap.

leads to positive affect and/or to a return to matching-to-standard. An unfavorable outcome perception leads to negative affect and/or behavioral withdrawal.

Following presentation of the model, three specific points of conflict between this theory and self-awareness theory were addressed. In each of those cases, evidence was reviewed that indicated the present analysis to be the more internally consistent and parsimonious of the two. Comparisons also were undertaken between the present model and three other theories: Bandura's analysis of cognitive processes underlying fear-related behavior, helplessness theory, and social comparison theory. Each of those comparisons revealed the possibility for considerable integration between models.

There are many other potential links that could be developed between the present theory and information-processing ideas currently under investigation elsewhere in social and cognitive psychology. A fairly obvious example is the similarity between the propositions of the present model that concern salience of oneself versus the environment and research conducted on other kinds of salience phenomena (see, e.g., Taylor & Fiske, 1978). Although there simply is not the space to pursue those interrelationships here, the existence of such commonalities among theories suggests that the ideas on which the present model is based have a good deal of integrative potential. Indeed, this seems apparent even if one considers only the material discussed in this paper. The model of self-attention processes presented here has proven to be applicable to a wide range of phenomena, topics as diverse as taste and test anxiety, in a way that is internally consistent. It is to be hoped, moreover, that this perspective will prove to be useful in generating additional hypotheses, beyond those considered in the preceding pages. Thus might the specific ideas advanced here, and the concepts of control theory more generally, come to be seen as important tools for the analysis of a broad range of longstanding problems in human behavior.

Reference Notes

1. Slapion, M., & Carver, C. S. *Self-directed attention and facilitation of intellectual performance among persons high in test anxiety*. Manuscript submitted for publication, 1979.

2. Carver, C. S., & Scheier, M. F. *The self-attention-induced feedback loop and human motivation: A control-systems analysis of social facilitation*. Manuscript submitted for publication, 1979.

References

Abramson, L. Y., Seligman, M. E. P., & Teasdale, J. D. Learned helplessness in humans: Critique and reformulation. *Journal of Abnormal Psychology*, 1978, *87*, 49–74.

Appley, M. H. Derived motives. In P. H. Mussen & M. R. Rosenzweig (Eds.), *Annual review of psychology* (Vol. 21). Palo Alto, Calif.: Annual Reviews, 1970.

Apter, M. J. *The computer simulation of behavior*. New York: Harper Colophon, 1970.

Bandura, A. Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 1977, *84*, 191–215.

Bandura, A., Adams, N. E., & Beyer, J. Cognitive processes mediating behavioral change. *Journal of Personality and Social Psychology*, 1977, *35*, 125–139.

Barker, R. G., *Ecological psychology: Concepts and methods for studying the environment of human behavior*. Palo Alto, Calif.: Stanford University Press, 1968.

Berlyne, D. E. Motivational problems raised by exploratory and epistemic behavior. In S. Koch (Ed.), *Psychology: A study of a science* (Vol. 5). New York: McGraw-Hill, 1963.

Borkovec, T. D., & O'Brien, G. T. Relation of autonomic perception and its manipulation to the maintenance and reduction of fear. *Journal of Abnormal Psychology*, 1977, *86*, 163–171.

Brehm, J. W. *A theory of psychological reactance*. New York: Academic Press, 1966.

Brehm, J. W., & Brehm, M. L., In J. W. Brehm (Ed.), *A theory of psychological reactance*. New York: Academic Press, 1966.

Buckley, W. *Modern systems research for the behavioral scientist*. Chicago: Aldine, 1968.

Buss, D. M., & Scheier, M. F. Self-awareness, self-consciousness, and self-attribution. *Journal of Research in Personality*, 1976, *10*, 463–468.

Cantor, N., & Mischel, W. Traits as prototypes: Effects on recognition memory. *Journal of Personality and Social Psychology*, 1977, *35*, 38–48.

Carroll, J. S. The effect of imagining an event on expectations for the event: An interpretation in terms of the availability heuristic. *Journal of Experimental Social Psychology*, 1978, *14*, 88–96.

Carver, C. S. Facilitation of physical aggression through objective self-awareness. *Journal of Experimental Social Psychology*, 1974, *10*, 365–370.

Carver, C. S. Physical aggression as a function of objective self-awareness and attitudes toward punishment. *Journal of Experimental Social Psychology*, 1975, *11*, 510–519.

Carver, C. S. Self-awareness, perception of threat, and

the expression of reactance through attitude change. *Journal of Personality*, 1977, *45*, 501–512.

Carver, C. S., & Blaney, P. H. Avoidance behavior and perceived arousal. *Motivation and Emotion*, 1977, *1*, 61–73. (a)

Carver, C. S., & Blaney, P. H. Perceived arousal, focus of attention, and avoidance behavior. *Journal of Abnormal Psychology*, 1977, *86*, 154–162. (b)

Carver, C. S., Blaney, P. H., & Scheier, M. F. Focus of attention, chronic expectancy, and responses to a feared stimulus. *Journal of Personality and Social Psychology*, 1979, *37*, 1186–1195.

Carver, C. S., Blaney, P. H., & Scheier, M. F. Reassertion and giving up: The interactive role of self-directed attention and outcome expectancy. *Journal of Personality and Social Psychology*, in press.

Carver, C. S., & Scheier, M. F. Self-focusing effects of dispositional self-consciousness, mirror presence, and audience presence. *Journal of Personality and Social Psychology*, 1978, *36*, 324–332.

Carver, C. S., & Scheier, M. F. Self-consciousness and reactance. *Journal of Research in Personality*, in press.

Dabbs, J. M., Jr., Johnson, J. E., & Leventhal, H. Palmar sweating: A quick and simple measure. *Journal of Experimental Psychology*, 1968, *78*, 347–350.

Davis, D., & Brock, T. C. Use of first person pronouns as a function of increased objective self-awareness and prior feedback. *Journal of Experimental Social Psychology*, 1975, *11*, 381–388.

Diggory, J. C. *Self-evaluation: Concepts and studies*. New York: Wiley, 1966.

Duval, S., & Wicklund, R. A. *A theory of objective self-awareness*. New York: Academic Press, 1972.

Duval, S., & Wicklund, R. A. Effects of objective self-awareness on attribution of causality. *Journal of Experimental Social Psychology*, 1973, *9*, 17–31.

Duval, S., Wicklund, R. A., & Fine, R. L. Avoidance of objective self-awareness under conditions of high and low intra-self discrepancy. In S. Duval & R. A. Wicklund (Eds.), *A theory of objective self-awareness*. New York: Academic Press, 1972.

Exner, J. E., Jr. The self-focus sentence completion: A study of egocentricity. *Journal of Personality Assessment*, 1973, *37*, 437–455.

Festinger, L. Informal social communication. *Psychological Review*, 1950, *57*, 271–282.

Festinger, L. A theory of social comparison processes. *Human Relations*, 1954, *7*, 114–140.

Franks, J. J., & Bransford, J. D. Abstraction of visual patterns. *Journal of Experimental Psychology*, 1971, *90*, 65–74.

Fredericksen, N. Toward a taxonomy of situations. In P. M. Insel & R. H. Moos (Eds.), *Issues in social ecology: Human milieus*. Palo Alto, Calif.: National Press Books, 1974.

Froming, W. J. The relationship of moral judgment, self-awareness, and sex to compliance behavior. *Journal of Research in Personality*, in press.

Geen, R. G., & Gange, J. J. Drive theory of social facilitation: Twelve years of theory and research. *Psychological Bulletin*, 1977, *84*, 1267–1288.

Geller, V., & Shaver, P. Cognitive consequences of self-awareness. *Journal of Experimental Social Psychology*, 1976, *12*, 99–108.

Gibbons, F. X. Sexual standards and reactions to pornography: Enhancing behavioral consistency through self-focused attention. *Journal of Personality and Social Psychology*, 1978, *36*, 976–987.

Gibbons, F. X., Carver, C. S., Scheier, M. F., & Hormuth, S. Self-focused attention and the placebo effect: Fooling some of the people some of the time. *Journal of Experimental Social Psychology*, 1979, *15*, 263–274.

Guilford, J. P. Motivation in an informational psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 20). Lincoln: University of Nebraska Press, 1965.

Gump, P. V. The behavior setting: A promising unit for environmental designers. *Landscape Architecture*, 1971, *61*, 130–134.

Hare, R. D., & Blevings, G. Conditioned orienting and defensive responses. *Psychophysiology*, 1975, *12*, 289–297.

Hull, J. G., & Levy, A. S. The organizational functions of the self: An alternative to the Duval and Wicklund model of self-awareness. *Journal of Personality and Social Psychology*, 1979, *37*, 756–768.

Hunt, J. McV. Intrinsic motivation and its role in psychological development. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 20). Lincoln: University of Nebraska Press, 1965.

Ickes, W. J., Wicklund, R. A., & Ferris, C. B. Objective self-awareness and self-esteem. *Journal of Experimental Social Psychology*, 1973, *9*, 202–219.

Johnson, J. E., & Dabbs, J. M., Jr. Enumeration of active sweat glands: A simple physiological indicator of psychological changes. *Nursing Research*, 1967, *16*, 273–276.

Kuhn, A. *The logic of social systems*. San Francisco: Jossey-Bass, 1974.

Kuiper, N. A., & Rogers, T. B. Encoding of personal information: Self–other differences. *Journal of Personality and Social Psychology*, 1979, *37*, 499–514.

Lacey, J. I. Somatic response patterning and stress: Some revisions of activation theory. In M. H. Appley & R. Trumbull (Eds.), *Psychological stress*. New York: Appleton-Century-Crofts, 1967.

Leventhal, H. Findings and theory in the study of fear communications. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. (Vol. 5). New York: Academic Press, 1970.

Lewin, K. *A dynamic theory of personality*. New York: McGraw-Hill, 1935.

Liebling, B. A., Seiler, M., & Shaver, P. Self-awareness and cigarette-smoking behavior. *Journal of Experimental Social Psychology*, 1974, *10*, 325–332.

MacKay, D. Mindlike behavior in artefacts. In K. M. Sayre & F. J. Crosson (Eds.), *The modeling of mind: Computers and intelligence*. Notre Dame, Ind.: University of Notre Dame Press, 1963.

Markus, H. Self-schemata and processing information about the self. *Journal of Personality and Social Psychology*, 1977, *35*, 63–78.

Martens, R. Palmar sweating and the presence of an

audience. *Journal of Experimental Social Psychology*, 1969, *5*, 371–374. (a)

Martens, R. Audience effects on learning and performance. *Journal of Personality and Social Psychology*, 1969, *12*, 252–260. (b)

McDonald, P. J. Reactions to objective self-awareness. *Journal of Research in Personality*, in press.

Miller, G. A., Galanter, E., & Pribram, K. H. *Plans and the structure of behavior*. New York: Holt, Rinehart & Winston, 1960.

Neisser, U. *Cognition and reality*. San Francisco: Freeman, 1976.

Neumann, P. G. An attribute frequency model for the abstraction of prototypes. *Memory & Cognition*, 1974, *2*, 241–248.

Newcomb, T. M. Attitude development as a function of reference groups: The Bennington study. In E. E. Maccoby, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Holt, Rinehart & Winston, 1958.

Paulus, P. B., Annis, A. B., & Risner, H. T. An analysis of the mirror-induced self-awareness effect. *Bulletin of the Psychonomic Society*, 1978, *12*, 8–10.

Posner, M. I. Abstraction and the process of recognition. In G. H. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 3). New York: Academic Press, 1969.

Posner, M. I., & Rogers, M. G. K. Chronometric analysis of abstraction and recognition. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 5). Hillsdale, N.J.: Erlbaum, 1978.

Powers, W. T. *Behavior: The control of perception*. Chicago: Aldine, 1973.

Powers, W. T. Quantitative analysis of purposive systems: Some spadework at the foundations of scientific psychology. *Psychological Review*, 1978, *85*, 417–435.

Pryor, J. B., Gibbons, F. X., Wicklund, R. A., Fazio, R. H., & Hood, R. Self-focused attention and self-report validity. *Journal of Personality*, 1977, *45*, 513–527.

Reitman, J. S., & Bower, G. H. Storage and later recognition of exemplars of concepts. *Cognitive Psychology*, 1973, *4*, 194–206.

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology*, 1977, *35*, 677–688.

Rogers, T. B., Rogers, P. J., & Kuiper, N. A. Evidence for the self as a cognitive prototype: The "false alarms effect." *Personality and Social Psychology Bulletin*, 1979, *5*, 53–56.

Schank, R., & Abelson, R. *Scripts, plans, goals and understanding*. New York: Halsted, 1977.

Scheier, M. F. Self-awareness, self-consciousness, and angry aggression. *Journal of Personality*, 1976, *44*, 627–644.

Scheier, M. F., Buss, A. H., & Buss, D. M. Self-consciousness, self-report of aggressiveness, and aggression. *Journal of Research in Personality*, 1978, *12*, 133–140.

Scheier, M. F., & Carver, C. S. Self-focused attention and the experience of emotion: Attraction, repulsion, elation, and depression. *Journal of Personality and Social Psychology*, 1977, *35*, 624–636.

Scheier, M. F., Carver, C. S., & Gibbons, F. X. Self-directed attention, awareness of bodily states, and suggestibility. *Journal of Personality and Social Psychology*, in press. (a)

Scheier, M. F., Carver, C. S., & Gibbons, F. X. Self-focused attention and reactions to fear. *Journal of Research in Personality*, in press. (b)

Scheier, M. F., Carver, C. S., Schulz, R., Glass, D. C., & Katz, I. Sympathy, self-consciousness, and reactions to the stigmatized. *Journal of Applied Social Psychology*, 1978, *8*, 270–282.

Scheier, M. F., Fenigstein, A., & Buss, A. H. Self-awareness and physical aggression. *Journal of Experimental Social Psychology*, 1974, *10*, 264–273.

Simonov, P. V. The informational theory of emotion. In M. Arnold (Ed.), *Feelings and emotions: The Loyola Symposium*. New York: Academic Press, 1970.

Singh, J. *Great ideas in information theory, language, and cybernetics*. New York: Dover, 1966.

Snyder, M. L., & Wicklund, R. A. Prior exercise of freedom and reactance. *Journal of Experimental Social Psychology*, 1976, *12*, 120–130.

Sokolov, Y. N. *Perception and the conditioned reflex*. New York: Macmillan, 1963.

Steenbarger, B. N., & Aderman, D. Objective self-awareness as a non-aversive state: Effect of anticipating discrepancy reduction. *Journal of Personality*, in press.

Stotland, E. *The psychology of hope*. San Francisco: Jossey-Bass, 1969.

Taylor, D. W. Toward an information-processing theory of motivation. In M. R. Jones (Ed.), *Nebraska Symposium on Motivation* (Vol. 8). Lincoln: University of Nebraska Press, 1960.

Taylor, S. E., & Fiske, S. T. Salience, attention, and attribution: Top of the head phenomena. In L. Berkowitz (Ed.), *Advances in experimental social Psychology* (Vol. 11). New York: Academic Press, 1978.

Turing, A. M. Computing machinery and intelligence. *Mind*, 1950, *59*, 433–460.

Vallacher, R. Objective self-awareness and the perception of others. *Personality and Social Psychology Bulletin*, 1978, *4*, 63–67.

Vickers, G. Motivation theory—a cybernetic contribution. *Behavioral Science*, 1973, *18*, 242–249.

Wicklund, R. A. *Freedom and reactance*. Potomac, Md.: Erlbaum, 1974.

Wicklund, R. A. Objective self-awareness. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 8). New York: Academic Press, 1975. (a)

Wicklund, R. A. Discrepancy reduction or attempted distraction? A reply to Liebling, Seiler, and Shaver. *Journal of Experimental Social Psychology*, 1975, *11*, 78–81. (b)

Wicklund, R. A., & Duval, S. Opinion change and performance facilitation as a result of objective self-awareness. *Journal of Experimental Social Psychology*, 1971, *1*, 319–342.

Wiener, N. *Cybernetics: Control and communication in the animal and the machine*. Cambridge, Mass.: M.I.T. Press, 1948.

Wiener, N. *The human use of human beings: Cybernetics and society*. New York: Avon Books, 1954.

Williams, R. B., Bittker, T. F., Buchsbaum, M. S., & Wynne, L. C. Cardiovascular and neurophysiologic correlates of sensory intake and rejection. I: Effect of cognitive tasks. *Psychophysiology*, 1975, *12*, 427–433.

Wine, J. D. Test anxiety and direction of attention. *Psychological Bulletin*, 1971, *76*, 92–104.

Wortman, C. B., & Brehm, J. W. Responses to uncontrollable outcomes: An integration of reactance theory and the learned helplessness model. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 8). New York: Academic Press, 1975.

Received August 10, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 8, 1282-1292

Some Effects of Deception and Level of Self-Monitoring on Planning and Reacting to a Self-Presentation

Gregory C. Elliott
Department of Sociology
University of Maryland

This experiment addressed issues in impression management. First, this study investigated whether type of impression intended (accurate or fabricated) and level of self-monitoring (high or low) affect the amount of information about a target person that individuals would acquire, at some cost to themselves, prior to interacting with that person. As predicted, high self-monitors planning a fabrication purchased more information than high self-monitors planning an accurate impression or low self-monitors planning either type of impression. In addition, reactions of both participants to the actor's performance were analyzed. Impression type affected both actors' and targets' reactions, whereas self-monitoring affected only the targets' ratings.

Erving Goffman, in his seminal work on impression management (1959), has discussed how individuals present an impression of self that influences the definition of the situation that others come to formulate. It is in the various parties' best interest to induce others to attribute qualities to them that are supportive of their own interaction goals. People can do this by expressing themselves in a manner that communicates impressions intended to lead others to support their plans voluntarily. The current experiment focuses on actors' preparation for an upcoming encounter and reactions to their performance.

The experiment reported herein was conducted as part of a doctoral dissertation for the Department of Sociology, University of Wisconsin, Madison. The research was supported by National Institute for General Medical Science Grant GM-01526-11 and by the Small Grants Committee, Department of Sociology, University of Wisconsin.

Special thanks are due Shalom H. Schwartz and Elizabeth E. DeCourcy-Wernette for their patient counsel and constructive criticism. I also thank H. Andrew Michener, William Ickes, Jane Piliavin, Mark Snyder, John Fleishman, and the anonymous reviewers for their comments on earlier drafts.

Requests for reprints should be sent to Gregory C. Elliott, Department of Sociology, University of Maryland, College Park, Maryland 20742.

Acquiring Information

Actors planning for an interaction can benefit from information about the other participants. They can use this information to help them tailor their impression to fit the particular characteristics of the encounter. In considering the problems of creating and sustaining an impression, one should distinguish between impressions that the actor believes to be accurate self-presentations and impressions that the actor knows are false (fabrications, in Goffman's, 1959, terms). Information about others is helpful in conveying an accurate impression (if only to ensure accuracy), but it is even more important in conveying a fabrication. Whereas an accurate impression can be generated from habit and previous experience, actors planning a fabrication must be wary of the possibility of being discredited. Behavior must be carefully monitored and, when necessary, accentuated or suppressed. Normal, habitual behavior that usually enhances an accurately descriptive performance could betray deceiving actors. Fabricators must be sensitive to elements of the encounter that could threaten or enhance their performance. In particular, information about the other participants in the interaction can indicate the kind of performance most likely to succeed.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3708-1282$00.75

There is evidence that individuals differ in the importance they attach to information about other participants. Snyder's (1974, 1978, in press) work on self-monitoring tendencies indicates that some individuals show a greater attentiveness to external cues (i.e., information from the environment in which an interaction occurs) when judging the appropriateness of an impression and the effectiveness of a performance. Labeled high self-monitors, they are concerned with the situational appropriateness of their behavior during a social encounter. As a result, they show greater sensitivity to the characteristics of others involved in the interaction, whether revealed by self-presentation or by inadvertence.

On the other hand, low self-monitors do not rely so much on social comparison information. Although they may also be concerned with their behavior in an encounter, they tend to rely on internal cues (such as internalized norms or attitudes) to guide their self-presentation.

It follows that high self-monitors should be eager to acquire information about other participants in the interaction. This should be especially true if they plan to convey a fabrication. The challenge of having to contradict their own attitudes or beliefs without revealing the contradiction makes them more careful than usual about their performance. This greater care manifests itself in part in an increased interest in information about other participants. Actors can use relevant information about others to plan their fabrication more precisely and to increase the likelihood of success.

On the other hand, individuals attentive to internal cues (low self-monitors) should feel little need for information about other participants, regardless of the type of impression to be conveyed. Their behavior is guided by their own beliefs about what is necessary to convey the intended impression, be it accurate or fabricated. Although they may feel increased concern in planning a fabrication, it is manifested in a more thorough introspection for guides to action, not in a need for information about others.

Prior research on the acquisition of information about other participants in an upcoming interaction is sparse. Snyder (1974) found that, compared to low self-monitoring subjects, high self-monitors spent more time looking at a sheet of majority responses to a personality inventory that they were filling out while preparing for an upcoming discussion. This information was obtainable merely by turning one's head to see the majority response sheet taped to the wall near the subjects.

Note that this activity involved no real effort on the subjects' part, because the information was readily available. Information more critical to the actor may be more difficult to obtain and may involve some cost. Two studies (Eiser & Eiser, 1976; Eiser & Tajfel, 1972) investigated subjects' preferences for different kinds of information about an opponent in upcoming bargaining sessions. Subjects could find out how much each potential outcome would cost the opponent in resources expended and how much the outcome was valued in terms of resources gained. Each piece of information was available at a set price. Subjects purchased more information about the values the opponents placed on outcomes than on the costs; this value–cost difference was larger in competitive than in noncompetitive groups.

These studies indicate that individuals who expect to interact with another do seek out information about them. Although the bargaining experiments do not explicitly involve impression management, self-presentational concerns could be one aspect determining subjects' behavior. The present experiment investigated the acquisition of information by individuals prior to conveying a specific impression of themselves to another person. It differs from Snyder (1974) in that the information was available only at some cost to the actors. Unlike the studies conducted by Eiser and associates (Eiser & Eiser, 1976; Eiser & Tajfel, 1972), the present study investigates information purchasing in an explicit context of impression management.

The arguments presented above lead to the following hypotheses. First, because of their greater sensitivity to external cues for appropriateness of behavior in everyday en-

counters, high self-monitors will seek out more information about another participant, even at some personal cost, than will low self-monitors. Second, the type of impression planned should influence the magnitude of the discrepancy between the information acquired by high and low self-monitors. A planned fabrication increases one's concern for performance during a self-presentation, which is reflected in a greater attentiveness to cues of situational appropriateness. For low self-monitors, this means more attention is paid to internal states, such as norms or personality characteristics, than would occur when planning an accurate impression. High self-monitors, by contrast, would be led to a greater reliance on information from the environment than would be required for an accurate performance. A two-way interaction between self-monitoring and impression type is therefore predicted: Prior to an accurate performance, high self-monitors should acquire more information about another than will low self-monitors, but the difference in the amount acquired should be even greater when planning a fabrication.

Reacting to the Encounter

This study also investigated the participants' reactions to the encounter. Actors could conceivably react to their behavior during the interaction on two distinct levels: their own personal feelings during the encounter (were they comfortable, confident, self-conscious?) and their feelings about the performance they have just completed (were they successful, organized, persuasive?). Individuals may be able to separate critiques of their own performance quality from the feelings they had while performing, although the two are not necessarily independent. In addition, other participants in the interaction should be able to make judgments about the actor on the same dimensions of performance quality and apparent feelings.

In prior research, Ickes and Barnes (1977) found that when high self-monitoring subjects were paired with self-monitoring partners in an informal interaction, high self-monitors rated themselves and their partners as more self-conscious than did low self-monitors. Regarding others' impressions of an actor's performance, Weiler and Weinstein (1972) discovered that accurate impression managers were judged more convincing than were fabricators.

Individuals conveying an accurate impression should feel little stress during the interaction. They have only to be themselves to keep the interaction proceeding smoothly. Fabricators, however, are under much more stress. They must constantly be alert for signals that their performance is being discredited. This leads to the hypothesis that actors who have just conveyed a fabrication will reveal more negative feelings about themselves and their performance than will those who have conveyed an accurate impression. Similarly, observers will evaluate the fabricating actors' performance more negatively than that of accurate impression managers, largely because of nonverbal leakage (Ekman & Friesen, 1969).

High self-monitors check their performance during an interaction by judging the others' reactions to them and tailoring their self-presentation accordingly; low self-monitors do not make much use of the reactions of others. Because high self-monitors are especially aware of the reactions of others, they are more critical of their performance than are low self-monitors. The hypothesis is that they will react more negatively than will low self-monitors. On the other hand, greater attentiveness to reactions of others spurs the high self-monitors to correct any problems in their performance quickly. As a result, the hypothesis is that observers will evaluate the performance of high self-monitors more positively than that of low self-monitors.

Method

Overview

Subjects were led to believe that they were participating in a study of impression *formation*. They were asked to help the experimenter discover how people build impressions of others by conveying a specific impression about themselves to a target person (ostensibly a naive subject, but actually a confederate). This impression was either in harmony with their own attitude (accurate impression condition) or contrary to it (fabricated impression condi-

tion). Subjects were told that the experimenter would be studying how the target person built an impression of them and what elements of the impression were important in the target person's mind.

A successfully formed impression would net both participants a monetary bonus. Before beginning the encounter, subjects were given the opportunity to purchase information about the target person by giving up some of the standard subject fee. After some time to prepare themselves, subjects spent 10 minutes in an informal face-to-face discussion with the confederate. Subjects and confederates then completed a questionnaire on which they rated the subjects' performance and feelings during the encounter.

Subjects

Sixty-six female students at a large Midwestern university served as subjects. All subjects had previously filled out preexperimental questionnaires. Two subjects were dropped from the analysis. One did not understand the instructions. A second refused in advance to be paid for her participation, commenting that this meant the information made available to her would cost her nothing. Since her orientation to the experiment was affected by her refusal to be paid, her responses were not comparable to the others.

Preexperimental Questionnaire

Six weeks before the experiment began, students from introductory sociology and education courses at the university filled out a preexperimental questionnaire. The questionnaire contained Snyder's (1974) Self-Monitoring Scale (labeled "personal reaction inventory") and a "critical issues inventory" that measured attitudes toward controversial issues. Two items on legalizing marijuana served as the basis for determining the accuracy of the impression to be conveyed. The items were scored from −3 (extremely unfavorable to legalization) to 3 (extremely favorable to legalization). Subjects receiving a summed score of 4 or more on these items were judged as definitely in favor of legalization and were eligible to participate in the experiment.

Design

The study utilized a 2×2 factorial design that varied type of impression (accurate vs. fabricated) and level of self-monitoring (high vs. low). The impression was accurate if the subject was asked to appear in favor of legalization; it was fabricated if the subject was asked to appear opposed to legalization. Subjects were high in self-monitoring if they scored above the median (12 in a possible range from 0 to 25) on the self-monitoring scale; they were low in self-monitoring if they scored below the median. Subjects were randomly assigned to the impression-type condition, with the constraint that in each condition half the subjects must be high and half must be low in self-monitoring. This yielded 16 subjects in each cell of the design. Confederates were randomly assigned to each condition.

Experimental Procedure

In a preparation room, the experimenter explained the (fictitious) nature of the study. The confederates were not present. The experimenter explained that this experiment was an attempt to discover how people form an impression about another when they are face-to-face with that person, and that the subjects would serve as a discussion partner with another woman. They were also told that the experimenter would find out if the other woman had formed specific impressions about the subject after the discussion had ended. Subjects were told they would be paid $2.50 for their participation. They then signed the participation consent form.

Next, the experimenter handed two instruction sheets to the subjects. The first described their task in general. Reading from the instructions, the experimenter explained that to ensure comparability across settings, the subjects' task was to convey a *specific* impression about themselves to their partner during a 10-minute, face-to-face conversation period. The specific impression dealt with marijuana legalization. The experimenter explained that he was interested in determining whether appearing to be for or against an issue influences how others build up an impression of people. Therefore, some would be randomly chosen to convey the impresson that they were in favor of legalizing marijuana use, whereas others would be randomly selected to appear opposed to it.

In addition, the experimenter told the subjects he was interested in whether the other person could build up an overall *favorable* picture of them at the same time she was learning whether they were for or against marijuana use. Subjects were asked to induce the other woman to like them, regardless of which side of the issue they were assigned. The purpose of this task was to keep the subject from merely stating her assigned views perfunctorily and to promote a more realistic conversation.

The experimenter explained that the other woman was the focus of the experiment. He said he would be analyzing exactly how she received the impression they would convey. The other woman (ostensibly) was with a (nonexistent) research assistant and had been told the study involved an analysis of what goes on in an informal, videotaped discussion of issues that affect college students' lives. The rationale for the 10-minute discussion was that it was a practice session to give the participants a chance to get used to the discussion.

Subjects were led to believe that the other woman had been given a list of possible discussion topics, including marijuana legalization. Since it was only one of several issues on the list, the subjects were told that they would probably have to bring up the topic themselves. They were told to do this in the manner most comfortable for them, as long as the

partner had a fair chance to build an impression of how they felt about marijuana.

The experimenter stressed that it was important that the partner not realize that the study was really about impression formation until it was actually over; otherwise, she might try too hard and do things she wouldn't do in normal, everyday life.

The second instruction sheet informed the subjects of some special considerations regarding their task. First, the experimenter was interested in more than whether their partner came to understand how they felt about legalization. Subjects were asked to build up in the other person's mind an organized, sensible set of reasons supporting their position. To help them, they would have available a list of some of the most common reasons given for the position they were assigned to convey.[1] The experimenter emphasized that their use of this list was entirely a matter for them to decide and that the quality of reasons used was at least as important as the quantity.

Second, subjects did not have to persuade the other person to agree with them. The experimenter was only interested in whether and how the partner came to understand what the subjects' position was.

Third, the subjects were told that success in forming impressions often carries some kind of benefit for those involved in the get-together. To simulate the real-life possibility of rewards, the subject and her partner would each earn a 50¢ bonus if the other woman formed *both* a clear impression of the subject's attitude toward marijuana legalization and an overall favorable impression of her. Success would be determined by the target person's responses to a questionnaire after the discussion.

Subjects were then taken to their separate interaction rooms, where the actual discussions were to take place. At this point, half the subjects learned that they were to convey the impression that they favored legalizing marijuana (an accurate impression), and the other half were to appear to oppose legalization (a fabricated impression). They were also given a list of 16 reasons supporting the assigned position. To this point, the experimenter had not known to which condition the subjects would be assigned. Throughout the study, the experimenter was blind to subjects' self-monitoring scores.

Dependent Variable

After considering their positions for a few minutes, subjects were given an opportunity to purchase information about their partner. Three kinds of information were available: biographical (e.g., hometown, religious preference), attitudinal (involving issues other than marijuana, e.g., abortion, the priority of conscience over law), and personality characteristics (e.g., assertiveness, tolerance for others). This information came from a self-descriptive questionnaire ostensibly being filled out by the other woman at that time. Actually, a standard response for each item was provided by the experimenter. These responses were neither extreme, in the case of the attitudinal and personality items, nor unusual, in the case of the background information.

For every 2 bits of information selected, 1¢ was deducted from the standard subject fee of $2.50. Ten bits of each type were available (a total of 30 bits), for a maximum possible cost of 15¢. Subjects were told that the cost was to simulate the real-life possibility that one could often find out about another only at some cost to the self. The experimenter emphasized that subjects were free not to purchase any information at all. They could have as little or as much as they wanted. Subjects were given a few minutes alone to decide, at which time the experimenter returned with the questionnaire (presumably filled out by the other person) and filled in any requested responses. The amount and kind of information bought by the subjects constituted the major dependent variable.

Performance Ratings

After being given 5 minutes to prepare, subjects engaged in an informal 10-minute discussion with the partner, who was brought to the subjects' room. The other woman was actually a confederate, chosen for her ability at improvisational drama.[2] She was trained to be easygoing, personable, and ambivalent regarding legalization. Although keeping up her end of the conversation, she did not direct nor dominate the conversation in any way. The confederates were blind to the impression-type manipulation and the premeasured attitudes of the subjects, as well as to the experimental hypotheses. Subjects were randomly assigned to a confederate.

Following the discussion, the subject and the other woman were separated, and each filled out the same questionnaire consisting of 30 bipolar adjectives in semantic differential form. Half of the items measured their impressions of the performance that the subject had just given (e.g., interesting, impressive); the rest measured their impression of what the subject had been feeling during the discussion (e.g., enthusiastic, self-conscious). These items constituted the measure of the second dependent variable.

At this juncture, the subjects were extensively debriefed. This involved assessing their suspicion level, discovering their perceptions about the experiment, revealing the true nature of the experiment, and ensuring that they felt comfortable about their participation. The experimenter then paid the subjects $2.50, cautioned them not to discuss the experiment, and dismissed them.

---

[1] The reasons were generated from two sources: Kaplan (1970) and the National Commission on Marijuana and Drug Abuse (1972).

[2] I thank Joyce Bizub, Mary Erdman, Laurie Franklin, Jane Homburg, Laura Raffe, and Dana Schwindt for serving as confederates. Their skill at improvisation added greatly to the believability of this experiment.

Results and Discussion

Internal Validity

During the debriefing, the degree of suspicion the subjects expressed was coded. Complete surprise upon learning the true nature of the experiment was assigned a value of 0; some vague feeling that something else was going on was assigned 1; the correct inference that the subjects themselves were at least part of the focus of the study received a 2; guessing in addition that the partner was a confederate received a 3; correctly discerning that the experiment investigated the subject's impression management behavior was scored 4.

Complete surprise was expressed by 60.9% of the subjects, and 26.6% had only a vague feeling that something else was going on. The mean level of suspicion was .61, indicating that most subjects believed the cover story they were given. Analysis of variance in suspicion using level of self-monitoring and impression type as factors revealed that no condition caused more suspicion than any other. In addition, dichotomizing subjects into nonsuspicious (those receiving a coded value of 0) and suspicious (those receiving any other score) created a factor that showed no main effects on the dependent variables, nor were any interactions between suspicion and the experimental factors in evidence.

Information Purchasing

The major dependent variable in this study was the amount of information purchased by subjects in preparing for their self-presentation. Recall that the available information was of several types. To see whether the effects of impression type and self-monitoring differed for each type of information, an analysis of variance was carried out, with the level of self-monitoring and impression type as between-subjects factors and information type as a within-subjects factor with three levels: biographical, attitudinal, and personality characteristics. The results of the analysis support the hypotheses. Analysis of variance in the total amount of information purchased revealed a significant main effect for self-monitoring, $F(1, 60) = 10.65$, $p < .002$, a marginal main effect for impression type, $F(1, 60) = 3.57$, $p < .07$, and a significant two-way interaction between these factors, $F(1, 60) = 4.25$, $p < .05$. The marginal values of Table 1 provide the means for each cell of the design. High self-monitoring subjects purchased more information than low self-monitoring subjects. Fabricating subjects showed a tendency to purchase more information than those conveying an accurate impression. High self-monitoring fabricators reliably bought more information than all others.[3]

An additional interesting finding was a three-way interaction between level of self-monitoring, impression type, and information type, $F(2, 120) = 3.213$, $p < .05$. Inspection of the means (provided in Table 1) revealed the pattern of this interaction. For low self-monitoring subjects, there were no differences in the amounts of information purchased. High self-monitoring subjects tended to buy the same amount of personality information, regardless of the truthfulness of the impression they were conveying; they tended to buy less biographical and attitudinal information when planning an accurate impression than when planning a fabrication.

It may be that, for the high self-monitors, what the other individual thinks she is like as a person is not differentially important in planning a performance. Personality characteristics are important elements of some baseline information about another, but they are probably not directly related to the issue at hand (at least not in this experiment). Personality information serves mainly to guide actors in choosing the general manner of self-presentation (i.e., orientation to the other)

[3] Inspection of the within-cell variances revealed that the assumption of homogeneity of variances was tenuous. The information purchased was transformed by the square root function, resulting in more homogenous cell variances. The analysis of variance for this transformed variable did not change the significance levels of any of the effects. Since the effects are more readily interpretable for the untransformed variable, results are discussed in terms of it. An analysis of variance was conducted using confederate as an additional between-subjects factor. Results showed no effects for confederate, alone or with any of the other factors.

Table 1
Mean Number of Information Bits About the Partner Purchased by Actors as a Function of Impression Type, Level of Self-Monitoring, and Information Type

Information type	Impression type: Accurate	Impression type: Fabricated
High self-monitoring		
Attitudinal		
M	$.375_a$	1.938_b
SD	.719	2.744
Biographical		
M	$.625_a$	1.875_b
SD	1.258	1.821
Personality		
M	1.125_a	1.188_{ab}
SD	1.360	1.424
Total		
M	2.125	5.000
SD	2.872	4.502
Low self-monitoring		
Attitudinal		
M	$.438_a$	$.250_a$
SD	.629	.683
Biographical		
M	$.438_a$	$.438_a$
SD	.629	.727
Personality		
M	$.375_a$	$.438_a$
SD	.619	.629
Total		
M	1.250	1.125
SD	1.612	1.668

Note. Means that do not share a subscript differ from each other at the .05 level of significance by the Newman–Keuls procedure.

and not specific elements of the performance. Therefore, personality information is equally important in conveying an accurate or a fabricated impression.

Attitudinal and biographical information is more directly related to the issue at hand. These data give indications of how one might feel about the issue around which the impression is being built. This would be especially important for a fabrication. High self-monitoring actors in a fabrication are feeling their way through unfamiliar territory and need relevant information about their partners. With an accurate impression, high self-monitors can rely on their abilities as performers. The luxury of being themselves may lend high self-monitors the confidence to do without this information and deal with contingencies as they arise. In short, a fabricated impression is more of a challenge to high self-monitoring actors than is an accurate impression.[4]

One might ask if the difference between high and low self-monitors was due to the fact that low self-monitoring actors felt no heightened concern when contemplating a fabrication. Although no data from this experiment directly addressed this question, some indirect evidence suggests that this was not the case. All subjects judged the accurate performance as easier ($M = 3.97$) than the fabricated one ($M = 1.87$), $F(1, 60) = 12.53$, $p < .001$, and there were no such differences between high and low self-monitors, $F(1, 60) < 1$. It is probable that all subjects anticipated difficulty with a fabrication. Because low self-monitors rely on their own internal states for resources, however, the challenge afforded by a fabrication did not motivate low self-monitoring actors to acquire more information about their partner.

Reactions to the Performance

The second set of hypotheses focused on the aftermath of the self-presentation. Actor subjects and observer confederates were given two sets of 15 bipolar items in semantic differential format to record their impressions following the encounter. Responses were made on a 9-point semantic differential scale. For purposes of analysis, a score of 9 was given to the positive extreme and 1 to the negative extreme.

It should be noted that confederate ob-

[4] This explanation is entirely post hoc and should be taken with some caution. One factor that adds to the tentativeness of this reasoning is that for the low self-monitors, a floor effect is operating: They cannot buy fewer than zero pieces of information. However, tests for the differences between accurate and fabricated impression type were conducted within each level of information type using only high self-monitors. In a test for "simple effects" of impression type (cf. Kirk, 1968, chap. 8), the results indicated significant differences ($p < .05$) in the amount of biographical and attitudinal information purchased by accurate and fabricating subjects; there was no difference with respect to personality information.

servers' ratings of the subjects' performance quality were their own personal judgments of what they saw; their ratings of the subjects' feelings during the interaction were what the subjects were *apparently* feeling, as revealed by their verbal and nonverbal behavior. Further, subjects and confederates completed these items while totally isolated from each other, with no way of influencing each other's ratings. Subjects were not told that they would be rating themselves until after the discussion was over, nor did they know that the confederate had rated them until the debriefing. Confederates were not told that the subjects had rated themselves until the last subject had been run. Finally, it should be recalled that the confederates were blind to all experimental conditions, so that any differences due to these factors must have been communicated through the performance.

Because of the large number of evaluations being made (30 each for subject and confederate), subjects' and confederates' ratings were factor analyzed separately using Rao's uniqueness rescaling procedure with varimax rotation. This served to organize the evaluations, reduce the number of analyses conducted, and ensure that the variables to be analyzed were relatively orthogonal.

Results for subjects' self-reports revealed three factors for the performance quality ratings and four factors for the feeling states ratings. For the performance quality items, Factor 1, an acting-quality factor, consisted of the items unconstrained–constrained, satisfied–dissatisfied, and graceful–awkward. The second factor reflected organization and included organized–disorganized, systematic–unsystematic, clear–unclear, impressive–unimpressive. Factor 3 indicated communicative quality, with consistent–inconsistent, understood–misunderstood, and persuasive–unpersuasive loading on this factor.

For the feeling states items, the first factor seemed to reflect satisfaction with the performance: positive–negative, capable–incapable, satisfied–frustrated, commendable–reprehensible, and enthusiastic–apathetic. The second factor was a security factor: calm–nervous, comfortable–uncomfortable, secure–insecure, and confident–unconfident. Factor 3 was a friendliness factor and consisted of friendly–unfriendly and likable–unlikable. Finally, Factor 4 seemed to reflect personal comportment: sensible–foolish, straightforward–devious, and free–constrained.

Turning to the confederates' evaluations of the subjects, four factors emerged from the ratings of performance quality. Factor 1 indicated performance competence: interesting–uninteresting, impressive–unimpressive, successful–unsuccessful, satisfied–dissatisfied, and believable–unbelievable. The second factor involved aesthetic reactions to the performance and consisted of organized–disorganized, graceful–awkward, flawless–flawed, and systematic–unsystematic. Factor 3 involved communicative clarity, with understood–misunderstood and clear–unclear. Factor 4 reflected performance coherence: persuasive–unpersuasive and consistent–inconsistent.

For the feeling states items, five factors were discovered. Factor 1 consisted of items dealing with the actors' perceived comfort in the role: calm–nervous, comfortable–uncomfortable, and free–constrained. Items loading on Factor 2 suggested personal accomplishment: commendable–reprehensible and capable–incapable. Factor 3 involved security, with secure–insecure, unself-conscious–self-conscious, confident–unconfident, and satisfied–frustrated. The fourth factor indicated friendliness: likable–unlikable, friendly–unfriendly, and sensible–foolish. Finally, Factor 5 suggests role commitment: enthusiastic–apathetic and positive–negative.[5]

A second-order factor analysis was run on

[5] As the reader may have noted, the factor structures for the two rating sources (subject and confederate) are somewhat different. The difference appears to reflect the perspective each brings to the performance. The factor structure for the actor subjects indicates their concern for the technical qualities of the performance. By contrast, the factor structure for the confederates reveals their critical reactions to the aesthetics of the performance. Such a performer/audience difference in perspectives is consistent with the dramaturgical approach to social interaction. However, it should be noted that the experimental design may have contributed to this difference. The observers were confederates of the experimenter and as such may have felt more like an audience than a coparticipant in the interaction.

the entire set of factors separately for subjects and confederates.[6] Variables used in the second-order analysis were obtained by summing the scores for the items loading on the same factors. For the subjects, analysis of the seven first-order factors yielded three second-order factors. Factor 1 involved performance competence, performance satisfaction, security, and personal comportment (performance quality first-order Factor 1 and feeling states first-order Factors 1, 2, and 4) and may be labeled a performance enactment factor. Factor 2 included organization and communicative quality (performance quality first-order Factors 2 and 3), which appears to deal with the mechanics of the performance. Factor 3 was the friendliness factor (feeling states first-order Factor 3).

For the confederates' ratings, analysis of the nine first-order factors also produced three second-order factors. Factor 1 involved performance competence, performance aesthetics, communicative clarity, performance coherence, and personal accomplishment (all the performance quality first-order factors plus feeling state first-order Factor 2) and indicates a critical evaluation of the performance. Factor 2 included comfort in the role, security, and role commitment (feeling states first-order Factors 1, 3, and 5), which appears to reflect a critical evaluation of the actors' personal fit to the role requirements. Finally, Factor 3 was the friendliness factor (feeling states first-order Factor 4).

Analysis of variance was conducted on the second-order factors. Variables were created by summing the scores for the items loading on the same factor. Results provide moderate support for the hypotheses. The subjects' second-order factors were relatively unaffected by the experimental conditions. No effects for level of self-monitoring achieved statistical significance. Impression type affected only the performance enactment factor (Factor 1). Subjects rated their performance enactment more positively for an accurate impression ($M = 100.31$) than for a fabrication ($M = 88.44$), $F(1, 60) = 6.64$, $p < .025$.

The experimental conditions had greater influence on the confederates' second-order factors. Level of self-monitoring affected all three factors. For the performance evaluation factor (Factor 1), confederates rated high self-monitors more favorably ($M = 108.41$) than low self-monitors ($M = 93.72$), $F(1, 60) = 6.82$, $p < .025$. High self-monitors were also seen to fit the role they enacted (Factor 3) better than low self-monitors ($M = 65.59$ for high self-monitors vs. $M = 59.53$ for low self-monitors), $F(1, 60) = 3.97$, $p < .06$. Finally, high self-monitors were seen as friendlier (Factor 3; $M = 23.62$) than low self-monitors ($M = 22.19$), $F(1, 60) = 4.05$, $p < .05$, a result also reported in Lippa (1976).

Impression type influenced the confederates' ratings on two second-order factors. Accurate impression managers were judged by the confederates to fit the role (Factor 2) better ($M = 65.69$) than fabricators ($M = 59.44$), $F(1, 60) = 4.22$, $p < .05$. In addition, they tended to see accurate impression managers as friendlier (Factor 3; $M = 23.59$) than fabricators ($M = 22.22$), $F(1, 60) = 3.71$, $p < .06$.

Although the targets seemed sensitive to many aspects of the actors' performances, one exception stands out. They did not seem to realize when the actors were deceiving them. Observers rated fabricators as no less straightforward or believable than accurate performers were. Believability *was* affected by the level of self-monitoring. High self-monitors were rated as more believable ($M = 7.47$) than low self-monitors ($M = 6.03$), $F(1, 60) = 5.95$, $p < .02$.

Apparently, actors were able to look fairly convincing during the discussion. They did not

[6] A second-order factor analysis was conducted using all 16 factors extracted for both subjects and confederates. The results yielded 4 second-order factors. Factor 1 consisted of all the confederates' first-order factors. Factors 2, 3, and 4 were identical to the 3 second-order factors reported in the analysis involving only the subjects' factors. Because there was no mixing of first-order factors across rating source in the second-order structure and because separate analyses of the subjects' and confederates' factors yielded identical results in the subjects' case and more differentiated results in the confederates' case, it was decided to use the second-order factors from the separate analyses as dependent variables in the analyses of variance.

fool themselves, however. Actors in the fabricating condition reported themselves as less straightforward ($M = 5.94$) than those conveying an accurate impression ($M = 7.34$), $F(1, 60) = 5.48$, $p < .05$, and slightly less believable ($M = 7.09$ for fabrication vs. $M = 7.81$ for accurate impression), $F(1, 60) = 3.26$, $p < .08$.

Compare these results to those in Lippa (1976), in which subjects sequentially role played teachers as introverts, extraverts, and themselves. Although rating high self-monitors as more technically proficient actors, observers were better able to detect their true tendencies toward extraversion. The correlation between perceived and actual (as assessed by premeasures) extraversion was greater for high than for low self-monitors. In other words, high self-monitors were perceived more accurately than low self-monitors. Perhaps the difference is due to differences in the content of the self-presentation. In the present study, attitudes were being conveyed; in Lippa's study, the content of the impression was a personality characteristic. Expressive control for the latter might be much more difficult.

Conclusion

The results of this study indicate that actors involved in planning a self-presentation exhibit differential interest in information about a prospective interaction partner. A person variable (self-monitoring) and a situation variable (impression type) combined interactively to influence the amount of information acquired by impression managers. Specifically, only high self-monitors who planned a fabrication tended to purchase information about their prospective partner. Further, whereas low self-monitors found all types of information unnecessary, high self-monitoring actors placed different values on the kind of information available. High self-monitors acquired personality information about their partner regardless of impression type but purchased attitudinal and biographical information only when planning a fabrication.

This study also investigated actors' and observers' reactions to the actors' performance. One striking result is the degree to which subjects' self-reports were unaffected by the experimental conditions. In particular, high self-monitors rated themselves no better or worse than low self-monitors did. But confederates clearly distinguished between them, consistently evaluating high self-monitors more positively. This suggests that although high self-monitors may have been as worried as their low-scoring counterparts, they managed to hide those worries from their partner.

Several aspects of this experiment may limit the generalizability of the results. First, it may be that the results presented above apply only to those individuals whose true attitudes favor the issue involved in the performance. It should be recalled that all subjects were selected for their strong support for legalizing marijuana, a popular stand on college campuses. Perhaps the behavior of fabricating actors was due to being seen as supporting an unpopular position as much as to giving a deceptive performance.

Second, the impression to be conveyed was assigned by the experimenter and was not chosen by the actors themselves. It may be that the effects due to impression type were actually a function of a forced fabrication. Actors assigned a truly self-descriptive impression may behave no differently than they would if they had selected the impression themselves; actors assigned a fabrication may behave differently, more because they had no choice in the matter than because the impression was not honest. There are situations in which demands for a particular impression are strong enough so that very little choice exists. If the goals desired from the interaction are important, actors may be induced to enact the impression demanded by the situation, even if it involves a fabrication. Their plight would be not unlike that of fabricating subjects in the present experiment.

In sum, the results of this study suggest that impression managers are differentially concerned with acquiring information about another participant in an upcoming interaction. High self-monitors planning a fabrication are the only actors who acquire a significant amount of information about the other. More important, they will endure some personal

costs to obtain this information. They apparently believe that the advantages of additional information in preparing a self-presentation outweigh the costs of obtaining it. In the present study, individuals acquired information without the other's knowledge. It would be interesting in future research to learn if high self-monitoring fabricators maintain their interest in purchasing information when the other knows what is being acquired and at what cost. Further insights into the planning of a self-presentation will provide greater understanding of the process guiding impression management.

References

Eiser, C., & Eiser, J. R. Acquisition of information in children's bargaining. *Journal of Personality and Social Psychology*, 1976, *34*, 796–804.

Eiser, J. R., & Tajfel, H. Acquisition of information in dyadic interaction. *Journal of Personality and Social Psychology*, 1972, *23*, 340–345.

Ekman, J., & Friesen, W. V. Nonverbal leakage and clues to deception. *Psychiatry*, 1969, *32*, 88–105.

Goffman, E. *The presentation of self in everyday life*. Garden City, N.J.: Doubleday, 1959.

Ickes, W., & Barnes, R. D. The role of sex and self-monitoring in unstructured dyadic interactions. *Journal of Personality and Social Psychology*, 197 *35*, 315–330.

Kaplan, J. *Marijuana: The new prohibition*. Ne York: Meridian, 1970.

Kirk, R. E. *Experimental design: Procedures for th behavioral sciences*. Belmont, Calif.: Brooks/Col 1968.

Lippa, R. Expressive control and the leakage of dis positional introversion-extraversion during role played teaching. *Journal of Personality*, 1976, 4 541–559.

National Commission on Marijuana and Drug Abus *Marijuana: A signal of misunderstanding* (Appen dix, Vol. 2). Washington, D.C.: U.S. Governmen Printing Office, 1972.

Snyder, M. Self-monitoring of expressive behavio *Journal of Personality and Social Psychology*, 1974 *30*, 526–537.

Snyder, M. Cognitive, behavioral, and interpersona consequences of self-monitoring. In P. Pliner, K. R Blankstein, I. M. Spigel, T. Alloway, & L. Krames (Eds.), *Advances in the study of communicatio and affect: Vol. 5. Perception of emotion in sel and others*. New York: Plenum Press, 1978.

Snyder, M. Self-monitoring processes. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 12). New York: Academic Press, in press.

Weiler, J., & Weinstein, E. A. Honesty, fabrication, and the enhancement of credibility. *Sociometry*, 1972, *35*, 316–331.

Received July 10, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 8, 1293–1303

Effects of Speech Style and Sex of Speaker on Person Perception

Nora Newcombe and Diane B. Arnkoff
Pennsylvania State University

Lakoff has suggested that men and women use different speech styles, with women's speech more polite but less assertive than men's. The assumption that three of Lakoff's linguistic variables (tag questions, qualifiers, and compound requests) affect person perception in these ways was tested. Sex of speaker was also varied. In Experiment 1, undergraduates rated the assertiveness, warmth, and politeness of two male and two female speakers who used or did not use the three linguistic forms. All three "female" linguistic forms were rated less assertive than corresponding "male" forms; qualified speech and compound requests were rated warmer and compound requests more polite. Sex of speaker was a significant factor in only one possible comparison. These results were substantially replicated in Experiment 2, in which older and/or less educated women acted as judges. These findings suggest that Lakoff's intuitions concerning effects of speech styles on person perception are largely correct and that modification of speech styles could allow men and women to affect how they are perceived.

Our impressions of other people are derived from many sources: what they say, what they do, what others think of them, and so on. What people say is surely an important source of information, but *how* they say it may also be of great importance. Thus, for example, a rapid speech rate leads to higher ratings of speaker competence (Smith, Brown, Strong, & Rencher, 1975) and increases the persuasiveness of a communication by enhancing speaker credibility (Miller, Maruyama, Beaber, & Valone, 1976). The present article focuses on three linguistic dimensions that are hypothesized by Lakoff (1975) to affect person perception. According to Lakoff, the three forms are used differentially by men and women, thus contributing to the maintenance of stereotypes concerning sex differences in personality.

Lakoff has suggested that men and women differ in their styles of speech in ways that both result from sex stereotypes and reinforce those same stereotypes. Specifically, sex differences in speech styles are said to contribute to maintaining images of men as assertive, self-confident, and definite, and images of women as vague and lacking in confidence. Women's speech is also said to be more polite, in the sense of being more formal and more deferential than men's. The differences in speech style Lakoff discusses include the following linguistic dimensions, among others:

1. Greater female use of tag questions at the end of declarative sentences, for example, "It's really cold in here, isn't it?" Tags request listener confirmation of the truth of an assertion and are thus said to indicate lack of

We thank Cindy Allen, Kathy Malloy, and Carolyn Brodbeck for administering procedures; Cynthia Marsh, Ann Greeley, and Lillian Shermaski for assistance in coding the informally reported speech production study; Rus Bauer, Don Hallett, Debbie Northwicke, and WMAJ Radio for help in making tapes; and Alvin Rosenthal and James Rosenberger for statistical advice. David Siegler discussed the studies with us, and Carolyn Sherif and Martha Zaslow provided critical comments on the manuscript. Portions of this research were presented in a paper titled "Speech Styles and Sex Stereotypes" given before the Association of Women in Psychology, Pittsburgh, March 1978.

Requests for reprints should be sent to Nora Newcombe, 522 Moore Building, Pennsylvania State University, University Park, Pennsylvania 16802.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3708-1293$00.75

confidence. They are also characterized as polite in that they allow the listener a graceful "out" in case of disagreement.

2. Greater female use of hedges or qualifiers, that is, words or phrases such as "y'know," "kinda," "I guess," or "maybe," which blunt the force or definitiveness of an assertion but, again, are polite in that they give the listeners options if they wish to disagree with or ignore the statement.

3. Greater female use of compound requests (e.g., "Won't you close the door?") rather than simple requests (e.g., "Close the door"). Such forms are said to soften the assertion of power entailed in commanding or requesting action, as well as being considered more polite.

While these are interesting hypotheses with a variety of social implications, Lakoff has presented no empirical evidence to support her claims. She makes at least three assumptions based on intuition. First is the assumption that there are, in fact, differences in the frequency of use of the above-mentioned words and syntactic constructions as a function of sex. Second, Lakoff assumes that these linguistic differences influence how people are perceived, in the ways she has suggested. Third, Lakoff seems to assume, although she is inconsistent on this point, not only that linguistic style affects person perception but that style is an *important* influence on person perception.

The experiments to be reported in this article concern the second and third assumptions. However, we will first review evidence concerning the first assumption, that there are sex differences in linguistic usage. While this is not critical for our purposes, since the effects of individual linguistic variations are of some interest in their own right, much of the current excitement concerning Lakoff's proposals depends on the assumption of sex patterning. The data to date are somewhat contradictory. Several studies have failed to find sex differences, but there is growing evidence that such differences do exist, although perhaps only in certain contexts or for certain populations.

While Dubois and Crouch (1975) found 33 tag questions from males, but none from females, on tapes of question periods following formal presentations at a professional meeting, they did not provide information on number of participants or total time talking for each sex. Perhaps very few females were present, or those present talked very little in any style. Furthermore, an academic population may have distinctive speech styles. Lakoff (1975, 1977) has discussed at some length her belief that academic men are exceptions to her rules and use a speech style generally identified as "female." Many of the same reservations can be expressed about another study reporting no sex differences in the use of tag questions (Baumann, 1976).

Another study finding no sex differences was conducted by the present authors (Newcombe & Arnkoff, Note 1). Pairs of unacquainted undergraduates were taped while talking about three topics of general interest. The tapes were coded for instances of tag questions and qualifiers, as well as for other language differences discussed by Lakoff: rising intonation on declarative sentences, *so* and *such* as intensifiers, "cute" adjectives of admiration, and expletives and euphemisms. No differences in speech style or trends to differences were observed due to sex of speaker, sex of listener, or the interaction, either for frequencies or frequencies divided by time speaking.

There might, of course, be a variety of reasons for the lack of sex differences in linguistic form usage in this study, with perhaps the most obvious being the use of a college student population. However, sex differences in speech style were found in a very similar study by Crosby and Nyquist (1977, Study 1) that coded use of empty adjectives, tag questions, hedges or qualifiers, and *so*. McMillan, Clifton, McGrath, and Gale (1977) also reported differences between the speech of male and female college students, with women using more tags, more imperatives in question form, more intensifiers such as *so* and *such*, and more modals such as *might have said* (a dimension discussed by Key, 1975). McMillan et al. placed their subjects in a group problem-solving situation. Either being in a group or being in a problem-solving situation could potentially account for the appearance of sex-related linguistic differences. McMillan et al.

also found that sex differences were especially marked in comparing speakers from mixed-sex rather than same-sex groups.

Hartman (1976) found that in individual interviews, women aged 70 years or over used more tag questions and qualifiers compared to men of similar ages. Crosby and Nyquist (1977, Study 3) found that female clients requesting aid or information in a police station used "female register" (i.e., tag questions, hedges or qualifiers, and politeness expressions) more than male clients, and that female police personnel also used this style more than male police personnel. However, Crosby and Nyquist (Study 2) did not find sex differences in requesting assistance at an information booth. Gleason and Weintraub (1978) reported that mothers produced twice as many tag questions as fathers in speech to preschoolers, that fathers produced more simple requests than mothers, and that male day-care teachers produced more simple requests than female teachers. Further work is clearly needed to identify whether, and in what ways, variations in context, age, and other considerations determine the appearance of sex-related differences in language; that they do exist in at least some instances seems fairly well established.

There is relatively little direct evidence bearing on the second of Lakoff's assumptions, that differences in usage contribute to how people are perceived. While Bates (1976) has provided evidence that interrogative requests in Italian are judged to be more polite than simple imperatives, and Siegler and Siegler (1976) have found a nonsignificant trend for tag questions to be judged less intelligent than assertions, both studies involved judgments of a small sample of written sentences, with attention implicitly or explicitly drawn to the dimension of interest. We do not know from these studies whether linguistic form affects person perception in a more realistic speech context when attention is not focused on the forms and dimensions of interest.

Erickson, Lind, Johnson, and O'Barr (1978) have compared reactions to simulated court testimony delivered either in "powerful" or "powerless" style. These styles contrasted three of the dimensions discussed by Lakoff, hedges or qualifiers, intensifiers, and question intonation in normally declarative contexts. However, in addition, powerless speech contained many hesitation forms, such as pauses, stutters, and "uhs," incorrect or informal pronunciation, using of *sir* in addressing the attorney, and instances in which the witness asked the lawyer questions. Speakers using powerless speech were rated less credible and less attractive than those using powerful speech. While these findings indicate the general importance of speech style in impression formation, unfortunately they do not allow specific identification of which of the many contrasting speech dimensions were responsible for the effects observed.

It should be noted that some of Lakoff's claims regarding the effects of speech style in person perception are not obvious. For instance, tag questions, rather than conveying uncertainty, can be used in a condescending or overbearing manner to forestall opposition (Dubois & Crouch, 1975). If it is said with appropriate intonation, the statement "That production of Hamlet was *dreadful*, wasn't it?" can in fact be intimidating rather than uncertain.

With regard to Lakoff's third assumption, it seems important to determine whether speech style is a reasonably large determinant of person perception compared to the effect of sex stereotypes per se. According to Valian (1977), Lakoff is inconsistent in her position on the relationship of linguistic change to social change. She sometimes argues, as in her discussion of the term *Ms.*, that linguistic change can only reflect social change, not produce it. However, at other points she implies that changes in speech style could lead to changes in person perception large enough to be socially important. Whether the latter is the case would seem to depend in part on whether sex stereotypes are so powerful that most men would be seen as more assertive than most women, with speech style a relatively minor source of individual variation within sex.

The present studies were designed to examine the second and third assumptions regarding the effect of linguistic form on person perception. Ratings were obtained of the as-

sertiveness, politeness, and warmth of male and female speakers who used (or did not use) tag questions, qualifiers, and compound requests. The existence and nature of the hypothesized influence of the linguistic variables on person perception could thus be assessed, and the size of any effect could be evaluated relative to the effect of sex of speaker. Warmth was included in part because Lakoff alludes to women's speech as sharing greater sympathy, friendliness, and person orientation than men's and also because McMillan et al. (1977) suggest that Lakoff has overemphasized the negative connotations to women's speech and underemphasized its value in conveying interpersonal sensitivity. It would have been interesting to obtain ratings of femininity also, since Lakoff maintains that women who deviate from female style pay the price of being perceived as unfeminine. However, inclusion of such a scale would probably have sensitized subjects to the variable of sex of speaker and the general purpose of the study, thereby creating demand characteristics that would have invalidated results.

Following the rating task, subjects were also asked to estimate the relative frequency with which male and female speakers used tags, qualifiers, and compound requests. Siegler and Siegler (1976) have found that when asked to guess about sex, raters attribute tag questions to women and strong assertions to men. One purpose of the present studies was to determine whether sex stereotypes affect people's abilities to judge frequency differences. As Thorne (1976) suggests, if sex stereotypes override what people actually hear, Lakoff may herself have been biased in assuming that there are linguistic differences between male and female speech.

Having subjects rate spoken material leaves open the possibility that any effects found for sex of speaker will be due to differences in intonation, phrasing, or other vocal factors, rather than to subjects' sex stereotypes. Two speakers of each sex were used in an attempt to reduce the likelihood that effects of sex of speaker would be attributed to one individual's vocal characteristics, but clearly both males might differ from both females in some crucial respects. Written presentation would avoid the problem but would be a far less naturalistic technique. In any case, in these initial experiments the concern was to evaluate the effects, if any, of a variable over which people have some control (speech style) as compared to the effects of a set of variables over many of which they have no control (their sex) or relatively little control (e.g., pitch of voice can be controlled, but only within certain limits).

Experiment 1

Method

Subjects

A total of 138 undergraduates, 75 males and 63 females, participated in order to gain extra credit for a course in introductory psychology. Data from 58 subjects were later discarded to create an orthogonal design (see *Results* section for justification). Subjects were randomly assigned to one of 13 mixed-sex groups. Each group listened to one of eight versions of the stimulus material, with more than 8 groups necessary in order to obtain a minimum of 5 subjects of each sex listening to each version. The groups ranged in size from 6 to 15 people.

Stimulus Material

There were three linguistic variables of interest: tag questions, qualifiers, and compound requests. For each category, 16 *speech segments* (speech excerpts that could be part of an ongoing conversation) were written. Each was one to three sentences long and could plausibly either include or omit the words or phrases of interest. Items were chosen from a pool written by four people, after first discarding any items judged by any one of the four to involve a sex-typed topic. All tag-question items involved an inanimate third-person subject, such as a car, a movie, or a building, about which some aesthetic or practical judgment was being made. A sample item is: "My typewriter's finally been fixed. It sure took long enough (didn't it?)." Qualifier items used one of four words or phrases: *maybe, sort of, probably,* or *I guess*. A sample item is: "That's (sort of) a crazy reason for quitting school." Compound requests involved the insertion of either *will you* or *would you* before a simple imperative. Neither form included the word *please*. A sample item is: "(Will you) order me a pizza; I'll be there soon." [1]

[1] A complete list of the 48 sentences used in the experiment is available from the first author.

Two male and two female speakers recorded eight versions of the 48-item script (16 items for each of the three linguistic variables). Speaker identity was treated as a separate factor in analysis, nested within sex of subject. The speakers were in their twenties and were chosen on the basis of having no accent to the Northeastern ear, typical voice pitch for their sex, and some acting ability, although none were professional actors.

Eight versions were recorded so that each segment would be available in each of the two possible linguistic versions and would be recorded by each of the four speakers. The order of topics spoken about was constant across versions. Each of the linguistic forms (tag questions or corresponding assertion, qualified assertion or corresponding unqualified assertion, and compound request or correspondingly simple request) appeared once in each block of 6 segments, so that linguistic forms would be evenly distributed over the tape and the salience of each would be reduced. Speakers were assigned to segments so that no more than 2 successive segments were recorded by the same speaker and so that each of the four speakers recorded each of the six linguistic types once in the first half and once in the second half of the 48-segment tape.

Each subject thus heard, in a random order, 16 simple assertions (which could be considered buffer items), 8 assertions with tag questions, 8 assertions with qualifiers (one of four words), 8 simple imperatives, and 8 imperatives expressed as questions (one of two ways). This linguistic variety, combined with the variety introduced by the use of four speakers, reduces the likelihood that subjects noticed the linguistic variations and recognized that the study concerned reactions to these variations.

Procedure

Subjects were recruited for a study of "effective telephone communication." On arrival, they were told that the experimenters were interested in how people can communicate effectively when speaker and listener cannot make normal use of nonverbal cues such as facial expressions and gestures. Subjects were told that they would listen to a tape of 48 short segments of conversation recorded by people simulating telephone communication and were informed: "After listening to each segment, you will be asked to give your impression of the person talking. Specifically, you will be asked to rate the person's assertiveness, politeness, and warmth on a scale running from 1 to 10." Assertiveness was explicitly defined as "different from aggression—an assertive person doesn't have to sound belligerent or mad. But they do sound definite about what they want, rather than unsure or tentative." Subjects then signed an informed consent form, noted their sex and group number on an answer sheet, and proceeded to listen to the tape and make ratings. The experimenter stopped the tape recorder after each segment and allowed all subjects as much time as they desired to make the ratings.

Following this, another answer sheet was circulated, on which *tag question, qualifier,* and *compound request* were defined. The subject was asked to indicate for each form whether it had been used more by males, more by females, or equally by both. Subjects were then told the true purpose of the study and were given extra-credit cards.

Design

The ratings part of the study consisted of three parallel "studies," of tag questions, qualifiers, and compound requests. Each consisted of a Sex of Subject (2) × Sex of Speaker (2) × Linguistic Type (2) × Identity of Speaker (2) × Sentence Topic (i.e., stimulus sentence) (16) multivariate analyses of variance design with three dependent variables. All independent variables except sex of subject were manipulated within subjects, and identity of speaker was nested within sex of speaker. The three dependent variables were the ratings of assertiveness, warmth, and politeness on a 10-point scale, with higher ratings indicating higher levels of each quality.

Results

Rating Data

A total of 138 subjects participated in the study, but there were not equal numbers of each sex listening to each of the eight versions. The number of subjects of each sex per version ranged from 5 to 14, with a median of 8. With such a highly unbalanced design, it is likely that the order in which the variables were considered would seriously affect the results. Therefore an orthogonal design was created by discarding some subjects. Fifty-eight of the 138 subjects were randomly eliminated to leave 5 subjects of each sex for each of the eight versions. The following analyses are based on data from 80 subjects, 40 males and 40 females. No effects of version of the tape were observed in preliminary analyses, and this factor was thus dropped in the analyses reported below.

The three rating scales were correlated with each other (for assertiveness and politeness, Pearson $r = .27$; for assertiveness and warmth, $r = .18$; for politeness and warmth, $r = .81$), and thus multivariate analyses were performed on these data. The analyses were performed using the analysis of variance procedure of the Statistical Analysis System (SAS) statistical package (Barr, Goodnight, Sall, & Helwig, 1976).

Tag questions. The multivariate analysis showed a significant effect of linguistic form, Hotelling $T^2 = 31.77$, $F(3, 76) = 10.32$, $p < .0001$, and of sentence topic, Hotelling-Lawley trace = 2.64, F approximation $(45, 224) = 4.38$, $p < .0001$. Univariate analyses showed that tag questions were seen as less assertive ($M = 6.1$) than corresponding nontags (M rating = 7.2), $F(1, 78) = 20.27$, $p < .0001$. Sentence topic had a significant effect in all three analyses, of assertiveness, warmth, and politeness ratings. However, it is important to note that this heterogeneity in stimulus materials did not interact with or qualify any of the variables of interest (all Fs < 1). The only other significant univariate effect was that female subjects rated speakers as warmer than did male subjects, $F(1, 78) = 4.49$, $p < .05$, with ratings of 5.3 from females and 4.9 from males. However, sex of subject was not significant in the multivariate analysis or in the following two analyses of qualifiers or compound requests.

Qualifiers. The multivariate analysis showed a significant effect of linguistic form, $T^2 = 78.39$, $F(3, 76) = 25.46$, $p < .0001$, and a marginally significant effect of sex of speaker, $T^2 = 8.16$, $F(3, 76) = 2.65$, $p = .0541$. There was also a significant effect of sentence topic, Hotelling-Lawley trace = 1.99, F approximation $(45, 224) = 3.30$, $p < .0001$. Univariate analyses showed effects of sentence topic on all three ratings, but again no interactions (all Fs < 1). The overall effect of linguistic form was due to effects in ratings of both assertiveness and warmth. Qualified speech was seen as less assertive, with a mean rating of 5.8 as compared to 7.5 for nonqualified speech, $F(1, 78) = 60.95$, $p < .0001$, but also as warmer (mean rating of 5.5 as compared to 5.0 for nonqualified speech, $F(1, 78) = 4.71$, $p < .05$. Sex of speaker was significant in only one univariate analysis, that for assertiveness ratings, with male speakers using either qualified or unqualified speech seen as more assertive (M rating = 6.9) than females (M rating = 6.3), $F(1, 78) = 6.90$, $p \simeq .01$. That this was not due to idiosyncratic characteristics of the voices or acting of one of the speakers is suggested by the absence of any significant main effects or interactions involving identity of speaker. However, the strength of association, omega-square (Hays, 1963) of linguistic form with assertiveness ratings was .12, while the strength of association for sex of speaker with assertiveness ratings was only .01. (The strength of association for the other significant effect, sentence topic, was .09.)

Compound requests. The multivariate analysis showed significant effects of linguistic form, $T^2 = 154.78$, $F(3, 76) = 50.27$, $p < .0001$; sex of speaker, $T^2 = 11.21$, $F(3, 76) = 3.64$, $p < .05$; and identity of speaker, Hotelling-Lawley trace = .47, F approximation $(6, 150) = 5.88$, $p < .0001$. There was no significant effect of sentence topic; apparently the requests were perceived as more homogeneous than the tag question or qualifier items. Univariate analyses showed effects of linguistic form on ratings of assertiveness, $F(1, 78) = 40.14$, $p < .0001$; politeness, $F(1, 78) = 50.50$, $p < .0001$; and warmth, $F(1, 78) = 70.31$, $p < .0001$). Compound requests were rated as less assertive (M rating = 6.5 vs. 8.2 for simple requests) but more polite (6.5 vs. 4.6) and warmer (6.0 vs. 4.0). Female speakers were considered more polite than males, $F(1, 78) = 5.13$, $p < .05$, with ratings of 5.9 and 5.2, respectively. Females were also considered warmer, $F(1, 78) = 7.84$, $p < .01$, with ratings of 5.3 and 4.7, respectively. These results, however, unlike the ones for qualifiers, did seem to be due to specific characteristics of one of the two male speakers. Identity of speaker was a significant factor for ratings of both politeness and warmth, $F(2, 78) = 7.27$, $p \simeq .001$, and $F(2, 78) = 15.00$, $p \simeq .0001$, respectively. One male speaker was seen as much less polite and less warm than the other three speakers. This male received a politeness rating of 4.5, as compared to 6.0 for the second male and 5.8 and 6.0 for the two females. He received a warmth rating of 3.7, as compared to 5.6 for the second male and 5.3 and 5.4 for the two females. There was also, for the univariate analysis of the warmth ratings, a significant interaction of identity of speaker with linguistic form, $F(2, 78) = 3.34$, $p < .05$. The differences between the "cold" male speaker and the other three speakers were especially pronounced when speakers

Table 1
Judgments of Relative Frequencies of Usage of Three Linguistic Forms by Male and Female Speakers in Experiment 1

Judgment	Tag questions			Qualifiers			Compound requests		
	Males	Females	Total	Males	Females	Total	Males	Females	Total
Males use more	11	15	26	13	11	24	23	19	42
Males and females use equally	23	21	44	37	29	66	23	28	61
Females use more	42	26	68	26	22	48	20	15	35
Total	76	62	138	76	62	138	65	62	138

were using compound requests. However, the difference between the two linguistic forms appeared for all four speakers.

Frequency Estimation Data

The frequencies of usage of tags, qualifiers, and compound requests were actually equal for male and female speakers on the tape. Judgments of frequencies by all 138 subjects are shown in Table 1. (The 58 subjects discarded in the previous analyses were included here, since there was no need for an orthogonal design in this analysis.) The significance of these data was evaluated relative to a null hypothesis that subjects do not have sex stereotypes and thus judge the frequencies equal except as affected by random error due to factors such as lack of attention. Thus the null hypothesis stated that the probability of choosing *males more* would be equal to the probability of choosing *females more*. A binomial test rejected the null hypothesis of equal probability for tag questions ($z = 4.23$, $p < .001$) and for qualifiers ($z = 2.71$, $p < .01$) but not for compound requests.

The biases in frequency estimation can be interpreted either as reflecting sex stereotypes or as being accurate reports of how subjects have heard men and women talk in general, although not of how speakers talked on the experimental tape. The issue will be discussed more fully below.

Inspection of Table 1 suggests that male subjects may be more prone to bias than female subjects in making frequency judgments of tag questions, although not more prone to bias in the case of qualifiers or compound requests. However, a chi-square test carried out on the 2×3 table of responses for tag questions was not significant, $\chi^2(2) = 3.08$.

Discussion

Experiment 1 provides substantial support for several of Lakoff's intuitions regarding the effect of speech styles on person perception. Tag questions, qualifiers, and compound requests all decrease the perceived assertiveness of speech. Tags and qualifiers do not seem to increase perceived politeness, but use of qualifiers leads to higher ratings of warmth. Compound requests correspond most closely to Lakoff's hypotheses, in that their use increases ratings of both politeness and warmth as well as decreasing ratings of assertiveness.

Only one instance of an effect of sex of speaker on ratings was found in these data, that occurring for ratings of assertiveness of qualified and unqualified assertions. This effect was small relative to the corresponding effect of linguistic form: Presence or absence of a qualifier made more differences in ratings of assertiveness than did the sex of the speaker. An implication is that, when men and women do differ in frequency of usage of these three linguistic forms, the contribution to the support of sex stereotypes could be substantial. Conversely, men and women could change the way in which they are perceived by changing their speech style. However, several reservations concerning these conclusions spring to mind, perhaps most notably the fact that Experiment 1 used college students as subjects.

Perhaps an older or less well educated sample would show stronger sex stereotypes. For this reason, Experiment 2 was conducted as a replication of Experiment 1, using a different subject population.

Experiment 2

Method

Subjects

A total of 39 female secretaries employed at the university participated in return for a $2 reimbursement. Subjects were recruited through friendship networks. That is, we contacted secretaries known to us or introduced to us by a faculty member known to us and asked them to participate in the study and to bring their co-workers.

Because we wished to obtain a sample as different as possible from the undergraduates of Experiment 1, subjects were asked about their age and education, and their data were eliminated if they were under 30 years of age and had either a BA or the credit equivalent (5 people) or under 30 and had a smaller number of credits in a definite major (2 people with 52 and 100 credits). Thus data from 32 women, 4 listening to each of eight tapes, were retained. These women ranged in age from 18 to 58 years, with 7 under 20 years old, 12 aged 20–29 years, 8 aged 30–39 years, and 5 aged 40 years or over. The number of college credits they had acquired ranged from 0 to 90, with 21 having 0 and the remaining 11 having a mean of 24 credits (less than a full-year equivalent) and none having a definite major.

Stimulus Material and Design

These were as in Experiment 1, except that the design did not involve the factor of sex of subject.

Procedure

This was substantially the same as in Experiment 1, with some minor exceptions. Subjects were asked to supply information on age and education, used a computer-readable form instead of a manually scored answer sheet for making ratings, and were given money instead of extra-credit cards. They were also specifically instructed that the study had been previously conducted using undergraduates and that therefore some of the content and vocabulary of the conversation was most relevant to a college audience, for example, comments on courses, summer jobs, and living in dorms, and terms such as *frat* for *fraternity*. They were asked as far as possible to avoid considering their reactions to these "collegiate" topics or phrases in making their ratings but to concentrate on "how the speaker is coming across in general."

Results

Rating Data

As in Experiment 1, no effects of version of the tape were observed in preliminary analyses.

Tag questions. Although the multivariate analysis did not show a significant effect of linguistic form, there was a trend in the univariate analysis for tag questions to be rated less assertive than nontags, $F(1, 31) = 3.32$, $p < .10$. The mean assertiveness ratings were 6.6 for tags and 7.4 for nontags. There was again a significant effect of sentence topic, Hotelling-Lawley trace = 2.93, F approximation $(45, 83) = 1.80$, $p < .05$. Sentence topic was a significant factor in univariate analyses of warmth and politeness, but no interactions with other variables were significant. Sex of speaker did not approach significance in any analyses.

Qualifiers. The multivariate analysis showed a significant effect of linguistic form, $T^2 = 11.99$, $F(3, 29) = 3.74$, $p < .05$. Qualified speech was seen as less assertive ($M = 6.4$) than nonqualified speech ($M = 7.6$), $F(1, 31) = 8.23$, $p < .01$, although not as warmer, as it had been in Experiment 1. Neither sentence topic nor sex of speaker was a significant effect.

Compound requests. The only significant effect in the multivariate analysis was that of linguistic form, $T^2 = 35.92$, $F(3, 29) = 11.20$, $p < .0001$. Compound requests were rated less assertive than simple requests (7.2 vs. 8.5) but also warmer (6.6 vs. 4.6) and more polite (6.7 vs. 5.2). In the univariate analyses corresponding to these comparisons, $F(1, 31) = 14.93$, $p < .001$; $F(1, 31) = 16.28$, $p < .001$; and $F(1, 31) = 8.14$, $p < .01$. The univariate analysis of the warmth ratings showed a trend to an effect of sex of speaker, $F(1, 31) = 3.48$, $p < .10$, but as in Experiment 1, this was qualified by a trend to an identity of speaker effect, $F(2, 31) = 2.93$, $p < .10$. Inspection of means again showed that one male speaker was rated less warm than the other male, who was comparable to the two females.

Comparison with Experiment 1. A quanti-

tative assessment of the apparent similarity between Experiments 1 and 2 was undertaken by analyzing both studies together, including the experiment number as an independent variable with two levels and including all interactions between the experiment variable and the other independent variables. Male subjects from Experiment 1 were not included in the analysis, and 1 female subject was randomly dropped from each version of Experiment 1, so that both experiments would be equally weighted in the analysis (32 female subjects in each experiment).

For tag questions and qualifiers, multivariate analyses of variance showed no significant main effects of experiments. The analysis for compound requests did produce a marginally significant main effect, $T^2 = 8.27$, $F(3, 61) = 2.67$, $p = .0545$. Univariate analyses indicated that the secretaries in Experiment 2 gave speakers higher warmth ratings than did the female students in Experiment 1, $F(1, 63) = 4.28$, $p < .05$, and tended to give higher assertiveness ratings, $F(1, 63) = 3.39$, $p < .10$. However, no significant interactions involving the experiment variable were found in any of the three multivariate analyses. Thus the findings do not depend on which subject population is being considered.

Frequency Estimation Data

Judgments of the frequencies with which male and female speakers used tag questions, qualifiers, and compound requests are shown in Table 2. There is, again, no bias in estimations for compound requests. There are tendencies to biases in estimations for tag questions and qualifiers, although the tendency is not very marked for tag questions, and binomial tests are not significant in either case.

Discussion

Experiment 2 substantially replicated the effects of linguistic form obtained in Experiment 1, despite the reduction in power due to the reduction in sample size. No hint was obtained that sex of speaker was a significant factor for this subject population. Thus, the idea that an older or less well educated sample would have stronger sex stereotypes than the undergraduates and would thus rate male and female speakers very differently was not supported.

Table 2
Judgments of Relative Frequencies of Usage of Three Linguistic Forms by Male and Female Speakers in Experiment 2

Judgment	Tag questions	Qualifiers	Compound requests
Males use more	8	5	8
Males and females use equally	13	17	16
Females use more	11	10	8
Total	32	32	32

General Discussion

The results of these two studies indicate that as Lakoff has suggested, variations in the use of tag questions, qualifiers, and compound request forms can affect how people are perceived. The sex of the speaker had small and largely nonsignificant effects on ratings. However, before concluding that sex differences in speech style could contribute to the support of sex stereotypes and that changes in speech style might allow men and women to modify how they are perceived, we should consider the fact that the frequency estimation data from both studies can be interpreted as demonstrating the existence of sex stereotypes. Subjects are probably aware that they have been rating tag questions and qualified speech as less assertive than declaratives. They may therefore guess that the female speakers use these forms more frequently because they have a stereotype of women as less assertive than men.

This way of accounting for the frequency estimation data implies a contrast between behavior in specific situations with specific individuals (i.e., rating the speech of a particular speaker on a particular topic) and behavior in more abstract situations involving thought about a class of individuals (i.e., the reasoning outlined above for the frequency

estimation task). Judgments about concrete instances of behavior may be less influenced by sex stereotypes than judgments regarding a general class of behaviors. A woman speaking in a direct and self-confident style might thus be seen as self-confident, but the same individuals who saw her this way might still maintain that women in general are not very confident and do not speak assertively. However, the cumulative impact of judgments of the self-confidence of individual women might be to change general stereotypes.

There is, however, an alternate interpretation of the distortion in frequency estimation. In guessing about frequencies of use on the tape, subjects may unconsciously average speech they have heard in real-world contexts, rather than considering only the experimental tape. Of course, if this second idea is correct and frequency estimations reflect unconscious averaging of speech heard by subjects outside the experiment, the interpretation of the data from the experiment as a whole is more straightforward. Men and women do speak differently, at least in certain contexts, in ways that affect how they are perceived. Changing one's speech style could potentially change these perceptions.

An important question in discussing sex-related variation in speech styles has been whether the important dimension governing choice of speech style may be status rather than sex. Thus, it has sometimes been suggested that "female speech style" is really the language of the powerless or the lower in status rather than the language of women, and that women high in status or power do not use female speech style, although low-status men do use this style (Dubois & Crouch, 1975; Erickson et al., 1978; Key, 1972; O'Barr & Atkins, in press). Other sociolinguistic work, for instance that of Ervin-Tripp (1976) on imperatives, supports the idea that status differences result in linguistic variation of a kind similar to that claimed by Lakoff to be a function of sex.

Lakoff (1977) disputes the "language of the powerless" idea and claims that women with power still speak in female style, that some men with power (i.e., academics and artistic–literary men) also speak this way, and that minorities (e.g., blacks) who lack power do not speak this way. Crosby and Nyquist's (1977) finding that role and sex both condition the use of "female register" but do not interact partially supports Lakoff's argument. Status (or role) does make a difference, but within roles, females differ from males.

While empirical work is necessary to evaluate these claims and resolve the issue, it may be helpful to clarify the question being asked by contrasting two views of the relationship of sex and status to linguistic variation. A straightforward "status" hypothesis claims that, while sex and status are confounded in the real world, all speakers control all forms and can switch easily and flexibly as a function of a particular status situation. Conversely, listeners expect such context-based variation and are comfortable with it; the sex of the person holding a particular status position is irrelevant. A contrasting view holds that the real-world correlation of sex and status results in relatively stable speech patterns and listener expectations not so easily changed as a function of situation. These might result in part from early socialization, since mothers and fathers and male and female day-care teachers model sex-differentiated language in speech to preschoolers (Gleason & Weintraub, 1978) and since work on imperatives suggests that situation-based variation in commanding and requesting is well established at least by 4½ or 5 years of age (James, Note 2) and perhaps as early as 2½ or 3 years (Bates, 1976; Ervin-Tripp, 1976, 1977; Newcombe & Zaslow, Note 3). How early one's own sex is used as a basis for choosing linguistic forms is not known, however. The ability to use linguistic cues to *infer* sex of speaker undergoes considerable change during the elementary school years (Edelsky, 1977). If speech patterns and listener expectations are relatively stable, high-status women might well be in the double bind described by Lakoff (1975): branded unfeminine and unlikeable if they adopt male style and vague and frivolous if they adopt female style.

Reference Notes

1. Newcombe, N. & Arnkoff, D. B. *Speech styles and sex stereotypes.* Paper presented at the meeting of

the Association of Women in Psychology, Pittsburgh, March 1978.

2. James, S. L. *The effect of listener and situation on the politeness of preschool children's directive speech.* Unpublished doctoral dissertation, University of Wisconsin, 1975.

3. Newcombe, N., & Zaslow, M. *Do 2½-year-olds hint? A study of directive forms in the speech of 2½-year-old children to adults.* Unpublished manuscript, Pennsylvania State University, 1978.

References

Barr, A. J., Goodnight, J. H., Sall, J. P., & Helwig, J. T. *A user's guide to SAS 76.* Raleigh, N.C.: SAS Institute, 1976.

Bates, E. *Language and context: The acquisition of pragmatics.* New York: Academic Press, 1976.

Baumann, M. Two features of "women's speech?" In B. L. Dubois & I. Crouch (Eds.), *The sociology of the languages of American women.* San Antonio, Tex.: Trinity University, 1976.

Crosby, F., & Nyquist, L. The female register: An empirical study of Lakoff's hypotheses. *Language in Society,* 1977, *6,* 313–322.

Dubois, B. L., & Crouch, I. The question of tag questions in women's speech: They don't really use more of them, do they? ↓ *Language in Society,* 1975, *4,* 289–294.

Edelsky, C. Acquisition of an aspect of communicative competence: Learning what it means to talk like a lady. In S. Ervin-Tripp & C. Mitchell-Kernan (Eds.), *Child discourse.* New York: Academic Press, 1977.

Erickson, B., Lind, E. A., Johnson, B. C., & O'Barr, W. M. Speech style and impression formation in a court setting: The effects of "powerful" and "powerless" speech. *Journal of Experimental Social Psychology,* 1978, *14,* 266–279.

Ervin-Tripp, S. Is Sybil there? The structure of some American English directives. *Language in Society,* 1976, *5,* 25–66.

Ervin-Tripp, S. Wait for me, roller skate! In S. Ervin-Tripp & C. Mitchell-Kernan (Eds.), *Child discourse.* New York: Academic Press, 1977.

Gleason, J. B., & Weintraub, S. Input language and the acquisition of communicative competence. In K. E. Nelson (Ed.), *Children's language* (Vol. 1). New York: Gardner Press, 1978.

Hartmann, M. A descriptive study of the language of men and women born in Maine around 1900 as it reflects the Lakoff hypotheses in "Language and woman's place." In B. L. Dubois & I. Crouch (Eds.), *The sociology of the languages of American women.* San Antonio, Tex.: Trinity University, 1976.

Hays, W. L. *Statistics.* New York: Holt, Rinehart & Winston, 1963.

Key, M. R. Linguistic behavior of male and female. *Linguistics,* 1972, *88,* 15–31.

Key, M. R. *Male/female language.* Metuchen, N.J.: Scarecrow Press, 1975.

Lakoff, R. T. *Language and woman's place.* New York: Harper & Row, 1975.

Lakoff, R. T. Language and sexual identity. *Semiotica,* 1977, *19,* 119–130.

McMillan, J. A., Clifton, A. K., McGrath, C., & Gale, W. S. Women's language: Uncertainty of interpersonal sensitivity and emotionality? *Sex Roles,* 1977, *3,* 545–559.

Miller, N. Maruyama, G., Beaber, R. J., & Valone, K. Speed of speech and persuasion. *Journal of Personality and Social Psychology,* 1976, *34,* 615–624.

O'Barr, W. M., & Atkins, B. K. "Women's language" or "powerless language"? In R. Borker, N. Furman, & S. McConnell-Ginet (Eds.), *Language and women's lives: A feminist perspective.* Ithaca, N.Y.: Cornell University Press, in press.

Siegler, D. M., & Siegler, R. S. Stereotypes of males' and females' speech. *Psychological Reports,* 1976, *39,* 167–170.

Smith, B. L., Brown, B. L., Strong, W. J., & Rencher, A. C. Effects of speech rate on personality perception. *Language and Speech,* 1975, *18,* 145–152.

Thorne, B. Review of "Language and woman's place." *Signs,* 1976, *1,* 744–746.

Valian, V. Linguistics and feminism. In M. Vetterling-Braggin, F. A. Elliston, & J. English (Eds.), *Feminism and philosophy.* Totowa, N.J.: Littlefield, Adams, 1977.

Received July 17, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 8, 1304-1313

Effects of Resource Availability and Importance of Behavior on the Experience of Crowding

Richard McCallum, Caryl E. Rusbult, George K. Hong,
University of North Carolina, Chapel Hill

Tedra A. Walden
University of Florida

John Schopler
University of North Carolina, Chapel Hill

According to the interference formulation, participants in a crowded setting will experience interference to the extent that behavioral goals conflict with environmental conditions. The importance of the behavioral goals directly affects not only the magnitude of the interference but also the mechanism by which people cope with interference. It was reasoned that important goals would induce a more active coping strategy in a crowded setting than in an uncrowded setting and would maintain task performance at the price of increasing crowding stress. When the behavioral goal is unimportant, decrements in task performance preclude a rise in stress. A laboratory study manipulated group size, in order to vary the availability of resources, and the importance of the task behavior. The predictions were confirmed, and partial confirmation was obtained for predictions involving the effects of the internal–external personality dimension. The meaning of the results is discussed in terms of other findings in the literature on crowding and the mediating role of the type of mechanism used to cope with interference.

A number of variables have been said to mediate between conditions of high physical density and the experiences of subjective crowding and crowding stress. The variables include excessive stimulation from social sources (Desor, 1972), excessive and unwanted social interactions (Valins & Baum, 1973), restricted behavioral choice (Proshansky, Ittelson, & Rivlin, 1972), decreased interaction distance (Worchel & Teddlie, 1976), and breakdowns in privacy-regulating mechanisms (Altman, 1975). In an attempt to subsume these mediating factors under a single generic variable, Schopler and Stockdale (1977) suggested that interference with goal-directed behavior sequences is the crucial mediator between density and the experience of crowding stress.

The behavioral interference formulation is a sequential model that views physical density as a necessary but not sufficient condition for the occurrence of crowding stress. When the presence of other people in large numbers or in close physical proximity leads to interference with goal-directed behavior, stress is experienced. The degree of stress experienced in a given setting is determined by several factors that affect the impact of interference produced in the setting. These include the anticipated duration of the interference, the importance of the vulnerable response sequences, and personality factors related to anticipated control. This framework accounts for interference in terms of the reward–cost ratio associated with particular behaviors in a given setting. The effect of density-induced interference is to increase the costs of enacting the desired behavioral sequences. This increased cost may be manifested in various forms, such as annoyance, embarrassment, anxiety, or increased effort for making appropriate responses. Implicit in this formulation is the assumption that crowding stress is jointly determined by the physical and social

Requests for reprints should be sent to John Schopler, Psychology Department, University of North Carolina, Chapel Hill, North Carolina 27514.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3708-1304$00.75

characteristics of the setting and the nature of the behavioral goals set by the actor. Interference is experienced to the extent that behavioral goals conflict with environmental conditions.

Schopler, McCallum, and Rusbult (Note 1) provided evidence for the relationship between behavioral interference and the subjective experience of crowding. Subjects in this study worked on a group decision task under conditions of either high or low density. This factor was crossed with a manipulation that imposed structure on the interactions of half of the groups. The imposition of structure reduced the potential for interference in the high-density setting by providing for the coordination and synchronization of responses among group members. Consistent with the interference model, subjects in the high-density conditions perceived less interference and reported less subjective crowding with the imposition of structure. Similar results have been reported by Baum and Koman (1976). These investigators manipulated anticipated group size and found that subjects expecting to meet with a large group felt less crowded when anticipating structured interactions than when no structure was expected.

The present experiment was designed to test the interference model prediction regarding the relationship between crowding stress and the importance to the actor of vulnerable response sequences. One direct way in which high-density settings produce interference with goal-directed behaviors is to place limitations on access to available resources. The term *resource* refers to those features of the physical/social environment that are necessary for the attainment of behavioral goals. In natural settings, resources may include doorways, seating, clear lines of vision or movement, physical or verbal access to others, adequate materials for task completion, and so forth. The designation of a given environmental feature as a *resource* requires specification of the particular behaviors enacted in the setting. Consequently, the setting features required as resources will vary not only from setting to setting but also across individuals within the same setting and across time for a given individual. Despite these potential variations, it is evident that conditions of resource scarcity should produce interference and consequent stress.[1] The first hypothesis of the present study was merely that individuals provided with inadequate resources to complete required tasks would report greater subjective crowding and crowding stress and would perform less well than individuals provided with sufficient resources. This hypothesis is consistent with many points of view, including the interference formulation. Predictions regarding the effects of resource availability and importance upon actual task performance and crowding stress, however, do stem directly from the interference formulation. These effects depend upon the strategy the participants select to cope with interference and their success in maintaining task performance. It is useful to distinguish two general coping strategies for dealing with interference in high-density settings. One strategy involves active intervention of the type required for coordination and synchronization of responses, whereas another involves passive withdrawal from interaction, altering behavioral goals, and lowering expectations (Thibaut & Kelley, 1959). Either of these general strategies may be successful in reducing the degree of conflict between behavioral goals and environmental conditions and may consequently reduce crowding stress. To the extent that the behavioral sequences to be enacted in a setting are important to the individual, however, passive withdrawal and lowered expectations should become less attractive as coping strategies. A second prediction was that given low importance, the scarce-resources condition would evidence a performance decrement relative to the adequate-resources condition. The average effect

[1] The availability of resources may be manipulated in an experimental paradigm by holding group size constant and varying resources or by holding resources constant and varying the number of people who must compete for them. Group size was manipulated in the present experiment. Because the functional importance of group size, in our experimental setting, was in relation to the availability of task resources, we have labeled the manipulation *resource availability* throughout the article, rather than *group size* or *density*.

of resource availability upon task performance predicted by Hypothesis 1 is therefore modified by the interaction contrasts described by Hypothesis 2.

The interference formulation predicts that crowding stress increases in magnitude with increases in the importance of vulnerable response sequences. This effect of course assumes that the conditions expected by Hypothesis 2 will pertain, that is, that raising the importance of the goal will induce an active coping strategy even in the face of high interference. Under these restricted conditions, variations in resource availability should have an effect when behavioral goals are important but not when behavioral goals are unimportant. Specifically, among individuals working for important goals, subjects given inadequate resources should experience more crowding stress than should subjects given adequate resources. In contrast, individuals working for unimportant goals should be less affected by variations in resource availability. The average effect of resource scarcity upon crowding stress predicted by Hypothesis 1 is therefore modified by the interaction contrasts described by Hypothesis 3.

In addition to variables associated with the setting characteristics and the nature of the participants' behavioral goals, the present experiment investigated the effect of a personality variable that is conceptually related to the interference analysis. Stockdale (1978) has emphasized the importance of the perception of environmental control in determining the degree of actual or anticipated interference experienced in high-density settings. To the extent that the potential for exerting control over the social environment (through coordination and synchronization of behaviors) is salient to participants, the negative effects of interference on subjective crowding and crowding stress should be mitigated. Individuals who typically anticipate a high degree of personal control in social situations should therefore experience fewer negative effects of density-induced interference. A scale designed to assess individual expectancy of internal or external control over the social environment (Schopler, Langmeyer, Stokols, & Reisman, 1973) was employed to locate subjects on this personality dimension. The fourth hypothesis was that within the scarce-resources (high interference) condition, individuals identified as externals should experience greater crowding and crowding stress than should internals. No internality–externality differences in subjective crowding and crowding stress were predicted within the adequate-resources condition because the existence of adequate resources should obviate the salience of control considerations. Such results would be consistent with the findings that externals report more subjective crowding under high-interference conditions and respond more favorably than do internals to manipulations designed to reduce interference experienced in high-density situations (Schopler, McCallum, & Rusbult, Note 1). These results parallel the findings of research that directly manipulated perceived control. For example, Glass and Singer (1972) found fewer behavioral effects and aftereffects of another environmental stressor (noise) when subjects believed they controlled the onset of the aversive stimulus, and Sherrod (1974) demonstrated that subjects who controlled when they could leave a crowded setting showed greater tolerance for frustration on a postcrowding task.

Because the interference formulation and many other conceptions of crowding speculate about the arousal and reduction of crowding stress, identification of an adequate measure of this kind of stress is of general importance. The obvious limitations of self-reports of stress, among other considerations, have led some investigators to seek physiological indices. Although it is by no means evident that a variable as complex as crowding stress will have a reliable physiological counterpart, the literature contains some suggestions, such as the Palmar Sweat Index (PSI; Johnson & Dabbs, 1967) or blood pressure (D'Atri, 1975). We undertook to explore the adequacy of PSI as a measure of crowding stress in the present study.

Method

Subjects

Seventy-two males and 72 females participated in the experiment in partial fulfillment of the re-

quirements for an introductory psychology course. Subjects were assigned to 16 same-sex six-person and 16 same-sex three-person groups. Four six-person and four three-person groups of each sex were randomly assigned to two importance conditions.

Procedure

The study employed four independent variables in a complete $2 \times 2 \times 2 \times 2$ factorial design: availability of resources (scarce or adequate), importance of behavior (high or low), locus of control (internal or external subject orientation), and sex of subject.

Upon arrival at the laboratory, subjects were escorted to a waiting room by one experimenter who was of the same sex as the subjects and were given numbered clipboards. (Two women and two men served as experimenters for the study.) The experimenter briefly described the procedure as a simulation of working conditions in a large business organization. The subjects were told that they would be asked to perform clerical work under controlled conditions so that factors affecting work quality and job satisfaction could be evaluated. Subjects wishing to terminate participation in the experiment without penalty were given the opportunity to do so at this point. The experimenter then administered the North Carolina Internal–External Scale (Schopler et al., 1973). Instead of categorizing individuals relative to the sample median, locus of control assignments were determined for each subject on the basis of the standardization mean (if his/her score was 72 or lower, the subject was assigned internal status, and if his/her score was 73 or higher, the subject was assigned external status).

In order to obtain a measure of physiological stress (the Palmar Sweat Index) without arousing subject suspicion as to its purpose, the experimenter requested the subjects' cooperation in helping another graduate student (the second experimenter present at the session) pretest for an experiment he/she was preparing to run. The second experimenter stated that he/she wanted to pilot test a measure of concentration called the "PCI" and explained that in order to assess the stability of the measure across various activity levels, several measures would be obtained throughout the experimental session. The PSI was obtained by painting a special chemical on the fingertip and transferring the print to a slide with a piece of transparent tape. After demonstrating the administration of the PSI on the primary experimenter, the second experimenter obtained a baseline palmar sweat measure from each subject.

The primary experimenter then accompanied the group to a mock office containing six (or three) chairs placed in a semicircle facing three standard four-drawer file cabinets. Subjects were instructed to occupy the chair that corresponded to their respective clipboard numbers. Availability of resources was manipulated by variations in the size of the group. The resources available for work on the task (such as physical space and equipment) remained constant across all conditions. The mock office itself measured 2.86 m by 4.11 m and contained the subjects' chairs as well as the three file cabinets placed against one wall and spaced 60 cm apart. While these resources were sufficient for three-person groups, they were insufficient when divided among the members of larger groups. That is, while members of smaller groups seldom experienced difficulty in obtaining access to the files, resources were insufficient and led to problems of interference and a greater need for coordination of behaviors in the larger groups.

The experimenter distributed a different list of names to each subject and explained the experimental task. The file cabinets contained a set of alphabetically ordered folders, and the subjects' task was to locate the file for each name on their lists and to record the home address listed for each of those names. Subjects were told that they would receive "credit" for names only if their records of each address corresponded exactly to that on the experimenter's master list. Subjects were also asked to complete the names in the order in which they appeared on the lists.

Subjects in the low-importance condition were instructed to work as quickly and as carefully as possible and were told that individuals were typically able to complete from 20 to 60 names during the experiment. The other half of the subjects, those in the high-importance condition, were told that in order to simulate piecework schedules they would be paid 15 cents for each name they located and correctly identified. These subjects were informed that typical earnings in the experiment ranged from \$2 to \$8, the average being about \$5. The offer of monetary incentives contingent on high quality and quantity performance was intended to increase the relative importance of task-related behaviors.

After 5 minutes subjects' task work was interrupted to obtain a second PSI reading from each subject. Subjects then resumed work for an additional 5 minutes, after which the primary experimenter administered the experimental questionnaire. Subjects were then informed that the task work was completed, were escorted back to the waiting room, and were thoroughly debriefed. Subjects in the high-importance condition were paid \$5 each for participation in the experiment (this amount was greater than that actually earned by any subject).

Dependent Variables

The experimental dependent variables comprised four separate conceptual groupings. The first group was designed to measure crowding stress and included ratings of subjects' reported stress, per-

sonal feelings, attitudes toward other group members, and intentions concerning task performance. The second set consisted of measures of subjective crowding, whereas the third and fourth sets consisted of single measures—task performance and Palmar Sweat Index, respectively. Items were also included to evaluate the effectiveness of the manipulations of availability of resources and importance of the behavior.

Nine items were included in the crowding stress group. Affective reactions were assessed by requiring the subjects to rate both their "normal" feelings and their feelings during the experiment on four pairs of 9-point semantic differentials (pleasant–unpleasant, cooperative–uncooperative, nervous–calm, friendly–unfriendly). These ratings were then converted to difference scores (experimental rating less normal rating) to facilitate statistical analyses. Subjects also rated the degree to which the experience of working in the experimental setting was stressful, the extent to which task work was distressing compared to the subject's experience during a typical day, the extent to which their expectations concerning task performance decreased during work on the task, the difficulty of coordination with other group members, and the extent to which others interfered with task work.

Subjective crowding was assessed by six 9-point bipolar scales. These items asked: "Is there (too much – too little) room for you to work on the task?" "To what extent do you feel (extremely crowded – extremely uncrowded) by the other group members?" "Is there (too much – too little) room for you to feel comfortable while working on the experimental task?" "Do you feel (extremely unconfined – extremely confined) physically while working on the task?" "Do working conditions in this room make task work (much harder – much easier)?" "To what extent is the experimental room (congested – uncongested)?"

Task performance was evaluated by simply counting the number of items correctly completed by each subject. A baseline and task-related PSI score was computed for each subject. Two raters ignorant of the experimental conditions scored the PSI measure by counting the number of open sweat pores within a standard 4-mm square area for both the baseline and the task-related measures. Scores of the two raters were reliable and yielded a correlation of .97.

Two 9-point bipolar scales assessed the effectiveness of the availability-of-resources manipulation and required that subjects estimate the extent to which they would have performed better had they been alone in the room and indicate whether there were too many people in the room.

Four 9-point bipolar scales assessed the effectiveness of the manipulation of the importance of the behavior. These measured liking for the task, importance of task performance, difficulty of the task, and amount of effort expended on task work.

Results

Manipulation Checks

Except where otherwise indicated, the individual responses on each dependent measure were averaged and the group mean employed as the unit of analysis, avoiding problems associated with possible response dependence within experimental groups. To assess the effectiveness of the availability-of-resources manipulation, a two-factor analysis of variance was performed on the measures associated with this factor. Subjects in the large groups were more likely than those in small groups to report that they would have performed better had they been alone in the room and that there were too many people in the room. The means on the ability-to-perform measure were 8.07 for the scarce-resources condition and 7.44 for the normal-resources condition, and the respective means for the number-of-people measure were 7.20 and 6.23, multivariate $F(2, 27) = 16.92$, $p < .001$. Thus it appears that the manipulation of availability of resources was successful.

The success of the importance-of-behavior manipulation was assessed by four items designed to measure subjects' reactions to the experimental task. Subjects in the high-importance condition reported greater liking for the task, stated that it was more important for them to perform well on the task, perceived task work as more difficult, and estimated that they had expended greater effort on task work than did subjects in the low-importance condition. A two-factor analysis of variance revealed that the difference between the two conditions on these measures was only marginally significant, multivariate $F(4, 25) = 2.31$, $p < .086$. Although the importance manipulation was judged satisfactory, the marginal significance of the mean differences should be recalled in evaluating the results of the experiment.

Tests of the Major Hypotheses

Initial analyses revealed that no significant effects were obtained with the nervous–calm and friendly–unfriendly measures, and these

Table 1
Mean Values of all Dependent Measures as a Function of Importance of Behavior and Resource Availability

Item	Importance of behavior			
	High		Low	
	Scarce resources	Adequate resources	Scarce resources	Adequate resources
Behavior				
Performance score	15.40	16.92	14.40	16.83
Crowding stress				
Unpleasant	1.71	.92	1.81	1.21
Uncooperative	.94	.21	.92	.42
Difficulty of coordinating with others	5.35	4.42	4.85	4.67
Stress	4.77	3.13	3.90	3.13
Distress	4.90	4.17	5.31	3.96
Others interfere	5.92	5.33	5.88	4.63
Lowered expectations	5.27	4.75	5.10	4.71
Subjective crowding				
No room to work	7.21	6.17	7.10	6.21
Crowded	6.96	6.08	6.96	6.00
No room to feel comfortable	6.90	5.92	6.77	5.92
Confined	6.27	5.67	6.13	5.54
Congested	6.81	5.71	6.79	5.75
Working conditions poor	7.58	6.50	7.67	6.33
Physiological reactions				
Task-related PSI score	58.39	53.33	51.52	55.46

Note. PSI = Palmar Sweat Index.

variables were not considered further. Although no experimental hypotheses were advanced with respect to sex of subject, the influence of this factor was examined in order to assess the universality of the experimental results. A three-factor analysis of variance was performed on each of the four sets of dependent variables. There were no significant multivariate main effects nor interactions associated with sex of subject, and this factor was not included in the remaining analyses.

The first experimental hypothesis predicted that scarce resources would produce greater subjective crowding, crowding stress, and task decrements than would adequate resources. A two-factor analysis of variance revealed a significant effect of resource availability on participants' subjective crowding, multivariate $F(6, 23) = 9.90$, $p < .001$; crowding stress, multivariate $F(7, 22) = 2.48$, $p < .05$; and task performance, $F(1, 28) = 8.66$, $p < .006$. Participants in the scarce-resources condition reported greater subjective crowding and greater crowding stress and performed less well on the experimental task than did participants in the adequate-resources condition. The resource-availability manipulation did not significantly affect participants' physiological reactions to the setting (as measured by the PSI score, covarying initial PSI levels). The significant effects, in essence, confirm that the resource-availability manipulation established the circumstances required to test the hypotheses involving importance. Induction of scarce resources by an increase in group size did create the required crowding and interference with actual task performance. The lack of an effect for PSI scores precluded consideration of this measure. We will return to this regrettable fact in the discussion.

Hypotheses 2 and 3, stated as planned comparisons, were tested by a series of simple effects analyses because standard analysis of variance does not provide the specific contrasts necessary to test the hypotheses. It is

necessary to examine the influence of variations in availability of resources within both the high-importance and the low-importance conditions. It will be recalled that the second hypothesis expected high, but not low, importance to maintain task performance in the face of scarce resources. Consistent with the experimental hypothesis, scarcity of resources significantly affected performance among low-importance subjects, $F(1, 28) = 6.57$, $p < .02$, but did not significantly influence performance in the high-importance condition, $F(1, 28) = 2.56$, $p < .12$. Average performance scores are shown in Table 1. It is apparent that subjects in the high-importance condition produced a uniformly high level of performance (these subjects averaged 16.16 completed items per session), whereas subjects in the low-importance condition responded to variations in availability of resources. Within the low-importance condition, scarce-resources subjects averaged 14.4 completed items, whereas adequate resources subjects averaged 16.8 items. When motivated by the potential for monetary gain, subjects overcame the interference created by scarce resources and performed at a high level of productivity. When task behaviors were of relatively low importance, however, variations in availability of resources produced commensurate decrements in task performance.

In accord with the interference formulation, Hypothesis 3 predicted that subjects' crowding stress would be adversely affected by scarcity of resources when the task behavior was of high importance to the subject but not when it was of low importance. In support of this prediction, the crowding stress measures were influenced by variations in availability of resources within the high-importance condition, multivariate $F(7, 22) = 2.40$, $p < .05$, but not in the low-importance condition, multivariate $F(7, 22) = 1.62$, $p < .18$. Mean responses on these dependent variables are presented in Table 1, where it can be seen that for the seven stress items, only two (*distress* and *others interfere*) do not show the predicted pattern. With hindsight, these two items appear to be deficient in tapping events that were not strictly task related.

Locus of Control

It was predicted that among scarce-resources subjects in the high- and low-importance conditions, the stress responses, subjective crowding, and physiological reactions of external subjects would be more negative than those of internal subjects. This hypothesis was tested by simple effects analyses and was by necessity performed using the individual rather than the group as the unit of analysis. It should be noted that the use of locus of control as an independent variable produced a nonorthogonal experimental design. The reader is referred to Appelbaum and Cramer (1974) for a description of the series of ignoring and eliminating procedures that were employed to produce accurate tests of significance.

Analysis of subjects' stress responses indicated that externals reported more negative affect than did internals in the high-importance scarce-resources condition, multivariate $F(7, 130) = 2.11$, $p < .05$, but within the low-importance scarce-resources condition, internality–externality did not have the predicted effect, multivariate $F(7, 130) = 1.15$, $p < .34$. As expected, internality–externality did not significantly influence affective responses within the adequate-resources condition.

A similar pattern of results was obtained for the measures of subjects' subjective crowding. Among scarce-resources subjects, externals consistently responded more negatively than did internals, but the effect for internality–externality within the high-importance scarce-resources condition is significant, multivariate $F(6, 131) = 2.33$, $p < .04$, whereas the internality–externality effect within the low-importance scarce-resources condition was not significant, multivariate $F(6, 131) = 0.21$, $p < .98$. Again, internality–externality had no significant effect on subjects' responses in the adequate-resources condition.

There were no locus-of-control main effects nor interactions on the measures of performance or physiological stress.

Discussion

The major focus of the present research was to evaluate the effects of varying the importance of the behaviors vulnerable to interference by others. We successfully induced crowding and interference through the creation of scarce resources by manipulating group size. Individuals working in the scarce-resources condition, compared to those working in the adequate-resources condition, reported more crowding and crowding stress and completed fewer items on the task. These effects are hardly remarkable, although they do underscore the reason why some experiments, for example, those of Freedman, Klevansky, & Ehrlich (1971), find no performance decrements with high density, whereas other experiments, such as ours or that of Heller, Groff, & Solomon (1977), do find performance decrements. The difference appears to reside in whether or not the crowding manipulation and the particular task used intersect to produce scarce task resources and interference. Freedman, Klevansky, and Ehrlich used problem-solving tasks and manipulated room size, a combination that was selected explicitly to minimize task interference by unhinging high density from scarce resources. In our experiment and that of Heller, Groff, and Solomon, a performance task was combined with the heightening of density by increasing group size. This combination produces task interference and performance decrements.

Based on the interference model (Schopler & Stockdale, 1977), it was predicted that the aversive impact of interference would be greater when behaviors encountering interference are important to the actor. The second and third hypotheses expected efforts to maintain task performance, when the goal was important and despite scarce resources, by an active coping strategy that also produced increased stress. The data confirmed these predictions for task performance and for stress. Subjects responded more negatively to scarce resources on the stress measures only when they were provided with incentives to perform well on the task. In accord with the interference analysis, subjects who were highly motivated to complete goal-directed behavior sequences experienced more psychological costs than did less motivated subjects when these behaviors encountered interference from the physical/social environment. These results suggest that in understanding the impact of high-density settings, specifying the nature and importance of behaviors enacted in these settings is as necessary as specifying the nature of the settings themselves.

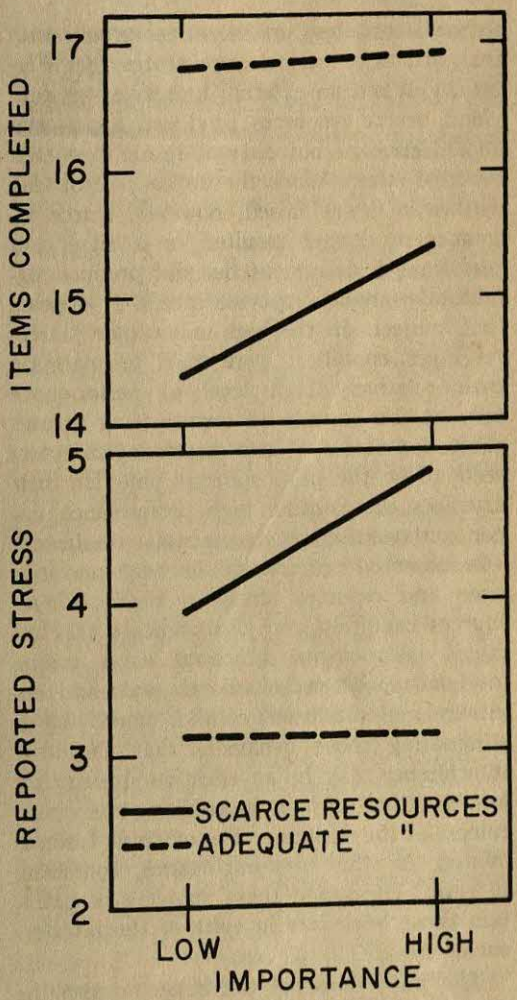


Figure 1. Mean number of items completed and reported stress as a function of resource availability and level of importance.

Further evidence of the relationship between reported stress and the interference produced by high-density settings is revealed in a comparison of the pattern of responses

to the stress item, in the stress group, with the pattern of performance scores (see Figure 1). When no external incentive was provided, scarce resources produced a performance decrement but did not result in greater reported stress. When the importance of task performance was raised, however, scarce resources no longer resulted in a significant performance decrement but did produce significantly greater reported stress. It appears that subjects in the high-importance scarce-resources condition were able to maintain their relatively high level of performance only at the expense of experiencing greater subjective stress. These psychological costs seem to be the price subjects paid for their attempts to maintain high performance under unfavorable environmental conditions. The observed relationship between performance and reported stress in the scarce-resources condition may also indicate that reduced task output functions as a coping mechanism. When behavioral goals and environmental conditions conflict, modifying or eliminating those behaviors that encounter interference may be an effective strategy for low-incentive subjects in reducing the stressfulness of the situation. The extrinsic rewards offered in the high-importance condition, however, encourage these subjects to maintain those behaviors in spite of the interference.

The importance of the behavior and the scarcity of task resources may also be the critical features determining when crowding and sex of subject interact with respect to affective feelings. Increases in density have raised positive feelings for women while lowering them for men, both when the measures are taken in the crowded situation (Freedman, Levy, Buchanan, & Price, 1972; Ross, Layton, Erickson, & Schopler, 1973) and when they are taken after the participants have left the crowded situation (Epstein & Karlin, 1975). Affective ratings are included in our crowding stress group. It will be recalled that we found no interactions with sex of subject on these measures. We would suggest that the sex interaction effect will not occur when the mechanism used to cope with interference does not lower crowding stress. The cohesion and cooperativeness that characterizes groups of women, according to Epstein and Karlin, will probably not survive the combination of scarce resources and high importance.

Although it was predicted that externals would respond more negatively to the scarce resources present in the large group under both high and low importance, this effect was observed only in the high-importance condition. A possible explanation of this result is suggested by reference to our previous discussion regarding the most likely coping strategies in the high- and low-importance conditions. A passive withdrawal strategy, resulting in reduced performance, may seem equally attractive to internals and externals when incentive is low. The high-importance condition favors internals because it requires a more active strategy aimed at coordinating and synchronizing responses to keep performance high. Thus, individual differences in expectancies for control over the social environment produce differing affective responses when the desirable coping strategy requires active intervention.

No significant effects were obtained for the physiological measure of stress. One of the problems with attempting physiological measures in crowding studies is the disrupting and distracting influence of the measurement process itself. The PSI was selected for use because it allowed subjects the necessary mobility and could be administered relatively quickly. To further reduce the disruptions inherent in taking the measure, the PSI was administered only twice. The data, which reflect generally higher scores on the initial administration, suggest that a longer baseline period with several administrations would have been desirable. The longer baseline period would allow subjects an opportunity to adapt to the measurement procedure itself and might have resulted in a greater sensitivity to the experimental manipulations. However, the nonsignificant results for the PSI must also lead us to entertain the possibility that physiological stress does not in actuality adhere to the pattern obtained for self-reported stress and negative affect. This seems plausible in view of the relatively brief

duration of the experimental session. In any case, we feel that the development of a sensitive and easily administered measure of physiological stress would add much to a literature that has relied heavily on self-report measures.

The results of this study have been interpreted in support of the interference model of crowding. This model locates the aversive impact of high-density settings in the increased costs of enacting behaviors when environmental conditions are in opposition to the actor's behavioral goals. Aversive psychological consequences were shown to be a joint function of the degree of density-related interference and the importance to the actor of his or her goal-directed behaviors. The relationship between density and actual task performance was also qualified by the importance of the task-related behaviors. When highly motivated individuals are successful in maintaining high levels of performance under high-density conditions, the increased costs of enacting the necessary behaviors are reflected in their negative affective responses to the situation.

Reference Note

1. Schopler, J., McCallum, R., & Rusbult, C. E. *Behavioral interference and internality–externality as determinants of subject crowding.* Unpublished manuscript, 1978. (Available from J. Schopler, Psychology Department, University of North Carolina, Chapel Hill, North Carolina 27514.)

References

Altman, I. *The environment and social behavior: Privacy, personal space, territory and crowding.* Monterey, Calif.: Brooks/Cole, 1975.

Appelbaum, M. I., & Cramer, E. M. Some problems in the nonorthogonal analysis of variance. *Psychological Bulletin,* 1974, *81,* 335–343.

Baum, A., & Koman, S. K. Differential response to anticipated crowding: Psychological effects of social and spatial density. *Journal of Personality and Social Psychology,* 1976, *34,* 526–536.

D'Atri, D. A. Psychophysiological responses to crowding. *Environment and Behavior,* 1975, *7,* 237–252.

Desor, J. Toward a psychological theory of crowding. *Journal of Personality and Social Psychology,* 1972, *21,* 79–83.

Epstein, Y. M., & Karlin, R. A. Effects of acute experimental crowding. *Journal of Applied Social Psychology,* 1975, *5,* 34–53.

Freedman, J. L., Klevansky, S., & Ehrlich, P. R. The effect of crowding on human task performance. *Journal of Applied Social Psychology,* 1971, *1,* 7–25.

Freedman, J. L., Levy, A. S., Buchanan, R. W., & Price, J. Crowding and human aggressiveness. *Journal of Experimental Social Psychology,* 1972, *8,* 528–548.

Glass, D., & Singer, J. *Urban stress: Experiments on noise and social stressors.* New York: Academic Press, 1972.

Heller, J. F., Groff, B. D., & Solomon, S. H. Toward an understanding of crowding: The role of physical interaction. *Journal of Personality and Social Psychology,* 1977, *35,* 183–190.

Johnson, J. E., & Dabbs, J. M., Jr. Enumeration of active sweat glands: A simple physiological indicator of physiological changes. *Nursing Research,* 1967, *16,* 273–276.

Proshansky, H. M., Ittelson, W. H., & Rivlin, L. G. Freedom of choice and behavior in a physical setting. In J. F. Wohlwill & D. Carson (Eds.), *Environment and the social sciences: Perspectives and applications.* Washington, D.C.: American Psychological Association, 1972.

Ross, M., Layton, B., Erickson, R., & Schopler, J. Affect, facial regard, and reactions to crowding. *Journal of Personality and Social Psychology,* 1973, *28,* 69–76.

Schopler, J., Langmeyer, D., Stokols, D., & Reisman, S. The North Carolina Internal–External Scale: Validation of the short form. *Research Previews,* 1973, *20,* 3–12.

Schopler, J., & Stockdale, J. E. An interference analysis of crowding. *Journal of Environmental Psychology and Nonverbal Behavior,* 1977, *1,* 81–88.

Sherrod, D. R. Crowding, perceived control and behavioral aftereffects. *Journal of Applied Social Psychology,* 1974, *4,* 171–186.

Stockdale, J. E. Crowding: Determinants and effects. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 11). New York: Academic Press, 1978.

Thibaut, J. W., & Kelley, H. H. *The social psychology of groups.* New York: Wiley, 1959.

Valins, S., & Baum, A. Residential group size, social interaction and crowding. *Environment and Behavior,* 1973, *5,* 421–439.

Worchel, S., & Teddlie, C. The experience of crowding: A two-factor theory. *Journal of Personality and Social Psychology,* 1976, *34,* 30–40.

Received July 19, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 8, 1314–1324

Identifying Victims of Groupthink From Public Statements of Decision Makers

Philip E. Tetlock
Yale University

The present study attempts to test hypotheses derived from Janis's groupthink analysis of several foreign policy decisions of the American government. Content analyses were performed using the public statements of key decision makers involved in crises for which Janis's case studies revealed or did not reveal evidence of groupthink. Consistent with Janis's theory, it was found that, relative to non-groupthink decision makers, groupthink decision makers were more simplistic in their perceptions of policy issues and made more positive references to the United States and its allies (own group). Inconsistent with Janis's theory, groupthink decision makers did not make significantly more negative references to Communist states and their allies (opponents). Possible explanations for the findings are assessed, and methodological and practical obstacles to definitive testing of the groupthink model are examined.

Investigators have taken diverse theoretical approaches to the problem of how psychological factors influence political decision making. Most of this research has focused upon psychological processes that occur within individuals: unresolved psychodynamic conflicts (e.g., Glad, 1973); the needs for achievement, affiliation, and power (Winter, 1973); cognitive maps of the political world (e.g., Abelson, 1973; Axelrod, 1976; Holsti, 1976; Jervis, 1976); pressures toward cognitive consistency (e.g., Jervis, 1976); or the effects of stress on information processing complexity (e.g., Hermann, 1972; Suedfeld & Tetlock, 1977).

Political decisions are not, of course, made in a social vacuum. They are usually made in organized group contexts in which implicit and explicit norms regulate the conduct of the decision maker. However, relatively few psychologically oriented studies have considered how patterns of social interaction among policy makers can influence decision making. One noteworthy exception to this generalization is the work of Janis (1972) on the *groupthink* phenomenon. Janis argued that intense social pressures toward uniformity and in-group loyalty within decision-making groups can build to the point where they seriously interfere with both cognitive efficiency and moral judgment. Groupthink occurs when independent critical analysis of the problem facing the group assumes second place to group members' motivation to maintain group solidarity and to avoid creating disunity by expressing unpopular doubts or opinions.

In several case studies of major foreign policy decisions by the American government, Janis attempted to trace the effects of social pressures toward groupthink on decision making. He selected for analysis cases in which he felt the signs of poor decision making as a result of concurrence seeking were "unmistakable" (Janis, 1972, p. 10). These included the decisions to pursue the

The author is very grateful to the many individuals without whose help this research would have been impossible. Professors Irving Janis and Peter Suedfeld provided many extremely useful suggestions throughout the project. I have also received invaluable practical assistance from Terry Murray, Elizabeth Ballard, and Ariel Levi, among others.

Philip E. Tetlock is now at the Department of Psychology, University of California, Berkeley.

Requests for reprints should be sent to Philip E. Tetlock, Department of Psychology, University of California, Berkeley, California 94720.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3708-1314$00.75

defeated North Korean army beyond the 38th parallel, to launch the Bay of Pigs invasion of Cuba, and to escalate involvement in the Vietnam War. In each situation, Janis discovered similar predisposing conditions to groupthink. The policy makers invariably belonged to cohesive groups characterized by high levels of mutual attraction among the members. The policy-making groups were relatively insulated from the judgments of qualified outsiders and lacked systematic procedures for evaluating and searching out new evidence relating to the problem. During the deliberations on the particular issues, the group leaders tended to promote their preferred solutions rather than to encourage open-minded, careful analysis of policy alternatives. Finally, all decisions were made in highly stressful situations in which policy makers were not hopeful of finding an alternative superior to the one that the group currently favored.

Under these conditions, certain distinctive attitudes and patterns of interaction emerged within the groups (see Janis, 1972, pp. 197–198). Decision makers appeared to believe that the group was invulnerable to outside attack—a belief that encouraged excessive optimism and risktaking. Group members undertook collective efforts to discount warnings concerning the popularly agreed-upon solution and exhibited unquestioning belief in the inherent righteousness of the group's policies and the inherent immorality and incompetence of the enemy. Decision makers often self-censored any personal doubts about the group's policies. The group applied direct pressure to those few members who showed signs of deviating from the consensus. Self-appointed "mindguards" shielded the group from external sources of dissonant or counterattitudinal information.

In a social atmosphere so inhospitable to rigorous analysis of policy options, decision-making procedures fell far short of the ideal "rational actor" standard (cf. Allison, 1971). According to Janis and Mann (1977), decision making in groupthink situations failed to meet any of the major procedural criteria that they used to determine whether decision-making procedures were of high quality. Groupthink policy makers generally did not (a) adequately survey the full range of policy alternatives; (b) consider the full spectrum of objectives that might be affected by the chosen policy; (c) obtain adequate information for evaluating the alternative policies; (d) weigh the costs and benefits of each alternative carefully; (e) take proper account of information that contradicted prior beliefs and preferences; (f) reexamine evaluations of all known alternatives, including those previously regarded as unacceptable; (g) develop sufficiently detailed plans for implementing the chosen policy, with special reference to contingency plans in the event known risks materialized. As Janis (1972, p. 10) noted, the outcomes of groupthink decisions *deserved* to be fiascoes "because of the grossly inadequate way the policy makers carried out their decision-making tasks."

Janis contrasted the groupthink decisions with two examples of well-worked-out foreign policy decisions: the development of the Marshall Plan and the handling of the Cuban Missile Crisis. In these instances, the decision-making groups and their leaders gave high priority to critical appraisal and open discussion of policy options. "Decision makers had to undergo the unpleasant experience of hearing their pet ideas critically pulled to pieces" (Janis, 1972, p. 165). The policies ultimately developed within these groups were based upon careful analysis of the likely consequences of large numbers of policy options, with frequent attempts at proposing new solutions that maximized the advantages and minimized the disadvantages of options already analyzed.

Janis provided no quantitative evidence bearing on the validity of his classification of foreign policy deliberations into the groupthink and non-groupthink categories.

> For purposes of hypothesis construction—which is the stage of inquiry with which this book is concerned—we must be willing to make some inferential leaps from whatever historical clues we can pick up. (Janis, 1972, p. v)

Unfortunately, the most appropriate documents for testing the groupthink hypotheses

—verbatim records of formal and informal meetings among decision makers—were not available. The historical clues upon which Janis relied—observer accounts of private conversations and participant memoirs—were susceptible to serious retrospective distortion of a motivational or cognitive nature (cf. Fischhoff & Beyth-Marom, 1976, on the "certainty-of-hindsight" bias). Moreover, Janis did not specify the criteria that he used in deciding to include or exclude data. Thus, he may inadvertently have given too much weight to evidence that supported group think hypotheses and too little weight to contradictory evidence.

A wide range of behavioral research methods can be used to test Janis's analysis. Two possible approaches are (a) laboratory simulation studies that factorially manipulate hypothesized determinants of groupthink and then observe effects on social interaction and quality of decision making, and (b) content analysis studies of archival records that cast light on how decision makers think in actual crises. Each approach has compensating methodological advantages and disadvantages.

Flowers's (1977) research exemplifies the first approach. In a 2 × 2 laboratory experiment, Flowers found that "open-leadership style" led groups to suggest more solutions to a problem and to use more available information than "closed-leadership style." This finding was compatible with Janis's (1972) analysis. Contrary to Janis, though, "degree of group cohesiveness" did not affect quality of decision making. This latter finding is, however, difficult to interpret. On the one hand, Janis's theory may need revision. On the other hand, Flowers's operational definition of cohesiveness (using groups of college students) probably differs from the cohesiveness displayed within foreign policy–making groups (who have typically known each other for many years and are confronted by extremely ego-involving tasks). The interpretive ambiguity reflects a fundamental limitation of simulations: the impossibility of ensuring that a simulation fully captures the processes operating in the original situation (cf. Bem, 1972).

The present study takes the second approach. It applies standardized content analysis procedures to statements of key decision makers involved in the groupthink and non-groupthink crises that Janis examined. Relative to laboratory studies, the primary advantage of this approach is the high external validity achieved by dealing with real-life decisions by top policy makers; the primary disadvantage is the relative lack of control over both independent and dependent variables.

With respect to independent variables, the laboratory researcher can often be confident that his or her experimental conditions differ in only a few theoretically relevant ways. Whether the same can be said for the hypothesized groupthink and non-groupthink foreign policy decisions is unclear. As is almost inevitable in comparing complex naturally occurring phenomena, possible confounding variables exist. For instance, the groupthink decisions all involved military interventions or escalations of questionable success; the non-groupthink decisions involved more restrained policies having desirable consequences (from the standpoint of the decision makers). It is debatable, however, whether these differences are independent of or attributable to groupthink. For example, Janis (1972) suggested that groupthink predisposes decision makers to seek rapid, clear-cut military solutions to complex problems. Nonetheless, it is important to keep in mind that the groupthink/non-groupthink contrast in this study can be reconceptualized as a contrast between unsuccessful military and successful nonmilitary national policies.

With respect to dependent variables, the laboratory researcher has relatively direct access to subjects' decision-making deliberations. Such records are much less accessible when one's subjects are heads of state and their advisors. Currently, the only standardized records of how groupthink and non-groupthink decision makers perceived policy options consist of decision makers' public statements during the relevant crises. This need not, however, be a serious liability. Although public statements are undoubtedly more influenced by efforts to manage polit-

ical impressions than are private statements, there are good reasons to assume that manifestations of groupthink will appear in both private and public statements. Political elites —for a variety of reasons—probably do not frequently endorse policy positions that are very dissimilar to those they endorse in private (cf. Graber, 1976). Political leaders are usually concerned with maintaining an image of trustworthiness that would be jeopardized by failure to maintain some consistency between their words and deeds (cf. Graber, 1976; Tedeschi, Schlenker, & Bonoma, 1971). To the extent that groupthink seriously interferes with the cognitive efficiency and moral judgment of decision makers, groupthink also probably affects how decision makers attempt to justify their policies to the world.

The study reported here selected content analysis procedures for their appropriateness in identifying two major manifestations of groupthink: (a) the tendency to process policy-relevant information in simplistic and biased ways, and (b) the tendency to evaluate one's own group highly positively and to evaluate one's domestic and international opponents highly negatively. These variables were selected partly because they are more likely to be revealed at an overt or public level than others (e.g., self-censorship of doubts, direct pressuring of deviants, or the emergence of mindguards) and partly because reliable and valid measurement techniques could be readily adapted for the purpose of assessing them.

The integrative-complexity coding system was used to determine the degree to which "simpler" (i.e., less differentiated and less integrated) modes of information processing prevailed in groupthink situations. The coding system was initially developed as a measure of the integrative-complexity dimension of personality (Harvey, Hunt, & Schroder, 1961; Schroder, 1971; Schroder, Driver, & Streufert, 1967). Recent research indicates that the coding system is also sensitive to situationally induced shifts in complexity of information processing and can be usefully applied to such documents as letters, essays, speeches, and diplomatic communications (Suedfeld, 1978; Suedfeld & Rank, 1976; Suedfeld & Tetlock, 1977; Suedfeld, Tetlock, & Ramirez, 1977). For instance, Suedfeld and Rank (1976) have shown that the integrative complexity of successful revolutionary leaders who subsequently held onto power was higher after victory than before victory, whereas leaders who failed to maintain power after victory did not show such a change. They argued that this finding reflects the need for change from the simple cognitive strategies appropriate to a revolutionary struggle toward the complex approach required of governments in power. Suedfeld et al. (1977) found that the complexity of Arab and Israeli speeches to the United Nations tended to decline prior to the outbreak of major wars. They argued that this finding probably reflects the disruptive effects of stress on complex information processing (see also Suedfeld & Tetlock, 1977). Such data strongly suggest the usefulness of the complexity coding system for unobtrusively assessing cognitive structural variables from archival data and, most relevant here, for testing Janis's groupthink analysis.

Evaluative assertion analysis was used to test the hypotheses concerning the effects of groupthink on "in-group" and "out-group" attitudes (see Osgood, 1959; Osgood, Saporta, & Nunnally, 1956). The technique has grown out of Osgood's research on the dimensions of meaning and the congruity principle of attitude change. It has been used successfully in a wide variety of research projects, including the coding of psychotherapeutic interviews (Osgood, 1959), the examination of John Foster Dulles's attitude toward the Soviet Union (Holsti, 1967), and the assessment of bias in the news coverage of presidential candidates (Westley et al., 1963). The common purpose underlying the diverse applications of evaluative assertion analysis has been "to extract from verbal communications the evaluations being made of significant concepts" (Osgood et al., 1956, p. 47). The significant concepts in the present context are defined as *groups with which the speaker identifies* and *domestic and foreign opponents*.

In summary, this study employed two independent techniques of content analysis to test hypotheses abstracted from Janis's case studies of groupthink in American foreign

policy making. The integrative-complexity coding system was used to assess the relative differentiation and integration of decision makers' public statements in groupthink and non-groupthink crises. Evaluative assertion analysis was used to assess the direction and intensity of in-group and out-group attitudes in these situations.

Method

Archival records of public statements by leading decision makers in five American foreign policy crises provided the data for the current study. Statements were drawn from primary sources that included the *United States Department of State Bulletin*, the *Congressional Record, Collected Papers of the Presidents of the United States,* and *Vital Speeches*. The decision makers for whom materials were scored included: the Marshall Plan—President Harry S. Truman, Secretary of State George C. Marshall, Undersecretary of State Dean Acheson; the invasion of North Korea—President Harry S. Truman, Secretary of State Dean Acheson; the Bay of Pigs invasion—President John F. Kennedy, Secretary of State Dean Rusk; the Cuban Missile Crisis—President John F. Kennedy, Secretary of State Dean Rusk; The Vietnam War escalation decisions—President Lyndon B. Johnson, Secretary of State Dean Rusk. Public statements were taken from the following time periods: the Marshall Plan—May 1, 1947 – December 19, 1947; the invasion of North Korea—September 10, 1950 – November 20, 1950; the Bay of Pigs invasion—January 20, 1961 – April 20, 1961; the Cuban Missile Crisis—October 16, 1962 – October 29, 1962; the Vietnam War escalation decisions—December 1, 1964 – July 1, 1965. A total of 12 (paragraph-sized) passages were randomly selected for each major decision maker within each crisis period.

A topic restriction was enforced. Public statements were analyzed only if the material was primarily concerned with the foreign policy decision on which the Janis case studies focused. The primary reason for this restriction is uncertainty regarding the degree to which groupthink in one topic area generalizes to other areas. The following guidelines were used:

1. The Marshall Plan. All statements pertained to the economic state of Europe and the role that American aid could play in facilitating recovery.

2. The invasion of North Korea. All statements pertained to the conduct of the war and political and moral justifications for the war.

3. The Bay of Pigs invasion. All statements pertained to the dangers presented by the Castro government in Cuba and to the role of the United States in helping prevent expansion of Communist influence in Latin America.

4. The Cuban Missile Crisis. All statements pertained to the threat posed by the actions of the Soviet Union in Cuba.

5. The Vietnam War escalation decisions. All statements pertained to the danger of the Communist threat in Southeast Asia and the need for American aid to defend the non-Communist governments in the region.

The Bay of Pigs invasion posed unique problems for obtaining data. Since the invasion was a secret operation, decision makers did not describe or justify the invasion policy in their public statements. In this sense, the Bay of Pigs case differs from the other four crisis situations examined. Data analyses have been performed both with and without the data collected from the Bay of Pigs time period.

Integrative-Complexity Coding

All material was scored for integrative complexity on a 7-point scale (see Schroder, Driver, & Streufert, 1967, Appendix 2, for detailed discussion of the coding rules). The scale defines complexity in terms of both differentiation and integration. Differentiation refers to the number of characteristics or dimensions of a problem situation that are recognized and taken into account in decision making. For instance, a decision maker may process information relating to policy options in an undifferentiated fashion by placing options into only one of two categories: the "good, patriotic" policies and the "bad, defeatist" policies. A more differentiated approach would recognize that policy options can have multiple, often contradictory effects that cannot be classified on a single evaluative dimension of judgment—for example, effects on different political constituencies, various sectors of the economy, military strength, and the strategies of one's opponents. Integration refers to the development of complex connections among differentiated characteristics. (Differentiation is thus a prerequisite of integration.) The complexity of integration depends on whether the dimensions of judgment employed by the decision maker are perceived as operating in isolation (low integration), in hierarchical interaction (medium integration), or according to multiple, complex, and perhaps flexible patterns (high integration). Scores of 1 reflect low differentiation and low integration. Scores of 3 reflect medium or high differentiation and low integration. Scores of 5 reflect medium or high differentiation and medium integration. Scores of 7 reflect high differentiation and high integration. Scores of 2, 4, and 6 represent transition points between adjacent levels.

Initial scoring for integrative complexity was performed by the author and Peter Suedfeld, who were, however, both aware of the hypotheses being tested and the sources of the material (interrater agreement $r = .91$). The reliability of this scoring was checked by having material rescored by trained coders associated with a research group under Suedfeld at the University of British Columbia. These coders were blind to both the hypotheses being tested and the

Table 1
Attitude Objects Included Under General Categories of "Domestic and Foreign Opposition" and "Groups With Which Speaker Identifies"

Marshall Plan
Opposition
Communism, USSR, East European Communist states, Communist parties of Western Europe, totalitarian regimes, Communist victories, fall of Europe
Groups with which speaker identifies
American government, American people
Invasion of North Korea
Opposition
Communism, USSR, North Korea, Communist China, forces of Communism, Communist imperialism, new colonialism, Soviet leaders
Groups with which speaker identifies
American government, American people, United Nations Command in Korea, South Korea, UN soldiers, we, American history to the present, policy of increasing strength
Bay of Pigs invasion
Opposition
Communism, USSR, Castro's Cuba, alien ideology, Communist agents
Groups with which speaker identifies
American government, American people, Cubans opposed to Castro government, those concerned with maintaining peace, our long-range aspirations, Kennedy, our (OAS) delegates, anti-Castro guerrilla fighters, Cuban people, security of Latin America
Cuban Missile Crisis
Opposition
Communism, USSR, Castro's Cuba, new Soviet moves, Khrushchev's decision, extracontinental power
Groups with which speaker identifies
American government, American people, nations of this hemisphere, unanimity of Western Hemisphere, free nations
Vietnam War escalation decisions
Opposition
North Vietnamese, Viet Cong, Red China, USSR, aggressors, yielding to aggression, policy of withdrawal
Groups with which speaker identifies
American government, American people, South Vietnamese government, policy of standing firm

sources of the material. High interrater agreement ($r = .88$) was obtained between the complexity ratings assigned by persons aware and unaware of the hypotheses. Analysis of disagreements revealed no relationship to the groupthink hypothesis.

Evaluative Assertion Analysis

Evaluative assertion analysis was performed on the same materials. (Osgood et al., 1956, present a more detailed discussion of how to perform evaluative assertion analysis.) The analysis involved four basic stages and was performed by two coders, one of whom was the author; the other was an individual otherwise uninvolved in the study. The first stage of the evaluative assertion analysis was the identification and isolation of attitude objects. Two general classes of attitude objects were selected for analysis: *political groups with which the speaker identifies* and *domestic and foreign opponents*. Table 1 presents the variety of specific terms taken from public statements that were considered to fall into one of these two general categories. It was possible to obtain extremely high interrater agreement for identifying and classifying these terms. Disagreements were resolved by using the judgments of the coder unaware of the hypotheses.

The second stage involved translating all statements in which these attitude objects appeared into one of two common sentence forms:

Attitude Object 1 / Verbal connector / Common meaning term
Attitude Object 1 / Verbal connector / Attitude Object 2.

For example, the sentence "An aggressive North Korea threatens freedom-loving South Korea" would be translated to read:

North Korea / is / aggressive
North Korea / threatens / South Korea
South Korea / is / freedom loving.

Here, again, coding disagreements were rare and were resolved by using the judgment of the blind coder.

In the third stage, the verbal connectors and predicates were rated for intensity and direction on 7-

point scales ranging from −3 to +3. A verbal connector received a negative score to the degree that it dissociated the subject from the predicate ("never is" would receive a score of −3) and a positive score to the degree that it associated the subject and the predicate ("always is" would receive a score of +3). A predicate received a negative score to the degree that it represented a negatively evaluated attribute or quality (e.g., evil, aggressive) within "the language community of the speaker" (Osgood et al., 1956). The predicate received a positive score to the extent that the attribute or quality cited was positively evaluated within the language community of the speaker (e.g., freedom, peace). Reliability checks were obtained to ensure that verbal connectors and common meaning terms were being assigned ratings that reflected common linguistic standards. The correlations of coder ratings of verbal connectors and common meaning terms were .93 and .94, respectively.

The direction and intensity of evaluative sentiment directed toward each of the two categories of attitude objects were computed for each paragraph unit. Within each unit, references to each of the two classes of attitude objects were separated. The direction and intensity of evaluative sentiment directed to each class were computed by (a) multiplying the values assigned to the verbal connectors and predicates used to describe attitude objects falling within the class and (b) summing the products thus obtained. Two scores were obtained in this way for each paragraph unit from the evaluative assertion analysis: One specified the intensity and direction of evaluative feeling toward political groups with which the speaker identified, and the other specified evaluative feeling toward the speaker's opponents.

Results

Unweighted-means analyses of variance were used to test the groupthink hypotheses. The analyses of variance always involved three independent variables: type of crisis (groupthink vs. non-groupthink), decision makers within crises, and the random ordering of the 12 passages of material sampled from each decision maker's statements.

The analyses of variance were complicated by the fact that individual decision makers were unevenly represented across different crisis situations. In analysis of variance terms, decision makers were partly "crossed" with and partly "nested" within crises. For instance, Truman and Acheson both appeared in the 1947 (non-groupthink) and 1950 (groupthink) crises (that is, they were crossed with these two crises); similarly, Kennedy and Rusk both appeared in the 1961 (groupthink) and 1962 (non-groupthink) crises. Neither pair of decision makers appeared in crises involving the other pair. Marshall appeared only in the 1947 crisis, and Johnson appeared only in the 1964–1965 (groupthink) crisis. Finally, Rusk appeared most frequently: in the 1961, 1962, and 1964–1965 crises.

To resolve this complicated situation, it was assumed that different decision makers were involved in each crisis (i.e., decision makers have been treated as nested within crises). This assumption was conservative in the sense that it reduced the likelihood of detecting significant differences between groupthink and non-groupthink crises. The assumption sacrifices the opportunity to control statistically for the effects of personality variables. The groupthink effect must overcome personality effects in order to be detected.

A second, conservative assumption was also made. In all analyses of variance, the groupthink and non-groupthink crises have been treated as random effects. The crises were viewed as a sample from the population of groupthink and non-groupthink decisions. For this reason, quasi-F ratios have been constructed to test the significance of crisis effects (cf. Clark, 1973; Winer, 1971). These quasi-F ratios are smaller than the ordinary F ratios that would have been appropriate if crises were viewed as a fixed variable.

Table 2 presents the mean integrative complexity scores of each major decision maker within each crisis. Consistent with the groupthink hypothesis, a planned orthogonal contrast of complexity scores revealed that decision makers in groupthink crises were significantly less complex than their counterparts in non-groupthink crises, quasi-$F(1, 17) = 18.29$, $p < .01$. This effect remained significant when the data obtained from the Bay of Pigs period were deleted, planned contrast, quasi-$F(1, 16) = 15.77$, $p < .01$. No other effects were significant.

Decision makers' mean evaluations of political groups with which they identified are also presented in Table 2. As predicted, decision makers in groupthink crises evaluated political groups with which they identified more positively than did decision makers in non-groupthink crises, planned contrast, quasi-

Table 2
Mean Ratings on the Three Dependent Variables

Decision maker	Integrative complexity	Evaluation of group with which speaker identifies	Evaluation of opponents
Non-groupthink crises			
1947			
Truman	3.66	5.5	− 1.75
Acheson	4.5	.666	− 1.0
Marshall	5.9	0	− .166
1962			
Kennedy	4.33	2.92	0
Rusk	3.16	2.0	− 4.08
Groupthink crises			
1950			
Truman	1.0	14.0	−10.5
Acheson	1.83	13.33	− 1.92
1961			
Kennedy	2.16	12.33	− 6.25
Rusk	2.58	9.833	− 4.0
1964–1965			
Johnson	2.16	4.83	− 9.83
Rusk	2.5	3.33	− 2.92

$F(2, 17) = 17.37$, $p < .01$.[1] This effect remained significant when the data obtained from the Bay of Pigs period were deleted, planned contrast, quasi-$F(1, 16) = 8.03$, $p < .05$. No other effects were significant.

Table 2 also presents decision makers' mean evaluations of domestic and foreign opponents. Contrary to theoretical expectations, there was no significant difference between groupthink and non-groupthink crises in negative evaluations of domestic and foreign opponents, planned contrast, quasi-$F(1, 13) = 2.9$, $p < .15$. The effect remained nonsignificant when data from the Bay of Pigs period were deleted, planned contrast, quasi-$F(1, 11) = 2.34$, $p < .25$. In both analyses, the difference was in the direction predicted by the groupthink analysis. The other effects were also nonsignificant.

Low significant correlations existed between integrative complexity and the evaluative assertion analysis measures. More complex statements tended to include fewer positive evaluations of political groups with which the speaker identified, $r(130) = -.27$, $p < .001$, and fewer negative evaluations of domestic and foreign opponents, $r(130) = .33$, $p < .001$.

A major disadvantage of treating decision makers as nested within crises is that it prevents comparison of how individuals who appeared together in two crises (e.g., Truman and Acheson; Kennedy and Rusk) reacted in the different situations. For example, one decision maker may have been highly consistent in the integrative complexity and evaluative intensity of his public statements, whereas another may have been highly inconsistent. To examine this hypothesis, matched-pairs t tests were performed on the data obtained from Truman and Acheson in the crises in which they appeared (1947 and 1950) and on data obtained from Kennedy and Rusk in the crises in which they appeared (1961, 1962, and in the case of Rusk, 1964–1965). The results indicated that public statements of individual decision makers did not always change in the directions predicted by the groupthink analysis. While Truman's, Acheson's, and Kennedy's statements were less

[1] Degrees of freedom for the quasi-F statistics were calculated using Satterthwaite's formula (see Winer, 1971, pp. 375–378). With this formula, it is possible for numerator degrees of freedom of a planned contrast to be greater than one.

complex in groupthink than in non-groupthink situations, the complexity of Rusk's statements remained relatively constant through the 1961, 1962, and 1964–1965 crises. Truman's evaluations of opponents were significantly more negative during the 1950 crisis (relative to the 1947 crisis), but his evaluations of political groups with which he identified were not more positive. Acheson's evaluations of political groups with which he identified were, however, more positive in the 1950 crisis (relative to the 1947 crisis), but his evaluations of opponents were not significantly more negative. Kennedy's public statements changed most in conformity with the groupthink model. Between the Bay of Pigs and the Cuban Missile Crisis, Kennedy showed significant or marginally significant changes in the predicted direction on integrative complexity, evaluations of opponents, and evaluations of political groups with which he identified. Rusk's public statements were least in conformity with the groupthink model. He showed no significant shifts on any of the three dependent variables between the 1961 and 1962 crises or the 1962 and 1964–1965 crises. In sum, evidence exists for individual differences in decision makers' reactions to groupthink and non-groupthink situations.

An important unresolved issue is how to determine whether groupthink is occurring in given situations. The present research suggests that relatively objective content analysis procedures applied to the public statements of decision makers may be useful for this purpose. As a first step toward developing clear criteria for distinguishing groupthink and non-groupthink decisions, discriminant analysis was used to evaluate the usefulness of the three dependent variables employed in this study in differentiating groupthink and non-groupthink decisions (cf. Cooley & Lohnes, 1971).

The discriminant function obtained was significant at the .001 level, $\chi^2(3) = 72.7$, and accounted for 45% of the total variation between groups. The standardized discriminant function coefficients for maximally distinguishing the two groups were (a) integrative complexity, .89; (b) evaluations of political groups with which speaker identifies, −.21; and (c) evaluations of domestic and foreign opponents, .08. This pattern indicates that public statements in groupthink crises can be best distinguished from public statements in non-groupthink crises on the basis of integrative complexity and, to a lesser extent, evaluations of political groups with which the decision makers identify. The "evaluations of opponents" variable appears to have no independent discriminatory power. The discriminant function correctly predicted the groupthink or non-groupthink origins of 78% of the public statements.[2]

Discussion

The results for two of the three dependent variables strongly supported the predictions derived from Janis's groupthink analysis. Public statements of decision makers in groupthink crises were characterized by significantly lower levels of integrative complexity than the public statements of decision makers in non-groupthink crises. Decision makers in groupthink crises evaluated political groups with which they identified more positively than did decision makers in non-groupthink crises. However, contrary to expectation, groupthink and non-groupthink decision makers did not significantly differ in the intensity of their negative evaluations of domestic and international opponents.

With the exception of the last dependent variable, the current findings converge impressively with the conclusions of Janis's intensive case studies. The convergence is impressive primarily because the findings of this study and those of Janis (1972) are based upon markedly different types of data (public statements versus the retrospective accounts of observers and participants) that have been processed in very different ways (quantitative content analysis vs. intuitive reconstruction of historical episodes). The present study underscores how multiple methods of investigation—ranging from case studies to content analysis studies to laboratory experiments—

[2] A separate discriminant analysis was performed without the Bay of Pigs data. The results did not differ substantially from those reported above.

an be brought together to test the validity of ne groupthink construct.

Nonetheless, it is appropriate to introduce cautious note. The evidence reported here loes not rule out all alternative explanations. The skeptic can point out that the groupthink and non-groupthink crises differed in ways that render interpretation of the results somewhat ambiguous. For instance, as noted earlier, the groupthink decisions all led to unsuccessful military interventions or escalations, whereas the non-groupthink decisions did not lead to significant military action. Differences in public statements prior to these decisions may reflect differences in propaganda strategies used to convince the American public to accept military or nonmilitary solutions to conflicts. Political leaders may justify militaristic decisions by emphasizing in simple terms their total opposition to the enemy and their total dedication to their own nation's values. The disadvantage of this interpretation is that, unlike the groupthink model, it cannot explain most of the available historical evidence on the behavior of policy makers in the relevant crises (see De Rivera, 1968; Graff, 1970; Janis, 1972; Neustadt, 1964). The propaganda explanation accounts only for the results of the content analyses of public statements.

A decisive test of the groupthink and propaganda explanations will probably elude us for a long time. At least two lines of further research can, however, help to clarify the applicability of the competing positions in particular historical settings. One (still not feasible) approach is to compare the private statements of decision makers in the hypothesized groupthink and non-groupthink crises. If groupthink is operating, manifestations of the syndrome should appear in private and public statements. If the propaganda position is correct, there is no reason to expect the private statements of groupthink and non-groupthink decision makers to differ, even though significant differences would emerge in their public statements. A second approach is to draw upon case studies in the historical and political science literatures to identify a much larger sample of probable groupthink and non-groupthink decisions, some of which have led to war and some of which have not. If the propaganda position is correct, symptoms of groupthink should appear in public statements prior to the decision to go to war, independently of how the decision was made. If the groupthink position is correct, manifestations of groupthink should appear only in the public statements of the groupthink decision makers.

Regardless of whether the war propaganda or groupthink interpretation of this study is correct, the results reported here have intriguing practical implications. The propaganda interpretation suggests that it should be possible to predict in advance the occurrence of military interventions or escalations using content analysis indicators of integrative complexity and attitude polarization. The groupthink interpretation suggests that it should be possible to use the same content analysis indicators to monitor the quality of decision making of governmental leaders. Such monitoring activity may have the beneficial effect of sensitizing policy makers to the manner in which they make decisions and may, in the most optimistic of scenarios, lead to improvements in the quality of their decision making.

Finally, a few comments are in order concerning the lack of significant differences in negative evaluations of opponents in the groupthink and non-groupthink crises. This result is puzzling for both the groupthink and war propaganda explanations (indeed, perhaps especially puzzling for the propaganda explanation). There may, however, be a methodological reason for the result. Content analyses of public statements may be insensitive to differences in evaluations of opponents in groupthink and non-groupthink crises. In this regard, the fact that the public statements sampled contained 33% fewer negative than positive evaluative assertions is suggestive. This finding is not surprising in view of the considerable social psychological evidence indicating that people are more likely to offer positive than negative evaluations and that persons giving positive evaluations are themselves regarded as more attractive than those giving negative evaluations (see Folkes & Sears, 1977). It is probably effective impression management for politicians to "accentuate the positive" in their

public statements. Diplomatic norms may add further pressure in this direction. An advocate of the groupthink position could simply argue that groupthink decision makers were reluctant to express fully their negative feelings toward opponents.

The argument above points again to the need for research that analyzes verbatim records of actual group deliberations. A more conclusive test of the groupthink analysis awaits the declassification of such documents.

References

Abelson, R. P. The structure of belief systems. In R. C. Schank & K. M. Colby (Eds.), *Computer models of thought and language*. San Francisco: Freeman, 1973.

Allison, G. T. *Essence of decision: Explaining the Cuban Missile Crisis*. Boston: Little, Brown, 1971.

Axelrod, R. (Ed.). *The structure of decision*. Princeton, N.J.: Princeton University Press, 1976.

Bem, D. Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6). New York: Academic Press, 1972.

Clark, H. H. The language as fixed effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 335–339.

Cooley, W. W., & Lohnes, P. R. *Multivariate data analysis*. New York: Wiley, 1971.

De Rivera, J. *The psychological dimension of foreign policy*. Columbus, Ohio: Merrill, 1968.

Fischhoff, B., & Beyth-Marom, R. Failure has many fathers. *Policy Sciences*, 1976, *7*, 388–393.

Flowers, M. A laboratory test of some implications of Janis's groupthink hypotheses. *Journal of Personality and Social Psychology*, 1977, *35*, 888–896.

Folkes, V. S., & Sears, D. O. Does everybody like a liker? *Journal of Experimental Social Psychology*, 1977, *13*, 505–519.

Glad, B. Psychobiography. In J. Knutson (Ed.), *Handbook of political psychology*. San Francisco: Jossey-Bass, 1973.

Graber, D. *Verbal behavior and politics*. Urbana: University of Illinois Press, 1976.

Graff, H. *The Tuesday cabinet*. Englewood Cliffs, N.J.: Prentice-Hall, 1970.

Harvey, O. J., Hunt, D. W., & Schroder, H. J. *Conceptual systems and personality organization*. New York: Wiley, 1961.

Hermann, C. F. *International crises: Insights from behavioral research*. New York: Free Press, 1972.

Holsti, O. R. Cognitive dynamics and images of the enemy. In D. J. Finlay, O. R. Holsti, & R. R. Fagen (Eds.), *Enemies in politics*. Chicago: Rand McNally, 1967.

Holsti, O. R. Foreign policy formation viewed cognitively. In R. Axelrod (Ed.), *The structure of decision*. Princeton, N.J.: Princeton University Press, 1976.

Janis, I. *Victims of groupthink*. Boston: Houghton Mifflin, 1972.

Janis, I. L., & Mann, L. *Decision making*. New York: Free Press, 1977.

Jervis, R. *Perception and misperception in international politics*. Princeton, N.J.: Princeton University Press, 1976.

Neustadt, R. E. *Presidential power: The politics of leadership*. New York: Signet, 1964.

Osgood, C. E. The representational model and relevant research methods. In I. de S. Pool (Ed.), *Trends in content analysis*. Urbana: University of Illinois Press, 1959.

Osgood, C. E., Saporta, S., & Nunnally, J. C. Evaluative assertion analysis. *Litera*, 1956, *3*, 47–102.

Schroder, H. M. Conceptual complexity. In H. Schroder & P. Suedfeld (Eds.), *Personality theory and information processing*. New York: Ronald Press, 1971.

Schroder, H. M., Driver, M. J., & Streufert, S. *Human information processing*. New York: Holt, Rinehart & Winston, 1967.

Suedfeld, P. Measuring integrative complexity in archival materials. In H. Mandl & G. L. Huber (Eds.), *Kognitive Komplexität*. Göttingen: Hogrefe Verlag, 1978.

Suedfeld, P., & Rank, A. D. Revolutionary leaders: Long-term success as a function of changes in conceptual complexity. *Journal of Personality and Social Psychology*, 1976, *34*, 169–178.

Suedfeld, P., & Tetlock, P. E. Integrative complexity of communications in international crises. *Journal of Conflict Resolution*, 1977, *21*, 169–184.

Suedfeld, P., Tetlock, P. E., & Ramirez, C. War, peace and integrative complexity: U.N. speeches on the Middle East problem, 1947–1976. *Journal of Conflict Resolution*, 1977, *21*, 427–441.

Tedeschi, J. T., Schlenker, B. R., & Bonoma, T. V. Cognitive dissonance: Private ratiocination or public spectacle? *American Psychologist*, 1971, *26*, 685–695.

Westley, B. H., Higbie, C. E., Burke, T., Lippert, P. J., Maurer, L., & Stone, V. A. The news magazines and the 1960 conventions. *Journalism Quarterly*, 1963, *42*, 591–595.

Winer, B. J. *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill, 1971.

Winter, D. G. *The power motive*. New York: Free Press, 1973.

Received July 24, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 8, 1325-1341

The Birth Order Puzzle

R. B. Zajonc, Hazel Markus, and Gregory B. Markus
University of Michigan (Ann Arbor)

Studies relating intellectual performance to birth order report conflicting results, some finding intellectual scores to increase, others to decrease with birth order. In contrast, the relationship between intellectual performance and family size is stable and consistently replicable. Why do these two highly related variables generate such divergent results? This birth order puzzle is resolved by means of the confluence model that quantifies the influences upon intellectual growth arising within the family context. At the time of a new birth, two opposing influences act upon intellectual growth of the elder sibling: (a) his or her intellectual environment is "diluted" and (b) he or she loses the "last-born's handicap" and begins serving as an intellectual resource to the younger sibling. Since these opposite effects are not equal in magnitude, the differences in intellectual performance among birth ranks are shown to be age dependent. While elder children may surpass their younger siblings in intellectual performance at some ages, they may be overtaken by them at others. Thus when age is taken into consideration, the birth order literature loses its chaotic character and an orderly pattern of results emerges.

> Reuben, you are my first born, my might, and the fruits of my strength, pre-eminent in pride and pre-eminent in power. (Genesis 49:3)

In most societies some individuals come to control great wealth, to wield enormous power, and to acquire coveted honors, titles, and privileges just by virtue of their position in the family. Faith in superior endowment of the first offspring has fostered laws of succession, inheritance, and intestacy favoring the firstborn child—especially the male child—that still prevail in many countries. In fact, it was not until the turn of the century that this faith and the corresponding laws of primogeniture were challenged. At this time, geneticists, pathologists, and psychologists began to explore the question from a variety of perspectives, and indeed, the first studies appeared to provide empirical substance for the belief in the superiority of the eldest (cf. Ellis, 1904; Galton, 1874; Gini, 1915). For example, Cattell and Brimhall (1921) published their observations on the birth rank of scientists and concluded that the firstborn is to be found in greater frequency among scientists than the later born, and that this overrepresentation occurs for all family sizes.

With continued attention to this question, however, contradictory results soon began to appear. Thurstone and Jenkins (1929) examined a large number of children and concluded quite explicitly:

> On the whole the *later* born siblings tend to be on the average brighter than the first-born. Not only does this seem to be the case in the comparison of the first-born with the subsequent children, but the rise in intelligence with the order of birth seems to continue as far as the eighth-born child. (p. 649)

In a subsequent study, Steckel (1930) substantiated Thurstone's results.

Other studies followed. Some reported increments with birth order, some decrements, and several failed to find any relationship whatever. Among studies that found decreasing intelligence or scholastic scores with birth

This research was supported by Grant BNS77-13973 from the National Science Foundation. We wish to thank John Bargh, Michael Berbaum, and Beth Shinn for their help.

Requests for reprints should be sent to P. B. Zajonc, Institute for Social Research, University of Michigan, Ann Arbor, Michigan 48106.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3708-1325$00.75

order were those of Altus (1965), Bayley (1965), Breland (1974), Belmont and Marolla (1973), Lunneborg (1968, 1971), and Schachter (1963). Increases in intelligence scores with birth order were reported by Arthur (1926), Commins (1927), Hill (1936), Koch (1954), and Willis (1924). In a study by Hsiao (1931), some samples showed a positive relationship with birth order and others a negative relationship. And Bayer (1966) and McCall and Johnson (1972) failed to find any relationship whatever between birth order and intelligence. The latter authors suspected, in fact, that the correlations of IQ with birth order approach zero "in those studies where more careful attention is given to sample design and to subsequent controls" (p. 208). The research literature on birth order appeared so confused that one serious reviewer was prompted to declare birth order unworthy of further research efforts (Schooler, 1972).

In contrast to the inconsistencies among the birth order data are the results on the intellectual consequences of family size (Terhune, 1976). Figure 1 shows the results of surveys of the intellectual levels of six large populations. In spite of differences in age, sex, nationality, and type of test given, they all show a strikingly similar pattern of decline in scores with family size.

This, then, is the birth order puzzle. Why is the effect of family size such a consistent one, showing itself over and over again in the literature, while at the same time the effect of birth order—a closely related factor—presents such a chaotic picture? The birth order problem is especially troublesome, for none of the variables examined in the literature can organize these results into an orderly set of generalizations. For example, the suspicion that birth order effects are found primarily on verbal proficiency tests is seriously undermined by the strong effects found also with the Raven Progressive Matrices Test (Belmont & Marolla, 1973). The possibility that socioeconomic factors may mediate these effects is dispelled by several studies (e.g., Institut National d'Etudes Démographiques [INED], 1973; Claudy, Note 1) showing parallel effects for a variety of socioeconomic status categories. Cohort effects, too, are soon ruled out as a possibility because Galton's cohort of the 19th century showed effects no different from those of the much later cohort examined by Cattell and Brimhall (1921) nor from those of the most recent cohort of high school seniors examined by Claudy (Note 1).

A recent analysis of intellectual development may supply a solution to the birth order puzzle. A model termed *the confluence model* (Markus & Zajonc, 1977; Zajonc, 1976; Zajonc & Markus, 1975) was developed to explain a large body of intelligence data published in 1973 by Belmont and Marolla. These data on birth order and family size exhibited five important features: (a) intelligence scores declined with family size; (b) within each

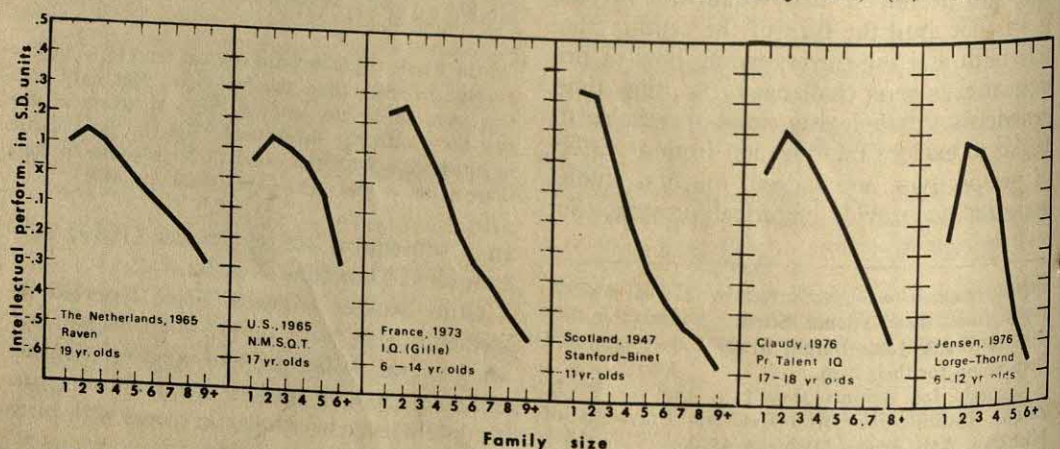


Figure 1. Family size (number of children) and intellectual performance in six large samples. (S.D. = standard deviation; N.M.S.Q.T. = National Merit Scholarship Qualifying Test.)

family size they declined with birth order; (c) if the last child was ignored, the decline with birth order seemed to be decelerated; (d) the decelerating birth order trend was not followed by the last child, who showed a discontinuous drop in intellectual performance; and (e) the only child, too, showed a discontinuity in that if the family factor were systematically negative in influencing IQ, the only child should have had the highest average of all, which was not the case, as can be clearly seen in Figure 1. The confluence model was constructed to reflect these features of the Belmont–Marolla data, and it has been refined since its original publication to accommodate new data (see Appendix).

The Confluence Model

The basic assumption of the confluence model, both in its earlier and its reparametrized version presented here, is that family influences that contribute significantly to intellectual growth can be divided into two sources. Specifically, the rate of intellectual growth is hypothesized to be a function of the intellectual environment within the family, α, and a factor, λ, associated with the special circumstances of last children.

The intellectual environment is defined to be a function of the absolute intellectual levels of all family members, including the individual whose intellectual development is being analyzed.[1] These intellectual levels are not to be confused with IQ, which is a measure relative to age. Nor are they to be confused with mental age, whose population means are commonly set equal to chronological ages. Rather, they are absolute quantities that reflect each individual's mental maturity. One could think here, for example, of the individual's vocabulary, which is accumulated not in constant increments but according to a sigmoid function.

As the family members mature, the intellectual environment changes, and the most pronounced changes are likely to occur when new children are born into the family or when grown members leave. Since infants are intellectually immature, their births depress the average intellectual level in the family. Moreover, the smaller the family, the greater is the impact of a new birth (or a departure of an adult) on the resultant intellectual environment. Therefore, the major family configuration variables that affect α are family size and length of intervals between the births of successive siblings.

The term λ reflects the two discontinuities found in the Belmont–Marolla data for only and last children. These discontinuities exist, it is hypothesized, because of a particular disadvantage suffered by all last children, including only children. The intellectual responsibilities of older and younger children differ within the family. Specifically, only and last children are not teachers in the broad sense of the word, and it is assumed that the relatively low performance scores often reported for them in the literature are associated with this lack of teaching opportunity.[2]

Since last-born children do not normally serve as intellectual resources to their siblings, a condition assumed to enhance intellectual growth (Zajonc & Markus, 1975), and since, except for twins, each child is at some age the last child in the family, the rate of intellectual growth during such periods is affected accordingly. Specifically, λ_t will be equal to zero throughout the period when the child is last born and will take on a positive value once a younger sibling is born. It is assumed, furthermore, that the beneficial effects of the teaching function vary with age too, and that they vary not only with the age of the individual who provides tutoring but also with the age of the individual who receives it. One cannot

---

[1] We shall use the term *family* to include also individuals who are not related to each other (such as permanent guests, domestic help, etc.) but who live in the home and can be considered to be members of the immediate intellectual environment.

[2] While there are no studies on the teaching function in the family, some recent research has shown the substantial benefits that are derived by tutors in school cross-peer teaching programs (Devin-Sheehan, Feldman, & Allen, 1976; Duff, 1974; Morgan & Toy, 1970; Richer, 1973; Wu, 1973). In most cases the benefits accruing to the tutor exceed those accruing to the tutee. In one study, for example, there was a 9-month gain on the Wide Range Achievement Test for tutors and only a 3-to-5-month gain for their charges (Morgan & Toy, 1970).

do much tutoring at an age less than 2 years, and little tutoring can be administered to an infant. The term λ_t reflects these considerations, in the reparametrized version of the model, representing them as the product of the two increments—that pertaining to the individual whose mental maturity is being calculated and that pertaining to the next sibling. All members of the family, except the last born, serve as intellectual resources to one another. The benefits that derive from this function are therefore divided among the parents and all the children but the last one. Thus, λ_t has $n - 1$ in its denominator.

Birth order has no independent effect on intellectual development within the confluence model. Instead, effects previously thought to be associated with birth order are posited to be mediated substantially by family size and the spacing of births. Short intervals force later-born children to spend a larger portion of their period of growth in a family environment "diluted" by the presence of young children. As intervals between successive children increase, new children enter relatively more advanced environments that enhance their growth. Hence the birth order effects may be mitigated or even reversed by variations in birth intervals (Markus & Zajonc, 1977; Zajonc, 1976; Zajonc & Markus, 1975). The reparametrized confluence model provides a fairly exact fit (obtained by means of iterative procedures) for all of the large data sets thus far analyzed, responding well to the wide variety of patterns of empirical results, some of which show a negative effect of birth order on intelligence, others a U-function or no effect at all.

Dynamics of the Intellectual Environment and Individual Patterns of Growth

According to the confluence model, the intellectual environment changes continually as family members mature and as the family expands (or contracts). Hence, entries at different points into this dynamic environment must result in radically different patterns of growth. Some children may grow up in a family environment marked by abrupt shifts that are occasioned by many additions or departures, whereas their siblings might enter a different time segment of the same environment, where changes are much more gradual. As a consequence, comparisons of intellectual levels at one point in time in the family history or at a given age may yield different values from those obtained from comparisons made at another time. The age at which birth order effects are evaluated is thus of pivotal significance, and the major concern of the remainder of this article is with these variations in birth order effects that occur as a function of the age of testing.

The dependence of birth order effects on age is one of the unanticipated discoveries of the confluence model. This dependence can be readily understood if we consider the growth patterns of only children and children who have just one sibling. Note that, other things constant, only children should develop undisturbed until maturity. There are no events within the family during this period that would cause sharp discontinuities in their rate of growth. Yearly increments in growth are only those that are associated with the maturation of all the family members. For the same reason, the second of two children will also develop undisturbed until maturity (provided that his or her older sibling does not leave home and there are no other major changes in the family situation). In principle, the above statement holds for *all* last children, and it holds *only* for them. The growth rates for these children will show no discontinuities. In contrast, the older of two siblings should manifest a discontinuity in his or her growth pattern, and such discontinuity should coincide with the acquisition of the younger sibling. From his or her own birth until the birth of a younger brother or sister, the first of two children grows at the same rate as the only child. The family conditions of their growth are identical until that age. But when the second child is born, two dramatic changes take place: (a) The average of the intellectual level within the family is decreased by the addition of the immature member and (b) the firstborn acquires a teaching function (i.e., the value of his or her λ_t changes from zero to a positive number).

Dunn (Note 2) reports findings consistent

with these hypothesized changes for young children (18 to 43 months) shortly after they acquire a younger sibling. She observed several specific examples of teaching and caregiving by the older sibling but at the same time found that "when the young siblings play together the firstborn tends to play relatively simple games; he approaches the level of play more appropriate for the younger child" (p. 19).

The birth of a second child can have one of several effects on the firstborn, depending on the assumptions made: (a) If the decrement occasioned by the dilution of the environment due to the new birth exactly equals the gain derived from the teaching function, the growth rate of the firstborn will not be affected at the time of the second child's birth; (b) if the teaching function has an effect more pronounced than the downward change in the value of α, the growth rate of the firstborn will be accelerated; (c) if the initial dilution is more than can be compensated for by the change in λ, however, we may observe an initial depression in the firstborn's growth rate relative to that of the only child.

The first two outcomes would require that the benefits accumulated from teaching a newborn sibling be immediate and large. This possibility seems fairly implausible. The third possibility, which presumes that the benefits from teaching accrue more slowly and gradually, seems more likely. For the range of values of the parameters w_1, w_2, and k (see Appendix), which fit aggregate data such as those of Belmont and Marolla (1973), it is the third effect that is most consistent with the model.

If this third possibility were to obtain, then according to the confluence model the decrement of the firstborn vis-à-vis the only child must be age bound. As the children develop and their intellectual skills increase, the environment within the family gains in quality. The effects of the teaching function, too, cumulate for the firstborn but not for the only child—a condition that could allow the firstborn eventually to surpass the only child, for whom λ always remains equal to zero. This prediction is illustrated in Figure 2, in which

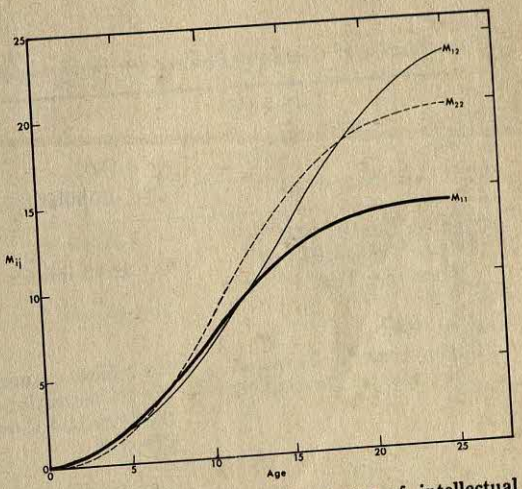


Figure 2. Theoretical growth curves of intellectual levels of only children (M_{11}) and of firstborns (M_{12}) and last borns (M_{22}) in families of two children.

the intellectual levels of the only child and of both siblings in a two-child family are plotted as a function of chronological age.[3] It can be seen in these growth curves that the only child and the first of two grow at the same rate until the latter acquires a sibling at age 5 years. Subsequently, the rate of growth of the firstborn drops below that of the only child. Eventually, the firstborn will surpass the only child, however, and will retain this superiority until maturity, because his or her teaching function continues to cumulate and eventually compensates for the decrement in the α component originally brought about by the acquisition of a younger sibling.

The comparison of the growth rates of the firstborn and second born shown in Figure 2 is also of interest. At birth the second born comes into an intellectual environment that is relatively more impoverished than the one entered by the firstborn. However, the rate of change in these environments is quite different, and the second born can soon exceed the firstborn's rate of growth. Consider, for example, a family with two children whose births are separated by 3 years. At age 8, the intellectual environment of the oldest will reflect the intellectual level of him-/herself (an 8-

[3] The differences between the curves have been exaggerated for illustrative purposes.

Table 1
Test Results of Children Ranked by Birth Order

Author and test	Population	Age	Results		
			Only	First-born	Second born
Bryant & Davies, 1974					
Denver Developmental Screening Test	668 Cardiff infants	16–375 days	Firstborn shows a faster gross and fine motor development.		
Bayley, 1965					
Bayley Scales of Motor and Mental Development	1,260 infants examined in the 10 institutions of the Perinatal Research Project[a]	1–15 months	Firstborn significantly superior on mental score in Months 5, 7, 12, and 13, on motor score in Months 6, 7, 12, 14, and 15. Second born shows no significant superiority over firstborn in any of the 15 months.		
Abe, Tsuji, & Suzuki, 1963					
Draw-A-Person-Test	302 Osaka children	3 years		2.0[b]	2.5[b]
Koch, 1954					
SRA Primary Abilities Test	360 Chicago children	6 years		108.2	111.1
Arthur, 1926					
Kuhlmann-Binet	92 pairs of siblings from Minnesota with Finnish, Russian, and South European names	5–7 years		93.0	99.1
	190 pairs of siblings from the same populations, all tested at the same age			90.1	96.9
Hsiao, 1931					
Stanford-Binet	133 Berkeley elementary school sibling pairs	7.5 years		104.4[c]	105.2[c]
	258 unrelated subjects	8 years		103.5[c]	108.1[c]
National Intelligence Test	118 sibling pairs	10 years		103.0	104.9
Jensen, 1974[d]					
Lorge-Thorndike	1,959 unrelated subjects	6–14 years	104.7	109.9	109.1
Willis, 1924					
Stanford-Binet	219 sibling pairs	Elementary school		97.1	99.7
Schoonover 1959					
Stanford-Binet	59 sibling pairs	Elementary school		24.8[e]	25.7[e]
INED, 1973					
INED Test	33,339 elementary school children	6–14 years	103.6	107.0	106.7
Tabah & Sutter, 1954					
Gille IQ Test	2,488 elementary school children	6–13 years		100.2	100.5
Commins, 1927					
McCall Multi-Mental	142 elementary school sibling pairs	Grades 3–8	In 70% of the pairs, firstborn has lower scores		

Table 1 (*continued*)

Author and test	Population	Age	Results		
			Only	First-born	Second born
Chittenden, Foan, & Zweil, 1968 Iowa Basic Skills	62 sibling pairs in Massachusetts	Grades 4–5	In 58% of the pairs, firstborn has higher scores; in 35%, lower scores.[f]		
	67 sibling pairs in Massachusetts	Grades 7–8	In 64% of the pairs, firstborn has higher scores; in 31%, lower scores.[e]		
Richardson, 1936 Stanford-Binet	101 sibling pairs tested at the same age	11.5 years		95.5	94.5
	101 sibling pairs tested at different ages	Firstborn, 12 years Second born 8.5 years		94.3	102.1
SCRE, 1949 Scottish Council Verbal Test	23,403 Scottish children	11 years	42.0	41.9	41.5
Eysenck & Cookson, 1970 Verbal Reasoning	4,000 primary school children in England	11 years	96.4	96.1	96.1
Cicirelli, 1967 California Short Form Test of Mental Maturity	609 suburban Detroit children	Grade 6			
	Males			108.2	114.4
	Females			117.9	110.8
Hill, 1936 Otis, Form A	New Jersey children, 66 sibling pairs	12 years		83.6	87.6
	72 unrelated matched pairs	12 years		84.2	87.5
Oberlander et al., 1970 SRA High School Placement	153 students from a northern suburb of Chicago	Grade 8		113.3	114.9
Svensson, 1972 Opposites, Number Series, Spatial	10,025 Swedish school children (national sample)	13 years	21.3	21.5	21.3
Hsiao, 1931 Terman Group Test	366 sibling pairs	13 years		109.1	106.3
	569 unrelated subjects	13 years		108.8	108.1
Davis, Cahan, & Bashi, 1977 Quantitative Skills Test Z scores	82,689 Israeli children of Western origin	14 years	.48	.58	.47
	109,304 Israeli children of Oriental origin	14 years	−.09	.16	−.01

Table 1
Test Results of Children Ranked by Birth Order

Author and test	Population	Age	Results		
			Only	First-born	Second born
Bryant & Davies, 1974					
Denver Developmental Screening Test	668 Cardiff infants	16–375 days	Firstborn shows a faster gross and fine motor development.		
Bayley, 1965					
Bayley Scales of Motor and Mental Development	1,260 infants examined in the 10 institutions of the Perinatal Research Project[a]	1–15 months	Firstborn significantly superior on mental score in Months 5, 7, 12, and 13, on motor score in Months 6, 7, 12, 14, and 15. Second born shows no significant superiority over firstborn in any of the 15 months.		
Abe, Tsuji, & Suzuki, 1963					
Draw-A-Person-Test	302 Osaka children	3 years		2.0[b]	2.5[b]
Koch, 1954					
SRA Primary Abilities Test	360 Chicago children	6 years		108.2	111.1
Arthur, 1926					
Kuhlmann-Binet	92 pairs of siblings from Minnesota with Finnish, Russian, and South European names	5–7 years		93.0	99.1
	190 pairs of siblings from the same populations, all tested at the same age			90.1	96.9
Hsiao, 1931					
Stanford-Binet	133 Berkeley elementary school sibling pairs	7.5 years		104.4[c]	105.2[c]
	258 unrelated subjects	8 years		103.5[c]	108.1[c]
National Intelligence Test	118 sibling pairs	10 years		103.0	104.9
Jensen, 1974[d]					
Lorge-Thorndike	1,959 unrelated subjects	6–14 years	104.7	109.9	109.1
Willis, 1924					
Stanford-Binet	219 sibling pairs	Elementary school		97.1	99.7
Schoonover 1959					
Stanford-Binet	59 sibling pairs	Elementary school		24.8[e]	25.7[e]
INED, 1973					
INED Test	33,339 elementary school children	6–14 years	103.6	107.0	106.7
Tabah & Sutter, 1954					
Gille IQ Test	2,488 elementary school children	6–13 years		100.2	100.5
Commins, 1927					
McCall Multi-Mental	142 elementary school sibling pairs	Grades 3–8	In 70% of the pairs, firstborn has lower scores		

Table 1 (*continued*)

Author and test	Population	Age	Results		
			Only	First-born	Second born
Chittenden, Foan, & Zweil, 1968 Iowa Basic Skills	62 sibling pairs in Massachusetts	Grades 4–5	In 58% of the pairs, firstborn has higher scores; in 35%, lower scores.[f]		
	67 sibling pairs in Massachusetts	Grades 7–8	In 64% of the pairs, firstborn has higher scores; in 31%, lower scores.[e]		
Richardson, 1936 Stanford-Binet	101 sibling pairs tested at the same age	11.5 years		95.5	94.5
	101 sibling pairs tested at different ages	Firstborn, 12 years Second born 8.5 years		94.3	102.1
SCRE, 1949 Scottish Council Verbal Test	23,403 Scottish children	11 years	42.0	41.9	41.5
Eysenck & Cookson, 1970 Verbal Reasoning	4,000 primary school children in England	11 years	96.4	96.1	96.1
Cicirelli, 1967 California Short Form Test of Mental Maturity	609 suburban Detroit children	Grade 6			
	Males			108.2	114.4
	Females			117.9	110.8
Hill, 1936 Otis, Form A	New Jersey children, 66 sibling pairs	12 years		83.6	87.6
	72 unrelated matched pairs	12 years		84.2	87.5
Oberlander et al., 1970 SRA High School Placement	153 students from a northern suburb of Chicago	Grade 8		113.3	114.9
Svensson, 1972 Opposites, Number Series, Spatial	10,025 Swedish school children (national sample)	13 years	21.3	21.5	21.3
Hsiao, 1931 Terman Group Test	366 sibling pairs	13 years		109.1	106.3
	569 unrelated subjects	13 years		108.8	108.1
Davis, Cahan, & Bashi, 1977 Quantitative Skills Test *Z* scores	82,689 Israeli children of Western origin	14 years	.48	.58	.47
	109,304 Israeli children of Oriental origin	14 years	−.09	.16	−.01

Table 1 (*continued*)

Author and test	Population	Age	Results		
			Only	First-born	Second born
Lunneborg, 1968					
Washington Precollege Battery	High school seniors 2,878 males	17–18 years	Firstborns scored higher than later borns or only children on 18 of 18 tests		
	2,523 females		Firstborns scored higher than later borns or only children on 14 of 18 tests		
Lunneborg, 1971					
Washington Precollege Battery	896 siblings of two-child families	16 years	Firstborns score higher on 15 of the 16 ability measures		
Breland, 1974					
National Merit Scholarship Qualification Test	794,589 high school students from a national competition	17 years	103.7	106.2	104.4
Burton, 1968					
Project TALENT IQ Measure	43,352 high school seniors	18 years			
	males			50.1	48.7
	females			49.7	48.9
Claudy, Note 1					
Project TALENT IQ Measure in *z* scores	81,175 12th graders (national sample)	17–18 years	.08	.30	.07
Altus, 1965					
SAT	1,878 University of California students	17–18 years		538.9	518.8
Schachter, 1963					
GPA	628 Minneapolis high school students	14–18 years	2.31	2.35	2.23
Rosenberg & Sutton-Smith, 1969					
ACE (total)	Students at Bowling Green State University	19 years			
	355 males			52.0	49.5
	658 females			53.2	52.4
Belmont & Marolla, 1973					
Raven	386,114 Dutch males born 1944–1946	19 years	2.688[g]	2.575[g]	2.678[g]

Note. SRA = Science Research Associates. INED = Institut National d'Études Démographiques. SCRE = Scottish Council for Research on Education. SAT = Scholastic Aptitude Test. GPA = Grade Point Average. ACE = American College Entrance Examination.

[a] Boston, Providence, New Haven, Buffalo, New York City, Philadelphia, Memphis, New Orleans, San Francisco, and Portland, Ore.

[b] High scores indicate superior performance.

[c] Corrected for age.

[d] Arthur R. Jensen kindly made his data from the Berkeley Unified School District study available for the purposes of this analysis.

[e] Months over average norms.

[f] Composite score.

[g] 1 = high score; 6 = low score.

year-old), the parents, and the sibling, aged 5. When the last born is 8 years old, however, the environment will consist of him-/herself, the parents, and a sibling aged 11 years. Thus, at a given age the second born's environment may be superior to the environment that sur-

rounded the older sibling at the same age.[4] The accelerated intellectual growth of the second born is temporary, however. Although at a given age the second born may enjoy a superior environment, he or she does not accrue the benefits of the teaching function. The firstborn will benefit by virtue of being a teacher, and the cumulation of this benefit may eventually lead the firstborn to surpass his or her sibling.

The differences in the rates of growth of the first and second children apply equally to all next-to-last and last children. In summary, given the values of the parameters w_1, w_2, and k such as we found for aggregate intelligence data, the confluence model leads to the following expectations:

1. The rate of intellectual growth of the only child and of the first of two is the same until the age when the firstborn acquires a younger sibling.

2. After the birth of the younger sibling, there is a decline in the intellectual environment, and relative to the only child, the rate of growth of the firstborn of two is temporarily depressed.

3. A few years after the birth of his or her younger sibling, the firstborn acquires a teaching function, and thus his or her rate of growth accelerates and eventually surpasses the rate of growth of the only child. At maturity, the first of two will have a higher intellectual level than the only child.

4. Immediately after birth, because of a diluted environment, the second born develops more slowly than did the firstborn at the same age. He or she continues at that lower rate of growth until an age equal to at least the length of the interval separating him/her from the older sibling.

5. The second born can surpass the firstborn at some age equal to at least the interval separating him or her from the older sibling.

6. The advantage of the second born over the firstborn that exists at age levels greater than the birth interval is temporary, and at maturity the firstborn exceeds the second born in intellectual level.

7. The changes in growth rates of first and second children given in 4, 5, and 6 occur for all next-to-last and last children.

Research Evidence

Data that could best reveal the dependence of birth order effects on age predicted by the confluence model would entail the *repeated* observations of entire families. Data of this type should exhibit the expected discontinuities in the growth curves. Unfortunately, information on entire families over time is not available. However, cross-sectional comparisons of individuals of different ages may be made from the existing literature, and although they are indirect, they are quite informative. Table 1 summarizes the literature on birth order effects arranged according to the age of subjects. The studies selected were those in which there was information about sibships of Sizes 1 or 2 and in which birth order effects were not confounded with family size. Studies comparing firstborns and second borns averaged over families of various sizes were thus not included.

Table 1 shows the pattern of results that we would indeed expect on the basis of the confluence model. For very young children the firstborn surpasses the second born. For children of older ages (from about age 3 to 13 years) the birth order effect seems to be reversed. From age 3 years to about age 13 or 14, with almost no exceptions, it is the second born who is higher in intellectual performance. The few exceptions occur at about age 11 (the age of the crossover) or in samples that span several ages. Homogeneous age groups, however, show a consistent pattern. From age 14 up, the superiority of the firstborn reappears and is maintained thereafter without exception. Studies that examine adults never show the second born surpassing the firstborn. For example, Cattell and Brimhall (1921) found firstborns to be overrepresented in *American Men of Science* relative to second borns in two-child families. Gini (1915) reported an overrepresentation of firstborns among Italian university professors, as did Poole and Kuhn (1973) among British university graduates.

Studies on birth order effects with children

[4] It should be clear that the second born's environment will be necessarily superior only at ages greater than the length of the birth interval.

of intermediate ages that were not included in Table 1 because of the confounding of birth order with family size or because of insufficient information, but in which inferences can be made about the likely form of this confounding, are in general agreement with the studies that were included (Kellaghan & MacNamara, 1972; Marjoribanks & Walberg, 1975; McCall, 1973; McCall & Johnson, 1972; Murray, 1971; Oberlander & Jenkins, 1967; Orum, 1971; Steckel, 1930; Thurstone & Jenkins, 1929). Interestingly, many of these studies, especially those having populations of young teenagers, show no birth order effects and often cite their results as counterexamples to other studies that claim birth order to be a significant factor in intellectual development.

Since some studies show negative effects, others positive, and still others no differences at all among the various birth orders, it is clear why birth order acquired such an unsavory reputation as an explanatory variable. But Table 1 reveals that the data on birth order are far from chaotic and that these data can be ordered fairly systematically according to age. The predicted crossover effect is clearly seen in the pattern of the reported

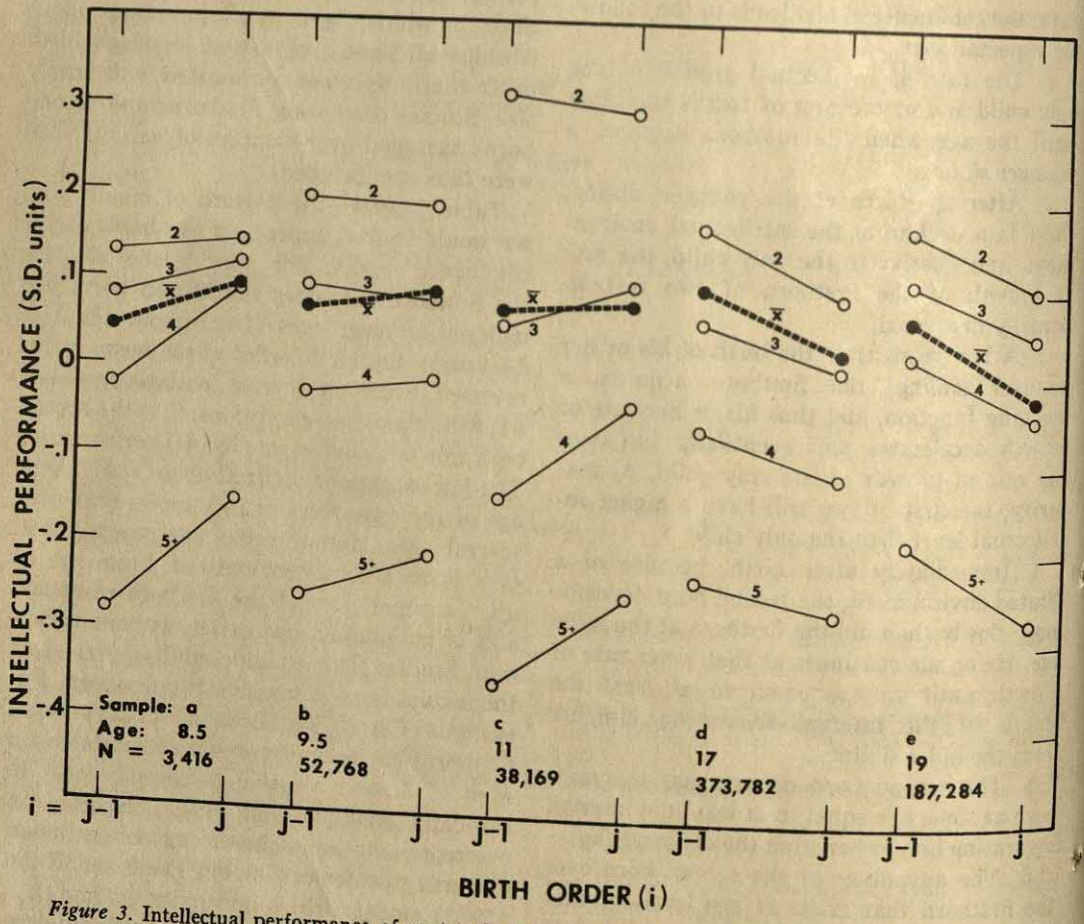


Figure 3. Intellectual performance of next-to-last and last children of different family sizes in five populations differing in age. (Family size is indicated by each curve. $\bar{X}$ indicates the average of all family sizes. The five populations are [a] Berkeley Unified School District study [Jensen, 1974]; [b] *Enquête nationale sur le niveau intellectuel des enfants d'âge scolaire* [INED, 1973]; [c] The trend of Scottish intelligence [Scottish Council for Research on Education, 1949]; [d] National Merit Scholarship Qualification Test Study [Breland, 1974]; [e] Survey of Dutch Recruits [Belmont & Marolla, 1973]. S.D. = standard deviation.)

results. There are two results for children less than 3 years of age. For both, the firstborn shows superiority. There are 12 studies with children aged 3 to 10 years. Of course, 10 show the second born as superior. The remaining 2 results in which the difference in intellectual performance favors the firstborn come from samples that spill over to an older age category, that is, 6 to 14 years (INED, 1973; Jensen, 1974). For the crossover interval 11 to 12 years, 4 results favor the firstborn and 5 favor the second born. Finally, for children over age 12 all results but 1 (18 out of 19) favor the firstborn.

The logic of the predictions indicating how the differential rates at which α and λ contribute to intellectual growth applies to all last and next-to-last children. In this regard, Figure 3 shows the results of five separate surveys of intellectual performance in which ages of the subjects vary from 8.5 to 19 years. The figure shows the scores of the next-to-last and last children in various sibship sizes. It is clear that the pattern found for children in sibships of Size 2 in Table 1 is duplicated quite systematically on all last and next-to-last children. Of the 12 observations of children 11 years of age or younger, the aggregate data favor the second born in 9 cases. Of the 8 observations of older children, all aggregate data show the opposite: The firstborn has a higher score.

Although the pattern of results in Table 1 and Figure 3 is consistent, there remains some ambiguity. Age differences in these results are confounded with a variety of other differences. A homogeneous population of children of different ages tested with the same materials would be virtually free of this confounding. Data of this type were secured from the Institut National d'Etudes Démographiques, which a few years ago conducted extensive surveys of intellectual performance. The second survey carried out by this institution (INED, 1973), examined a sample of more than 125,000 French schoolchildren. It utilized an intelligence measure containing subtests on vocabulary, comprehension, analogies, differences, series completion, and proverbs.[5]

Figure 4 shows intellectual performance of only children and of first- and second borns in two-child families of eight successive age groups.[6] It is clear from the age changes in birth order effects that the predicted patterns are borne out. Among younger children aged 6 to 8 years, the second born surpasses the firstborn. Among the older children this pattern is reversed. Moreover, as predicted, only children score higher than the firstborns among the youngest group, but from age 7 they show a consistent disadvantage. Analyses of variance showed both interactions to be significant, $F(7, 22881) = 2.59$; $p < .01$; $F(7, 22187) = 1.97$, $p < .05$. Figure 4 also shows comparisons for all next-to-last and last children. The pattern of these data is similar to those of two-child families, although the difference in favor of the next-to-last over the last child is not present among other children. Possibly, the full reversal of the birth order effect does not occur for these children until a later age.

In the above data, the firstborns and the second borns are unrelated. Data on first- and second borns from the *same* family would be preferable because these data would not be confounded with extraneous environmental factors such as socioeconomic status. Tabah and Sutter (1954) published an analysis of two-child families based on an earlier French survey (Tabah & Sutter, 1954), and calculations of the scores of firstborns and second borns of different ages can be made directly from their tables. For children who are 9 years old or younger, the second borns are higher in IQ than the firstborns (standard scores of $-.175$ vs. $-.003$ for firstborns and second borns, respectively). For children who are older than 9, the pattern is reversed (.048 vs. .036). This interaction of birth order with age

---

[5] Professor A. Girard, Chief of the Department of Social Psychology. Institut National d'Etudes Démographiques in Paris was kind enough to make some of these data available for the present analysis.

[6] The rising trend of intellectual scores with age results from the manner in which the tests were constructed. Four different forms of the test were prepared. The lowest school grades received only the two most difficult forms, whereas the intermediate grade received various combinations of the other forms. For the purposes of this analysis, however, this overall rising trend can be ignored.

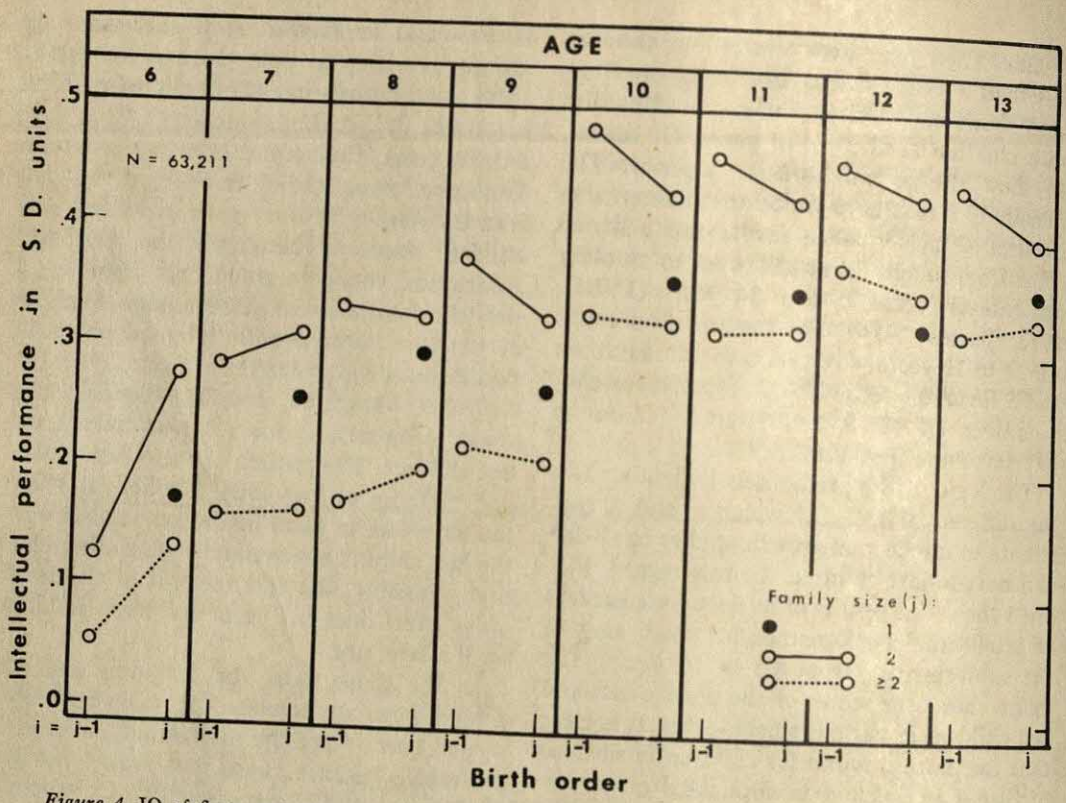


Figure 4. IQ of first- and second borns in families of two and of all next-to-last and last children as a function of age (calculated from data supplied by INED, 1973). S.D. = standard deviation.

is again significant, $F(1, 2476) = 6.65$, $p < .01$.

Conclusions

The predicted effects examined in this article clarify the seemingly conflicting reports on the relationship between birth order and intellectual performance. According to the confluence model, one should not be surprised to find birth order producing positive effects, negative effects, and no effects at all. Such findings, prevalent in the literature on birth order, lead some researchers to regard it as a fruitless field of study. Systematic patterns, however, can now be seen. It has been shown previously (Zajonc, 1976; Zajonc & Markus, 1975) that birth order effects can be better understood by considering the spacing among siblings. In the present article it has been demonstrated that birth order effects are also age dependent. Table 1 shows patterns of results that are entirely consistent with the inferences made from the confluence model. For very young children (i.e., children whose age is less than the typical interval in two-child families) the firstborn surpasses the second born. This advantage is then reversed, and a positive birth order effect prevails from age 3 or 4 years until the early teens. Finally, during the middle teens there is a return to an effect that favors the firstborn and that persists until maturity, and possibly permanently.

This complex pattern of results may be explained by the dynamic interplay between the contributions of the intellectual environment and the child's opportunity to serve as an intellectual resource. Certainly, other factors may contribute to the patterns examined in this article. The age dependence of birth order effects has been noted previously by Cicirelli (1967), who interpreted it as "some sort of trend where at an early age the later born child benefits from the stimulation of an older sibling, and at a later age (where the abstract verbal abilities come into play in the

school situation) the firstborn child profits from his closer exposure to adults" (pp. 482–483). It is also possible that the decline in the growth rate of the firstborn following the birth of a younger sibling may be due to a sudden shift in parental attention and care (Dunn, Note 2). Yet these alternative explanations lead to a variety of implications that are not borne out by the data. The first conjecture would lead us to expect an eventual equalization of performance when the second born reaches the school level, and the second explanation suggests a permanent disadvantage for the older child.

Among other explanations of family size and birth order effects are those couched in biological terms. For example, there is some evidence that perinatal asphyxia-anoxia may produce brain damage in children (Apgar, Girdany, McIntosh, & Taylor, 1955; Campbell, Cheeseman, & Kilpatrick, 1950) and that its incidence increases with parity. Nutritional deficit and hormonal and mineral depletion have also been singled out as factors associated with successive parities and especially with those following each other in close succession (Mertz & Cornatzer, 1971). In general, biological explanations of birth order and family size effects are at present not capable of dealing with the crossovers of the growth rates of only children with first- and second borns, with the crossovers between the siblings of two-child families, and with the crossovers between all last and next-to-last children. That birth order is strictly an environmental factor is strongly suggested by an interesting study of 845 subjects from 120 biological and 104 adoptive families carried out by Scarr and Weinberg (1977). These authors found no differences in birth order effects on IQ when biological and adoptive birth orders were compared. In fact, in a number of regression analyses, the coefficients for biological and adoptive birth order were remarkably similar.

Within the context of the confluence model, it is quite clear that the crossover patterns of growth found in the data cannot be accounted for without assuming some form of facilitation, such as the teaching function, that occurs following the birth of a younger sibling. Thus, both the theory presented here and the confirming data give additional support to the idea that older children do derive some benefit from the presence of a younger sibling.

It must be pointed out that the effects found here are for the most part rather weak. The largest difference in intellectual performance (found between extreme family sizes) is on the order of one standard deviation. Birth order differences are even smaller; differences between adjacent birth ranks are sometimes less than one tenth of a standard deviation. But if these differences had been large and systematic to begin with, the role of family configuration in intellectual development would have been understood long ago.

In a recent article, Grotevant, Scarr, and Weinberg (1977) applied the early outline of the confluence model to individual data and claimed that only 2% of the variance in IQ was accounted for by that version of the model. They concluded that the model was inappropriate for individual data and is useful only for population trends. The 2% of variance that Grotevant et al. found they could explain, however, is not an accurate reflection of what the model can do. Their figures were calculated on the basis of the nonparametrized version of the model and used IQ values rather than mental ages. The predictions, moreover, were made *without allowing* any free parameters whatever and without estimating them from the data. Additional peculiarities in the calculations of Grotevant et al. make their results less than useful in evaluating the performance of the confluence model. Berbaum and Moreland (Note 3) used the current parametrized version of the model (see Appendix) and mental ages as the basis for calculating predicted values in a more recent analysis of individual scores that accounted for nearly half of the variance in the mental ages of 257 children from 51 families.

As we have seen, the present version of the model fits aggregate data quite well, and the previous simpler linear version (Markus & Zajonc, 1977) performed quite well, too. As much as 95% and no less than 88% of the variance can be accounted for with three free parameters when the model is tested on all available large data sets. These predictions

were obtained for a variety of populations from different countries that differ in age, sibling spacing, and socioeconomic background (Zajonc & Bargh, Note 4). Moreover, different data sets involve different measures of intellectual performance. With only two free parameters (k and w_2), as much as 92% and no less than 78% of the variance in aggregate data can be accounted for by the predictions from the model.[7]

The performance of the model evaluated by Berbaum and Moreland (Note 3) must be evaluated against some expectation concerning the amount of variance in IQ that the model *should* explain. If the confluence model were constructed for the particular purpose of accounting for *all* the differences found in IQ —those of genetic origin, those associated with environmental factors, and those resulting from an interaction of these sources—then we would be quite justifiably disappointed in its performance. As much as 25% of the variance in the IQ of children can be attributed to the IQ of their parents alone. However, there is little sound theory *why* as much as 25% of variance in children's IQ is "controlled" by their parents' scores. Some theorists stress genetic factors and use this finding to support a genetic theory of intelligence, others interpret these data as support for an environmental theory, or one that focuses on the correlation and the interaction of these factors (Meredith, 1973). It is to be noted, however, that our analyses that involve cross-sectional aggregate data (that is, all except those in Table 2) have no confounding with genetic factors, since the contribution of parents in these data is held constant at the population average. Within these constraints, the pattern of results reviewed here conforms well to the basic predictions of the confluence model. A more appropriate test requires longitudinal within-family comparisons of cognitive growth, linked with observations of parent–child and child–child interactions. Since the model describes an individual process of intellectual growth, dynamic (temporal) rather than static (cross-sectional) tests are required in verifying its validity. Only repeated observations made at critical intervals on the same individuals can reveal whether the intellectual growth curves manifest the sorts of disturbance, coinciding with the births of younger siblings, that the model predicts. Data of this kind, including the necessary family information and an adequate sample size, are not available. When such research has been carried out, it will illuminate the transmission of environmental effects that emerge in the pattern of individual differences in intelligence described by the extensive literature on family configuration.

[7] For the sake of completeness, a residual factor $\delta_t = w_3\Delta f(t)$, encompassing all influences on intellectual growth not otherwise specified in the model, might also be included. Because this term would reflect all the influences on growth that are extraneous to the home, it could provide useful information. For example, when comparing different subpopulations that have access to different educational or child-rearing resources, comparisons of the relative magnitudes of α and δ might suggest the extent to which intellectual development in these subpopulations reflects factors such as the neighborhood, the school experience, or day-care facilities, for instance, and circumstances strictly related to the home.

Reference Notes

1. Claudy, J. G. *Cognitive characteristics, family size, and birth order*. Paper presented at the meeting of the American Psychological Association, Washington, D.C., August 1976.
2. Dunn, J. *Interaction between young siblings in the context of family relationships*. Unpublished manuscript, 1977. (Available from J. Dunn, Sub-Department of Animal Behavior, Madingley, Cambridge CB3 8AA, England.)
3. Berbaum, M., & Moreland, R. *Intellectual development within the family: A new application of the confluence model*. Paper presented at the meeting of the Midwestern Psychological Association, Chicago, May 1978.
4. Zajonc, R. B., & Bargh, J. *The confluence model: Predicting the relationship between family factors and intellectual performance*. Unpublished paper, 1978. (Available from the first author.)

References

Abe, K., Tsuji, K., & Suzuki, H. The significance of birth order and age difference between sibs as observed in drawings of prekindergarten children. *Folia Psychiatrica et Neurologica Japonica*, 1963, *17*, 315–325.

Altus, W. D. Birth order and scholastic aptitude. *Journal of Consulting Psychology*, 1965, *29*, 202–205.

Apgar, V., Girdany, B. R., McIntosh, R., & Taylor,

H. C. Neonatal anoxia: I. A study of the relation of oxygenation at birth to intellectual development. *Pediatrics*, 1955, *15*, 653–662.

Arthur, G. The relation of IQ to position in family. *Journal of Educational Psychology*, 1926, *17*, 541–550.

Bayer, A. E. Birth order and college attendance. *Journal of Marriage and the Family*, 1966, *28*, 480–484.

Bayley, N. Comparisons of mental and motor test scores for ages 1–15 months by sex, birth order, race, geographical location, and education of parents. *Child Development*, 1965, *36*, 379–411.

Belmont, L., & Marolla, F. A. Birth order, family size, and intelligence. *Science*, 1973, *182*, 1096–1101.

Breland, H. M. Birth order, family configuration, and verbal achievement. *Child Development*, 1974, *45*, 1011–1019.

Bryant, G. M., & Davies, K. L. The effect of sex, social class, and parity on achievement of Denver Development Screening Test items in the first year of life. *Developmental Medical Child Neurology*, 1974, *16*, 485–493.

Burton, D. Birth order and intelligence, *The Journal of Social Psychology*, 1968, *76*, 199–206.

Campbell, W. A. B., Cheeseman, E. A., & Kilpatrick, A. W. The effects of neonatal asphyxia on physical and mental development. *Archives of Disease in Childhood*, 1950, *25*, 351–359.

Cattell, J. K., & Brimhall, D. R. *American men of Science* (3rd ed.). Garrison, N.Y.: Science Press, 1921.

Chittenden, E. A., Foan, M. W., & Zweil, D. M. School achievement of first-and second-born siblings. *Child Development*, 1968, *39*, 1223–1228.

Cicirelli, V. G. Sibling constellation, creativity, IQ, and academic achievement. *Child Development*, 1967, *38*, 481–490.

Commins, W. D. The intelligence of the later born. *School and Society*, 1927, *25*, 488–489.

Davis, D., Cahan, S., & Bashi, J. Birth order and intellectual development: The confluence model in the light of cross-cultural evidence. *Science*, 1977, *196*, 1470–1472.

Devin-Sheehan, L., Feldman, R. S., Allen, V. L. Research on children tutoring children: A critical review. *Review of Educational Research*, 1976, *46*, 355–385.

Duff, R. E. Effects of pupil-tutoring on self-perception and academic achievement of primary grade tutors and tutees (Doctoral dissertation, Southern Illinois University, 1973). *Dissertation Abstracts International*, 1974, 5473A-5474A. (University Microfilms No. 74–6262)

Ellis, H. A. *A study of British genius*. Boston: Houghton Mifflin, 1904.

Eysenck, H. J., & Cookson, D. Personality in primary school children: 3. Family background. *British Journal of Educational Psychology*, 1970, *40*, 117–131.

Galton, F. *English men of science: Their nature and nurture*. London: Macmillan, 1874.

Gini, C. Superiority of the eldest. *Journal of Heredity*, 1915, *37*, 37–39.

Grotevant, H. D., Scarr, S., & Weinberg, R. A. Intellectual development in family constellations with adopted and natural children: A test of the Zajonc and Markus confluence model. *Child Development*, 1977, *48*, 1699–1703.

Hill, H. S. Resemblance of bilingual siblings in verbal intelligence. *School and Society*, 1936, *43*, 271–272.

Hsiao, H. H. The status of the first-born with special reference to intelligence. *Genetic Psychological Monograph*, 1931, *9*, 1–118.

Institut National d'Etudes Démographiques. *Enquête nationale sur le niveau intellectual des enfants d'âge scolaire*. Paris, 1973.

Jensen, A. R. Cumulative deficit: A testable hypothesis? *Developmental Psychology*, 1974, *10*, 996–1019.

Kellaghan, T., & MacNamara, J. Family correlates of verbal reasoning ability. *Developmental Psychology*, 1972, *7*, 49–53.

Koch, H. L. The relation of "primary mental abilities" in five- and six-year-olds to sex of child and characteristics of his siblings. *Child Development*, 1954, *25*, 209–223.

Lunneborg, P. W. Birth order, aptitude, and achievement. *Journal of Consulting and Clinical Psychology*, 1968, *32*, 101.

Lunneborg, P. W. Birth order and sex of sibling effects of intellectual abilities. *Journal of Consulting Clinical Psychology*, 1971, *37*, 445.

Marjoribanks, K., & Walberg, H. J. Ordinal position, family environment, and mental abilities. *Journal of Social Psychology*, 1975, *95*, 77–84.

Markus, G. B., & Zajonc, R. B. Family configuration and intellectual development: A simulation. *Behavioral Science*, 1977, *22*, 137–142.

McCall, J. N. Birth order differences in special ability: Fact or artifact? *Psychological Reports*, 1973, *33*, 947–952.

McCall, J. N., & Johnson, O. G. The independence of intelligence from size and birth order. *Journal of Genetic Psychology*, 1972, *121*, 207–213.

Meredith, W. A model for analyzing heritability in the presence of correlated genetic and environmental effects. *Behavior Genetics*, 1973, *3*, 271–277.

Mertz, W., & Cornatzer, W. E. *Newer trace elements in nutrition*. New York: Marcel Dekker, 1971.

Morgan, R. F., & Toy, T. B. Learning by teaching: A student-to-student compensatory tutoring program in a rural school system and its relevance to the educational cooperative. *Psychological Record*, 1970, *20*, 159–169.

Murray, C. The effects of ordinal position on measured intelligence. *British Journal of Social and Clinical Psychology*, 1971, *10*, 221–227.

Oberlander, M., & Jenkins, N. Birth order and academic achievement. *Journal of Individual Psychology*, 1967, *23*, 103–109.

Oberlander, M., Jenkins, N., Houlihan, K., & Jackson, J. Family size and birth order as determinants of scholastic aptitude and achievement in a sample of

eighth graders. *Journal of Consulting and Clinical Psychology*, 1970, *34*, 19–21.

Orum, B. *Social baggrund, intellektuelt niveau og placering i skilesystemet*. Copenhagen: Socialforskiningsinstituttet, 1971.

Poole, A., & Kuhn, A. Family size and ordinal position: Correlates of academic success. *Journal of Biosocial Science*, 1973, *5*, 51–59.

Richardson, S. K. The correlation of intelligence quotients of siblings of the same chronological age levels. *Journal of Juvenile Research*, 1936, *20*, 186–198.

Richer, H. M. Peer teaching as a facilitator of learning: Using conservation of substance as a measure (Doctoral dissertation, University of California, Los Angeles, 1973). *Dissertation Abstracts International*, 1973, 641A–642A. (University Microfilms No. 73-18,646.

Rosenberg, B. G., & Sutton-Smith, B. Siblings age spacing effects upon cognition. *Developmental Psychology*, 1969, 661–668.

Scarr, S., & Weinberg, R. A. Intellectual similarities within families of both adopted and biological children. *Intelligence*, 1977, *1*, 170–191.

Schachter, S. Birth order, eminence and high education. *American Sociological Review*, 1963, *28*, 757–768.

Schooler, C. Birth order effects: Not here, not now! *Psychological Bulletin*, 1972, *78*, 161–175.

Schoonover, S. M. The relationship of intelligence and achievement to birth order, sex of sibling, and age interval. *Journal of Educational Psychology*, 1959, *50*, 143–146.

Scottish Council for Research on Education. *The trend of Scottish intelligence*. London: University of London Press, 1949.

Steckel, M. L. Intelligence and birth order in family. *Journal of Social Psychology*, 1930, *1*, 329–344.

Svensson, A. *Hembakgrund och prestationsnivå*. Goteborg, Sweden: Pedagogiska Institutionen, 1972.

Tabah, L., & Sutter, J. Le niveau intellectuel des enfants d'une même famille. In *Le niveau intellectuel des enfants d'âge scolaire*. Paris: Institut National d'Etudes Démographiques with Presses Universitaires de France, 1954.

Terhune, K. W. *A review of the actual and expected consequences of family size*. (Publication No. NIH 76-779). Washington, D.C.: U.S. Department of Health, Education, and Welfare, 1976.

Thurstone, L. L., & Jenkins, R. L. Birth order and intelligence. *Journal of Educational Psychology*, 1929, *20*, 641–651.

Willis, C. B. The effects of primogeniture on intellectual capacity. *Journal of Abnormal Social Psychology*, 1924, *18*, 375–377.

Wu, P. P. C. The effects of a peer tutorial program on academic achievement and self-concept of low achieving high school mathematics students (Doctoral dissertation, Florida State University, 1973). *Dissertation Abstracts International*, 1974, 5612A–5613A. (University Microfilms, No. 74–6607).

Zajonc, R. B. Family configuration and intelligence. *Science*, 1976, *192*, 227–236.

Zajonc, R. B., & Markus, G. B. Birth order and intellectual development. *Psychological Review*, 1975, *82*, 74–88.

Appendix

The Confluence Model: Reparametrized Version

The level of mental maturity, $M_{ij(t)}$, attained at age t by the ith child of j children in a family (see Footnote 1) of n individuals, is expressed as a first-order difference equation

$$M_{ij}(t) = M_{ij(t-1)} + \alpha_t + \lambda_t, \tag{A1}$$

where $t = 1, 2, 3, \ldots$ years. The sum of the two components, α_t and λ_t, both measured at age t, represents the growth increment accumulating each year.

Since intellectual growth is not linear with age, the environmental factors that influence it do not have the same effects at different ages of the child. Thus, the acquisition of a sibling at age 3 years has a more pronounced effect than at age 12, when the child is much nearer intellectual maturity. In the original article (Zajonc & Markus, 1975), we supposed that the effects of family configuration cumulated as a sigmoid function of age, $f(t) = 1 - e^{-k^2t^2}$. The two components α_t and λ_t are therefore expressed as weighted yearly increments of this sigmoid function, $\Delta f(t) = (1 - e^{-k^2t^2}) - (1 - e^{-k^2(t-1)^2}) = e^{-k^2(t-1)^2} - e^{-k^2t^2}$:

$$\alpha_t = w_1 \Delta f(t) [\sum_{i=1}^{n} M^2_{in(t-1)}/n_{(t-1)} + 1]^{\frac{1}{2}}, \tag{A2}$$

and

$$\lambda_t = \frac{w_2 L_t \Delta f(t) \Delta f(\tau)}{(n_t - 1)^2}. \tag{A3}$$

In the above terms, w_1 and w_2 are weights associated with the two components, k is a constant, $f(\tau) = 1 - e^{-k^2\tau^2}$ where τ is the age of the adjacent younger sibling, and L_t is the last-child index, which is equal to 0 for children who at age t are not followed by a younger sibling and is equal to 1 otherwise. The weights w_1 and w_2 represent metric contributions of the component sources of influence upon intellectual growth and need to be estimated from the data. The constant k in $f(t)$ and $f(\tau)$ must also be estimated, and it reflects the rate at which intellectual effects cumulate. It will therefore vary with the population investigated, with the type of ability measured, or with the test employed for its measurement.

In Equation A2 the intellectual environment within the home is represented as a root mean square of the intellectual levels of all family members. This formulation is used, as previously (Markus & Zajonc, 1977), to reflect the greater contribution of the more mature members. Moreover, the denominator is $n + 1$ to allow the quality of the intellectual environment to rise with increasing number of family members. We would not want to consider an environment consisting of one adult to be of the same quality as one consisting of four adults. Thus, for constant values of M, the inequality $nM/(n + 1) < (n + 1)M/(n + 2)$ is always satisfied. It can be readily seen that this formulation of the intellectual environment also reflects the fact that an addition of a new birth will reduce its quality, whereas the addition of an adult will increase that quality.

Received July 26, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 8, 1342-1351

Cultural Difference in Attribution of Performance: An Integration-Theoretical Analysis

Ramadhar Singh, Meenakshi Gupta, and Ajit K. Dalal
Department of Humanities and Social Sciences
Indian Institute of Technology, Kanpur, India

Heider's suggestion that Performance = Motivation × Ability has been empirically confirmed by Anderson and Butzin and by Kun, Parsons, and Ruble, using American students as subjects. This multiplying process failed to appear in the present series of three experiments performed on Indian college students. Contrary to the predicted linear fan pattern, the plot of Motivation × Ability effect displayed clear parallelism. An equal-weight averaging rule was able to account for the results obtained in both group and single-subject analyses. Perhaps the integration rules underlying achievement judgments are culture-specific, and Indian college students average motivation and ability information in attribution of future scholastic performance. These results illustrate the potential power that information integration theory provides for the cross-cultural study of social perception and cognition.

How do people integrate information about motivation and ability when they predict performance? Heider (1958) made the following suggestion:

> The personal constituents, namely, power (ability) and trying (effort) are related as a multiplicative combination, since the effective personal force (performance) is zero if either of them is zero. For instance, if a person has the ability but does not try at all he will make no progress toward the goal. (p. 83)

If Heider's proposal is correct, then attribution of performance from motivation and ability information would follow the multiplying rule:

$$\text{Performance} = \text{Motivation} \times \text{Ability}. \quad (1)$$

Anderson and Butzin (1974) tested this multiplying rule in two experiments. They presented information about the motivation and ability of target persons, applicants to graduate school and athletes trying out for college track, and asked subjects to predict how the target persons would perform. The factorial plot of the Motivation × Ability judgments was a diverging fan of straight lines. By the logic of functional measurement (Anderson, 1974a, 1976), this linear fan pattern implies the operation of a multiplying rule.

In a developmental study, Kun, Parsons, and Ruble (1974) also obtained evidence for a multiplying rule. Although the youngest children (5–6 years) integrated ability and trying information by an adding-type rule, the linear fan pattern was already present in second graders. The results of these two studies suggest that there is considerable generality to the multiplying rule.

However, an anomalous result appeared

This work was supported by the Research Fund of the Department of Humanities and Social Sciences, Indian Institute of Technology, Kanpur, India. The authors thank K. A. Bohra for his able assistance in data analyses and Norman H. Anderson, H. S. Asthana, John H. Harvey, Lilavati Krishnan, Kent L. Norman, and Jai B. P. Sinha for their comments on an early version of this article. The data reported here were also presented at the 66th annual meeting of the Indian Science Congress Association held at Hyderabad, India, January 1979.

Ajit K. Dalal is now at the Department of Psychology, University of Allahabad, Allahabad-211002, India.

Requests for reprints should be addressed to Ramadhar Singh, Organizational Behavior Area, Indian Institute of Management, Ahmedabad, 380 015, Gujrat, India.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3708-1342$00.75

when Anderson and Butzin (1974) presented information about performance and ability and asked subjects to judge motivation. Equation 1 implies a dividing rule,

$$\text{Motivation} = \text{Performance} \div \text{Ability}. \quad (2)$$

However, the data from both experiments supported a subtracting rule,

$$\text{Motivation} = \text{Performance} - \text{Ability}. \quad (3)$$

Anderson and Butzin concluded, therefore, that these judgments obeyed a simple cognitive algebra but that this cognitive algebra was not consistent mathematically.

Perhaps there is an alternative interpretation, one in which the subjects are mathematically consistent. The linear fan pattern is not unique to the multiplying rule. This pattern can also be produced by a conjunctive averaging rule with differential weighting (Anderson, 1971, p. 85). If lower values of motivation and/or ability had greater weight, then the averaging model would produce an approximate linear fan. This conjunctive averaging idea seems reasonable, for it reflects the idea that no amount of motivation can compensate for low ability, contrary to the implications of the multiplying model. This possibility was noted by Anderson and Butzin (1974), who said, "Since averaging processes are pervasive in judgment, a conjunctive integration rule deserves more consideration" (p. 602). However, this possibility has not been considered further.

A test between the multiplying rule and the conjunctive averaging rule can be obtained by asking for judgments based only on ability information. The multiplying rule can only operate if the subject imputes some implicit level of motivation. With this implicit level of motivation, the ability-only curve will still form part of a linear fan. But if the averaging model holds, then the well-known crossover interaction will be observed (Anderson, 1974a; Lampel & Anderson, 1968). No such test was used by Anderson and Butzin (1974) or by Kun et al. (1974).

Experiment 1

Experiment 1 had two purposes. The first was to replicate previous results on the multiplying rule, Performance = Motivation × Ability. The second was to test the plausibility of the alternative interpretation as a conjunctive averaging rule.

Method

Stimuli and design. Sixteen stimulus persons were prepared according to a 4 × 4 design comparable to that of Anderson and Butzin (1974). The two factors were *interest in studies* and *IQ*. The four levels of interest in studies were *slight, less than most, more than most,* and *very high*; the four levels of IQ were 100, 110, 120, and 130. In addition, 4 stimulus persons were described with respect to their IQ alone; their interest in studies was not specified. The information about each of these 20 stimulus persons was typed on a separate sheet of paper, and all 20 sheets were arranged randomly in a booklet.

Four practice examples were also constructed, using stimuli more extreme than the regular levels of the two factors. Levels of interest in studies were *not at all interested* and *extremely interested*; IQ levels were 90 and 140. These practice examples were intended to serve as end anchors and to orient the subjects toward the use of the entire response scale (Anderson, 1974b, Note 1).

Subjects. Subjects were 32 male second-year engineering students enrolled in an introductory psychology course at the Indian Institute of Technology, Kanpur, India. Participation fulfilled a course requirement.

Procedure. Subjects, gathered in groups from three to five, received a typed sheet of instructions that described the nature of the experimental task and their role as subjects. The task was introduced as dealing with prediction of future academic performance of first-year engineering students at the Indian Institute of Technology, Kanpur. It was emphasized that prediction about future performance would be based on interest in studies and intelligence or on intelligence alone. Subjects were instructed to base their prediction only on the information given about each stimulus student.

After reading the instruction sheet twice, each subject worked with the four practice examples. He read the information about each stimulus student and then indicated how that stimulus student would perform at the institute. Prediction of performance was made along a 21-point scale, with endpoints labeled −10 (very bad performance) and +10 (very good performance).

Immediately following the practice period, the main points of the instructions were summarized to the subjects by the experimenter. All queries about the task were answered. Finally, each subject received the experimental booklet and rated all 20 stimulus students twice. In each case, he wrote the code number of the target person and his judgment of performance on the response sheet with the 21-point response scale typed at the top. The first replication

was considered as additional practice, and only the data from the second replication were analyzed.

Results

Figure 1 plots mean judgments as a function of interest in studies (curve parameter) and IQ. The IQ levels are spaced on the horizontal axis according to the marginal means of the factorial design. This spacing allows the linear fan pattern to appear. If the two pieces of information were integrated in accord with the multiplying rule, then the four solid curves would form a diverging fan of straight lines.

It is clear that the pattern of the data is contrary to the multiplying model. There is not any evidence for divergence at all. Instead, the four solid curves display parallelism. This parallelism indicates that subjects integrated the given information by an adding or averaging rule.

The statistical analysis also supported the graphical interpretation. The linear fan pattern predicted by the multiplying model requires a statistically significant Interest in Studies × IQ interaction. However, this interaction was statistically nonsignificant, $F(9, 279) = 1.44$. This result argues against the multiplying model. On the other hand, it supports the parallelism prediction of the adding and averaging models.

A distinguishing test between the adding and averaging models may be obtained by considering the dashed curve in Figure 1. This curve represents judgments based on IQ information alone, with interest in studies not specified. The adding rule requires that the dashed curve be parallel to the solid curves; the averaging rule implies that the dashed curve should have steeper slope than the solid curves (see Anderson, 1974b, Section 3). Visual inspection of Figure 1 thus argues against the adding rule and for the averaging rule.

To obtain a statistical test that would discriminate between adding and averaging, the dashed curve data, which are based on IQ information alone, were considered as a fifth level of the interest-in-studies factor. The interaction term in this 5 × 4, Interest in

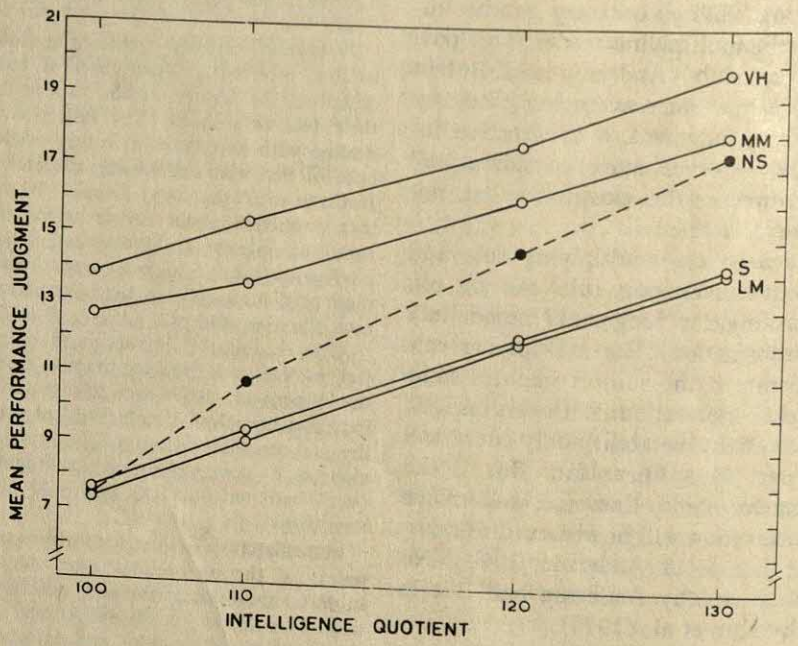


Figure 1. Mean judgment of performance as a function of interest in studies and IQ, Experiment 1. (Levels of *interest in studies* are listed as curve parameters: VH = very high; MM = more than most; LM = less than most; S = slight. The dashed curve is based on IQ information only; interest in studies is not specified [NS].)

Studies × IQ analysis of variance tests for nonparallelism in the set of all five curves in Figure 1. This interaction was highly significant, $F(12, 372) = 8.36$, $p < .01$. Since the four solid curves were essentially parallel, as shown above, it appears that the dashed curve is reliably steeper than the solid curves. This test thus supports the averaging hypothesis and rejects the alternative adding hypothesis.

Discussion

The main result of Experiment 1 is that the information about motivation and ability appears to be integrated by a simple, equal-weight averaging rule. There was no support for the linear fan pattern of the multiplying rule. This result is contrary to the prediction; it disagrees with results from American subjects.

One hypothesis is that this difference in integration rules is a consequence of cultural differences in outlook on social motivation. Before this hypothesis can be taken seriously, however, it is necessary to consider methodological differences between the present and previous experiments.

Methodologically, this experiment was different from those of Anderson and Butzin (1974) in two ways. First, they asked their subjects to describe in writing how they combined motivation and ability cues during the practice session. That was not done in the present work. This difference, however, does not seem to be very serious. Kun et al. (1974) obtained the linear fan pattern with children even without following the Anderson and Butzin procedure. Second, and perhaps more important, the motivation factor was defined here as interest in studies, and the subjects may not have interpreted that in terms of motivation and trying. This possibility was checked in the next experiment by using more explicitly motivational information.

Experiment 2

Experiment 2 was run chiefly as a reliability check on Experiment 1. There were, however, three notable changes from Experiment 1. First, information about laboriousness was used as a motivation cue to allow direct comparison of findings with those obtained in the United States. Second, the levels of the two factors, laboriousness and IQ, were drawn from a wider scale range to increase the opportunity for a linear fan pattern to appear. Finally, sufficient data were gathered from each subject to explore various integration rules on the individual level.

Method

Stimuli and design. A 4 × 4 (Laboriousness × IQ) factorial design was employed to construct 16 two-cue stimulus persons. The levels of laboriousness were *not at all laborious, slightly laborious, fairly laborious,* and *very laborious;* the levels of IQ were 90, 105, 120, and 135. In addition, 8 single-cue stimulus persons were described by one level of either factor.

The practice and filler stimulus persons were described by information more extreme than that used for the regular experimental stimulus persons. For example, laboriousness had *not at all laborious* and *extremely laborious* levels, and IQ had values of 85 and 140. These extreme stimuli were used to orient the subjects toward the use of the entire response scale and also to guard against the operation of floor and ceiling effects as in Experiment 1. Each person description was typed on a separate index card. The eight practice examples included four two-cue and four single-cue descriptions.

Subjects. Subjects were 12 male second-year engineering students fulfilling a course requirement in elementary psychology at the Indian Institute of Technology, Kanpur, India. Two subjects were run at a time in a session that lasted approximately 1.5 hours.

Procedure. The general procedure was the same as in Experiment 1 except that subjects read the entire set of descriptions before beginning their actual judgments. The complete set of 24 experimental and 2 filler descriptions was presented three times in different shuffled order for each subject. Data from all the three replications were analyzed.

Results

Parallelism prediction. The upper left panel of Figure 2 plots mean performance as a function of laboriousness (curve parameter) and IQ (horizontal axis) of the stimulus persons. If the result obtained in Experiment 1 is reliable, then the four solid curves should plot as parallel lines. Inspection shows that these curves are essentially parallel. The Laboriousness × IQ interaction in the 4 × 4 design was, however, significant, $F(9, 99) = 2.13$, $p < .05$. This corresponds to a slight tendency for the four curves to converge to the right. There is no sign of the diverging linear fan pattern reported in the American studies. Despite the

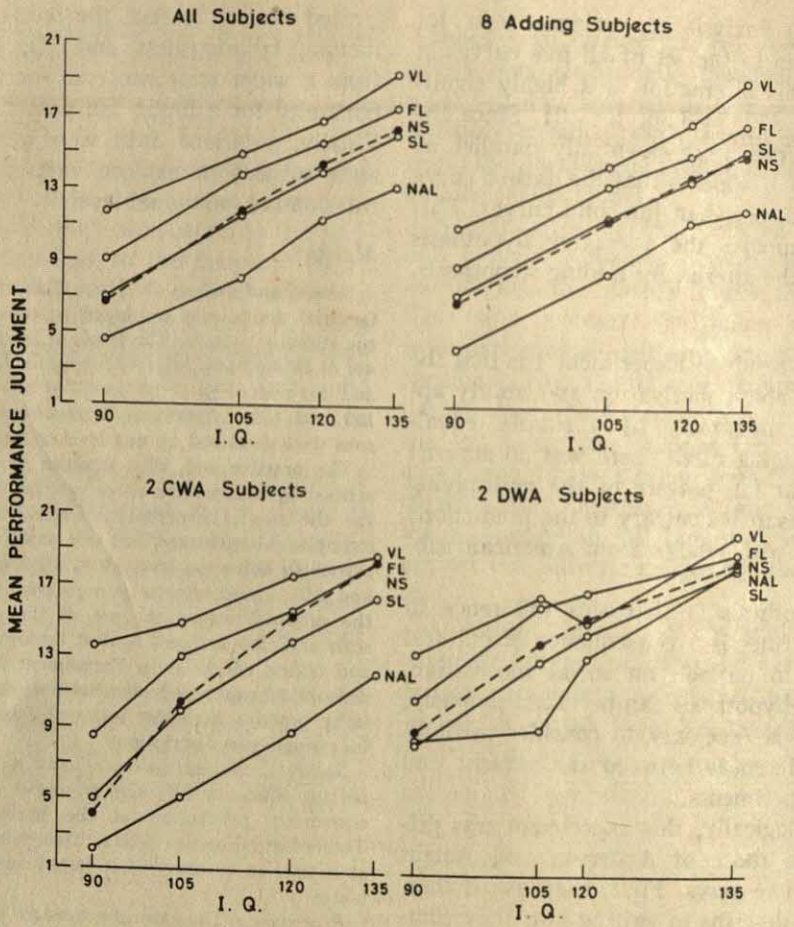


Figure 2. Mean judgment of performance as a function of laboriousness and IQ, Experiment 2. (Levels of *laboriousness* are listed as curve parameters: NAL = not at all laborious; SL = slightly laborious; FL = fairly laborious; VL = very laborious; NS = laboriousness not specified. Classification of subjects into adding, constant-weight averaging [CWA], and differential-weight averaging [DWA] subgroups was based on results from single-subject analyses.)

change in the nature of the motivation information, these results confirm those of the first experiment.

Averaging versus adding. The dashed curve in the upper panel of Figure 2 represents judgments based on IQ alone, with no information about laboriousness. This curve is nearly parallel to the solid curves. At face value, this seems to argue for the adding hypothesis and against the averaging hypothesis. The result raises some difficulty for both hypotheses, however.

The difficulty with the adding hypothesis is that the dashed curve has the same elevation as the solid curve for *slightly laborious*. Under the adding hypothesis, this equality in elevation requires that information that the student is slightly laborious has the same effect as lack of information about laboriousness. If *slightly laborious* has a value of zero, then of course there is no problem. However, if *slightly laborious* has a value different from zero, then *lack of information* must have the same value. In this case, therefore, it would be necessary to assume that the subject infers an implicit level of laboriousness when no explicit information is given and that he adds this value to the other information given.

The difficulty with the averaging model is the failure of the dashed curve to cross over the solid curve. However, the averaging model can account for the near-parallelism if it is

assumed that subjects infer an implicit value of laboriousness when no information is given and that they average that in with the other given information. If this inferred implicit value has the same weight as given laboriousness information, then the corresponding dashed curve will be parallel to the solid curves.

The assumption that subjects infer an implicit value for missing information about laboriousness seems reasonable for the present task. Logically, a prediction of performance cannot be made on the basis of IQ information alone; some information pertaining to degree of laboriousness is necessary. This question cannot be decided on the basis of the above data, but some evidence is available from Experiment 3.

Single-subject analyses. Separate analyses were made for each subject (Shanteau & Anderson, 1969, p. 315) to find out the various ways in which the experimental task was handled and also to explain the minor discrepancy from parallelism in the main analysis mentioned earlier. In these analyses, the subjects were broken down into three subgroups.

Mean performance judgments for these three subgroups of subjects are shown in Figure 2. The upper right panel displays results for eight subjects, all of whom showed nonsignificant interactions. Here parallelism is quite obvious. Perhaps these subjects followed an adding or an averaging rule. The lower left panel plots data from two subjects who apparently averaged, as shown by the crossover of the dashed and solid curves. The lower right panel shows two subjects who had a strong convergence tendency. However, the general shape of the curves is too irregular to be interpreted.

The most important information from these individual analyses is that not a single subject showed the diverging fan pattern obtained in the American studies. That supports the suggestion of Experiment 1 that Indian subjects make achievement judgments in a rather different way from their American counterparts.

Discussion

The parallelism pattern observed in Experiment 1 is replicated here. Although information about laboriousness, rather than interest in studies, served as the motivation factor, and although the laboriousness and IQ factors had levels from a wider range than in Experiment 1, the findings are basically identical. Most of the subjects obeyed the parallelism prediction. More importantly, not even one subject showed a diverging fan pattern predicted by the multiplying-type strategy. These results support the hypothesis that the differences between the American and Indian studies result from cultural differences.

The findings of the first and second experiments converge to make a single point: Information about motivation and ability is added or averaged in making attributions of future scholastic performance. The evidence leans toward the averaging hypothesis, but the results of Experiment 2 require the supplementary assumption that the subject imputes an implicit value for motivation when no information is given. This assumption is reasonable, since no performance is logically possible without some level of motivation. Nevertheless, more direct evidence is required.

Direct evidence can be obtained by using three stimulus cues. If the averaging model is correct, then the single-cue curves should be clearly steeper than the curves in the two-way factorial plots from the three-cue design. The reason for this is simple. When the single cue is given, any implicit inference about the absent motivation information will yield an effective two-cue response. But the averaging model implies that the slope of a curve based on two cues will be greater than that based on three cues. Accordingly, the crossover test remains valid even if the subject does impute an implicit value to the missing dimension of information. The adding hypothesis, of course, still predicts parallelism in such a three-cue design.

Experiment 3

Experiment 3 had two goals. The first was to extend the findings of Experiment 2 by including a third informational cue about past performance. The second was to resolve the ambiguity in the test of the averaging model in Experiment 2.

Method

Stimuli and design. There were two main designs. The first design was a 3 × 3 × 3 (Past Performance × Laboriousness × IQ) factorial. The three levels of past performance were Cumulative Performance Index (CPI) of 5, 7, and 9 during the first year of study at the Indian Institute of Technology, Kanpur. This information was from a 10-point scale. The three levels of laboriousness were *not at all laborious, fairly laborious,* and *very laborious;* the three levels of IQ were 90, 112, and 135. Orthogonal combinations of these three factors generated 27 three-cue stimulus persons.

The second design was a 3 × 3 (Laboriousness × IQ) factorial. The levels of laboriousness and IQ factors were the same as in the three-cue design. Pairing of the levels of these two factors produced nine two-cue stimulus persons.

There were also 9 single-cue stimulus persons, based on one of the three levels of each factor. In addition, 4 three-cue filler persons were constructed to serve as end anchors. All these 49 stimulus persons, including the 4 filler persons, were judged by each subject.

Procedure and subjects. The general procedure was the same as in Experiment 2. Each subject rated 49 stimulus persons three times. As in Experiment 2, a description of each stimulus person was typed on a separate index card.

Before rating the experimental stimuli, each subject worked with 10 practice examples prepared from extreme levels of the informational factors. Of the 10 examples, 4 had three cues, two had two cues, and the remaining four had just one cue.

There were 12 subjects from the same population as in Experiment 2. Two subjects were run at a time. Each subject spent approximately 2.5 hours on the task.

Results and Discussion

Three-cue design. The main purpose of Experiment 3 was to study the integration rule for the three cues. On the basis of their results, Anderson and Butzin (1974) suggested the compound averaging-multiplying rule:

$$\text{Predicted Performance} = \text{Past Performance} + \text{Motivation} \times \text{Ability}. \quad (4)$$

However, the results of the present Experiments 1 and 2 suggest the three-term averaging rule:

$$\text{Predicted Performance} = \text{Past Performance} + \text{Laboriousness} + \text{Ability}. \quad (5)$$

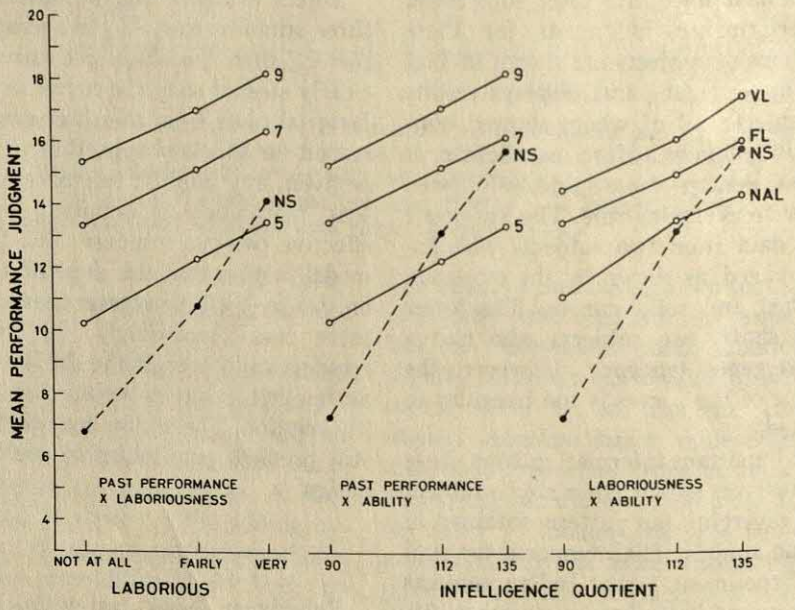


Figure 3. Two-way factorial plots of Past Performance × Laboriousness, Past Performance × Ability, and Laboriousness × Ability effects from the main three-cue, Past Performance × Laboriousness × Ability design. (The dashed curve of each panel is based on the single cue listed on the horizontal axis. Digits 5, 7, and 9 represent levels of past performance; NAL, FL, and VL represent *not at all laborious, fairly laborious,* and *very laborious,* respectively. NS means that the row information was not specified.)

The results of Figure 3 support the three-term averaging rule. Each panel shows one of the two-cue factorial graphs from the main design. In each panel, the solid curves are approximately parallel. The right panel, which corresponds to the Laboriousness × Ability interaction, is marginally significant, $F(4, 44) = 2.46$. However, the three solid curves show a slight convergence, as in Experiments 1 and 2, not the divergence predicted by the multiplying rule.

A test between the adding and averaging rules is obtained by comparing the dashed curve with the solid curves in each panel of Figure 3. The dashed curve represents judgments based on just the single cue listed on the horizontal axis. In each panel, the dashed curve has much steeper slope and crosses over the lowest solid curve convincingly. This crossover interaction is evidence against the adding rule and supports the averaging rule.

Two-cue design. The two-cue design was included to provide a further check on the two-cue results of Experiments 1 and 2. These data are shown in Figure 4. The solid curves are approximately parallel, but with a visible and barely significant tendency to converge to the right, $F(4, 44) = 2.85$, $p < .05$. The dashed curve is again nearly parallel to but slightly steeper than the solid curves. These results are essentially the same as those in Experiment 2.

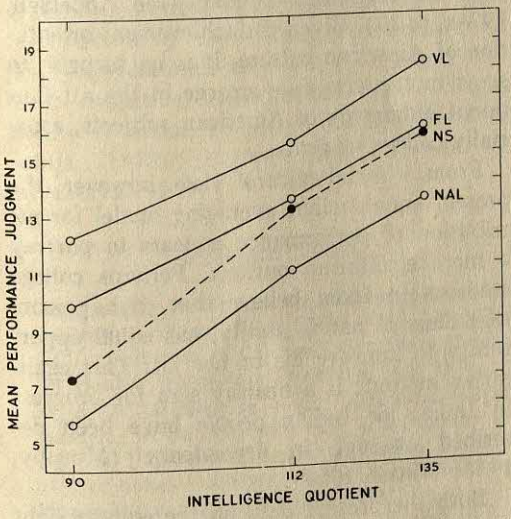


Figure 4. Mean judgment of performance as a function of *laboriousness* and IQ, Experiment 3, two-cue design. (NAL, FL, and VL represent *not at all laborious, fairly laborious,* and *very laborious,* respectively. NS means that the row information was not specified.)

Table 1
F *Ratios for Two-Way and Three-Way Interactions in Single-Subject Analyses (Experiment 3, Three-Cue Design)*

Subject	AB $F(4, 52)$	AC $F(4, 52)$	BC $F(4, 52)$	ABC $F(8, 52)$
1	2.69*	.58	.94	.84
2	2.23	3.33*	1.28	1.79
3	1.64	.74	1.77	.98
4	.84	1.19	3.73*	1.41
5	1.72	1.14	1.22	.87
6	.78	1.15	.77	.53
7	.35	.59	.95	.93
8	2.75*	1.09	1.18	3.50*
9	.42	.16	.91	.75
10	.51	.23	1.29	.91
11	1.36	.51	1.73	.27
12	1.65	.76	1.32	.68

Note. A = Past performance. B = Laboriousness. C = IQ.
* $p < .05$.

These two-cue data support the assumption of implicit inference about absent motivation information discussed in Experiment 2. The dashed curve in the center and right panels of Figure 3 is the same as the dashed curve of Figure 4. The clear crossover of the dashed curve in Figure 3 supports the averaging rule. Hence, the failure to obtain a clear crossover in Figure 4 would seem to reflect the presence of an inferred, implicit value for the missing motivation cue.

Single-subject analyses. As Shanteau and Anderson (1969) emphasized, individual analyses are important to check that the group averages are not hiding alternative integration strategies by different subjects. Table 1 presents the F ratios for the four interactions for each subject in the main three-cue design.

Table 1 shows that only five scattered interactions are statistically significant. Inspection of individual data failed to disclose any meaningful pattern to these deviations from parallelism. Instead, these deviations seemed to reflect individual idiosyncracies such as may be seen in the lower right panel of Figure 2. Almost all individuals showed the crossover

interaction for the single cue curves. Thus, the present results support the generality of the averaging rule at the level of the individual.

In single subject analyses of the two-cue, Laboriousness × IQ data, 11 of the 12 subjects showed parallelism. Only 1 subject had a converging type of nonparallelism. It can thus be said that present two-cue results are indeed similar to those of Experiments 1 and 2.

General Discussion

The results of the present set of three experiments do not agree with the multiplying model suggested by Heider (1958) and tested by Anderson and Butzin (1974) and by Kun et al. (1974). In the present experiments, attribution of performance from motivation and ability cues appears to have been governed by a process quite different from the multiplying. Strong evidence for parallelism pattern indicates that an equal-weight averaging rule could very well account for the data. It seems reasonable to suggest, therefore, that attribution of scholastic performance obeys an averaging rule in India.

Collateral work by Singh (Note 2) helps solidify this cultural-difference interpretation of the present results. Singh presented information about generosity and income of various persons and asked subjects to predict how much these persons would contribute to a family whose house had burned down. This experiment parallels that of Graesser and Anderson (1974) on Gift Size = Generosity × Income, and quite similar results were obtained. Judgments of gift size showed a diverging fan pattern, as if a multiplying model was operative, just as in the American studies. This shows that the present Indian subjects are able to use the same integration rule as the Americans. Accordingly, their present use of an equal-weight averaging rule would not seem to be attributable to the subjects' inappropriate understanding of the task, their attempt to simplify it (Anderson & Butzin, 1974), or their inability to use the same integration rule as the Americans. It can thus be said that the present discrepancy from the American results reflects a difference in cultural outlook.

Does the difference between the equal-weight averaging rule and the multiplying rule point to any important difference in cultural outlook between India and America? For attributions of achievement, given information about motivation and ability, the following argument suggests at least one important social difference. According to the equal-weight averaging model, effort or trying will be equally effective with persons of low or high ability. Within a cultural system of attributions that obeys the averaging model, therefore, it will appear that a person of lower ability can gain as much by trying as a person of higher ability.

The multiplying rule, in contrast, implies that effort or trying will be more effective with persons of higher ability. Within a cultural system of attributions that obeys the multiplying rule, therefore, it will appear that persons of lower ability have less to gain by trying.

This multiplying rule for attributions may be closer to the actual behavior processes, since just such a rule is found in most current behavioral theories of motivation, such as those of Hull and Tolman (see Anderson, 1974a, p. 29). Given the achievement orientation of American culture, it is no surprise to see a multiplying rule emerge in the attributional judgments of American subjects, especially college students.

From a social-cultural view, however, the present equal-weight averaging model for attribution of performance appears to portray a more egalitarian outlook. Perhaps college students in India believe that each person, regardless of native ability, has equal opportunity to improve his or her lot. This egalitarian attitude is a healthy sign for progress in India, for Indian people have been described as high in dependency (Murphy, 1953; Winter, 1969).

Both the present data on the equal-weight averaging rule and the previous American results on the multiplying rule, however, rest on a very narrow cultural sample, namely, college students. If the cognitive algebra of achievement attribution is to be useful in

cross-cultural research, then it is necessary to study many other cultural strata, both in India and in the United States.

To the authors, information integration theory is impressive because it provides a useful framework for cross-cultural studies of social perception and cognition. An important characteristic of the integration rules, it should be emphasized, is that they are concerned with *patterns* of responses, not with the numerical value of single responses. This characteristic is vital for cross-cultural comparisons because it bypasses uncertain assumptions that specific stimuli have comparable values across cultures. Indians and Americans undoubtedly have different value systems, but they can still be compared in terms of the pattern of their responses to a factorial set of stimuli. Indeed, the integration rule can provide a base and frame for measurement of meaning–value even for the individual within his or her culture (Anderson, 1976, 1977). The present results illustrate the potential power of this approach.

Because of its direct concern with the problems of multiple causation, integration theory is well suited for cross-cultural comparison. Of special interest and value is that the integration-theoretical approach can help shift the direction of cross-cultural research from a purely correlational approach (Whiting, 1968) to a causative, experimental base (Singh, Bohra, & Dalal, in press; Singh, Sidana, & Saluja, 1978a, 1978b; Singh, Sidana, & Srivastava 1978).

Reference Notes

1. Anderson, N. H. *Methods for studying information integration* (Tech. Rep. CHIP 43). La Jolla, Calif.: University of California, San Diego, Center for Human Information Processing, June 1974.
2. Singh, R. *Multiplying versus differential-weight averaging as an integration rule in attribution of gift size.* Unpublished manuscript, Indian Institute of Technology, Kanpur, India, Department of Humanities and Social Sciences, 1978.

References

Anderson, N. H. Integration theory and attitude change. *Psychological Review*, 1971, *78*, 171–206.

Anderson, N. H. Cognitive algebra. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 7). New York: Academic Press, 1974. (a)

Anderson, N. H. Information integration theory: A brief survey. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 2). San Francisco: Freeman, 1974. (b)

Anderson, N. H. How functional measurement can yield validated interval scales of mental quantities. *Journal of Applied Psychology*, 1976, *61*, 677–692.

Anderson, N. H. Note on functional measurement and data analysis. *Perception & Psychophysics*, 1977, *21*, 201–215.

Anderson, N. H., & Butzin, C. A. Performance = Motivation × Ability: An integration-theoretical analysis. *Journal of Personality and Social Psychology*, 1974, *30*, 598–604.

Graesser, C. C., & Anderson, N. H. Cognitive algebra of equation: Gift Size = Generosity × Income. *Journal of Experimental Psychology*, 1974, *103*, 692–699.

Heider, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

Kun, A., Parsons, J. E., & Ruble, D. N. Development of integration processes using ability and effort information to predict outcome. *Developmental Psychology*, 1974, *10*, 721–732.

Lampel, A. K., & Anderson, N. H. Combining visual and verbal information in an impression formation task. *Journal of Personality and Social Psychology*, 1968, *9*, 1–6.

Murphy, L. B. Roots of tolerance and tension in Indian child development. In G. Murphy (Ed.), *In the minds of men.* New York: Basic Books, 1953.

Shanteau, J. C., & Anderson, N. H. Test of a conflict model for preference judgment. *Journal of Mathematical Psychology*, 1969, *6*, 312–325.

Singh, R., Bohra, K. A., & Dalal, A. K. Favorableness of leadership situations studied with information integration theory. *European Journal of Social Psychology*, in press.

Singh, R., Sidana, U. R., & Saluja, S. K. Playgroup attractiveness studied with information integration theory. *Journal of Experimental Child Psychology*, 1978, *25*, 429–436. (a)

Singh, R., Sidana, U. R., & Saluja, S. K. Integration theory applied to judgment of personal happiness by children. *Journal of Social Psychology*, 1978, *105*, 27–31. (b)

Singh, R., Sidana, U. R., & Srivastava, P. Averaging processes in children's judgments of happiness. *Journal of Social Psychology*, 1978, *104*, 123–132.

Whiting, J. W. M. Methods and problems in cross-cultural research. In G. Lindzey and E. Aronson (Eds.), *The handbook of social psychology* (Vol. 2). Reading, Mass.: Addison-Wesley, 1968.

Winter, S. K. Case studies. In D. C. McClelland and D. G. Winter (Eds.), *Motivating economic achievement.* New York: Basic Books, 1969.

Received July 28, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 8, 1352-1363

The Effects of Reward Contingency and Performance Feedback on Intrinsic Motivation

Judith M. Harackiewicz
Harvard University

In this study the effects of reward contingency and positive performance feedback on subsequent intrinsic motivation for an enjoyable task were examined. It was hypothesized that rewards contingent only upon participation in the task would produce decrements in motivation (the overjustification effect) but that rewards contingent on performance quality would produce even greater decrements in intrinsic motivation relative to control conditions of no reward. It was also hypothesized that positive performance feedback would enhance intrinsic motivation and that this effect would be independent of reward effects. High school students were offered performance or task-contingent rewards, or no reward, for doing hidden-figures puzzles. Subjects offered performance-contingent rewards all received positive feedback concerning performance and half the subjects in task-contingent and no-reward conditions received the same positive feedback. Performance-contingent rewards were found to undermine intrinsic motivation more than task-contingent ones, which produced decrements relative to control conditions of no reward, supporting Deci's control model. Positive feedback enhanced intrinsic motivation and this effect was independent of reward effects. A recall measure indicated that subjects receiving performance-contingent rewards remembered fewer performance-irrelevant details about the task, suggesting that rewards may affect the process of task involvement as well as its motivational outcomes.

In a review of the literature concerned with the effects of extrinsic rewards on intrinsic motivation, Condry (1977) concludes that in certain contexts, subsequent interest in a task may be reduced by the imposition of task-extrinsic rewards. One aspect of context that apparently mediates this undermining effect is reward contingency. Rewards have been promised to subjects contingent only upon participation or for completion of the task (task contingent),[1] or contingent upon some level of performance of the task (performance contingent).

The undermining effects of task-contingent rewards have been well documented (Anderson, Manoogian, & Reznick, 1976; Lepper, Greene, & Nisbett, 1973). Rewards promised to subjects for engaging in an activity and implying no performance demands have consistently produced what Lepper and Greene (1976) termed the "overjustification effect,"

The author thanks John Kihlstrom, Bella DePaulo, Reid Hastie, and Robert Rosenthal for their help with the planning of this research and the data analysis, and Mark Lepper, Michael Ross, and Shelley Taylor for helpful comments on earlier versions of this article. I am grateful to Harold Rosen, Mary Harackiewicz, and Cliff Thurber for their cooperation and assistance in collecting the data. I would also like to thank Al Hirschfeld and the Margo Feiden Galleries for permission to use his drawings in this research.

Requests for reprints should be sent to Judith Harackiewicz, Department of Psychology and Social Relations, Harvard University, 33 Kirkland Street, Cambridge, Massachusetts 02138.

[1] These rewards have frequently been referred to as *noncontingent* rewards. Although not explicitly contingent on some level of performance, they are usually at least implicitly contingent on compliance with the experimenter or participation in the activity and are perhaps more appropriately considered task contingent.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3708-1352$00.75

or a decrease in subsequent intrinsic motivation. Their most recent theoretical statement specifies the context in which this effect should occur:

> The overjustification studies provide a demonstration of the negative effects on intrinsic motivation of making salient to a person the instrumentality of his behavior when rewards are presented which do not result in the acquisition of new skills or convey salient information to him concerning his ability at a task. (Lepper & Greene, 1976, p. 33)

Because the promise of a performance-contingent reward implies that information concerning task ability will be provided, specific predictions for the effects of performance-contingent rewards cannot be directly derived from the overjustification hypothesis. Other investigators, however, have attempted to examine the effects of such rewards and to relate them to overjustification effects. Two predictions for performance-contingent rewards have been advanced. Karniol and Ross (1977) suggested that performance-contingent rewards may maintain or increase intrinsic motivation to the extent that they convey information about effective performance at a task, because attainment of such rewards should provide tangible evidence of personal effectiveness. Deci (1975), on the other hand, predicted that a performance-contingent reward should decrease intrinsic motivation even more than a task-contingent one because a reward is perceived to be more controlling when it is contingent on some level of performance.

Karniol and Ross and Deci agreed that rewards have controlling and informational aspects, and both suggested that the more salient of the two will initiate either changes in perceptions of instrumentality of behavior (when controlling aspects of the reward are salient) or changes in feelings of competence and self-determination (when informational aspects are salient). Their predictions for the effects of performance-contingent rewards differed because of differences in the ascribed salience of controlling versus informational aspects of these rewards.

Previous Research

Deci (1971, 1972a) consistently found that subjects receiving performance-contingent rewards demonstrate less subsequent intrinsic motivation than unrewarded controls. Pritchard, Campbell, and Campbell (1977) and Pittman, Cooper, and Smith (1977) found that performance-contingent rewards undermined intrinsic motivation relative to control conditions of no reward. Pinder (1976) found that performance-contingent rewards produced greater decrements in intrinsic motivation than did task-contingent ones. Although these results all seem to disconfirm the hypothesis of Karniol and Ross, the studies do not represent adequate tests of it, because performance-contingent rewards are predicted to maintain interest only when subjects perceive that they have done well at the task. It is not clear that this criterion was met in these studies.

Two studies have compared task-contingent and performance-contingent rewards with a control condition of no reward in the same design. Greene and Lepper (1974) found that performance and task-contingent rewards produced equal decrements in intrinsic motivation relative to control conditions. Karniol and Ross (1977) found that when children received positive feedback concerning their performance, task-contingent rewards decreased subsequent interest relative to performance-contingent rewards and no rewards. Performance-contingent rewards did maintain interest, but they did not increase it above the control level as had been predicted. Karniol and Ross suggested that because the performance standard was established in the reward manipulation at the outset, children knew that they had performed well as soon as they finished the task. The information inherent in the receipt of the reward was thus redundant with that provided in the instructions. The researchers predicted that when the receipt of a reward provides additional rather than redundant information concerning competence, it should increase intrinsic motivation above the level of control subjects.

Methodological Issues

One aspect of reward contingency that has not been directly examined is the possible confounding of material reward with positive

feedback about performance quality. The promise of a performance-contingent reward implies that a subject's performance will be evaluated and that feedback will be available. Deci (1972b) and Anderson et al. (1976) demonstrated that positive feedback, in the absence of any reward manipulation, increases subsequent intrinsic motivation. Although it is difficult to separate reward manipulation from feedback in the case of performance-contingent rewards, it is possible to examine the effects of reward and feedback independently in task-contingent and control conditions because feedback is not implied by a task-contingent reward manipulation.

Several measures of intrinsic motivation have been used in the overjustification literature, ranging from behavioral measures (task persistence in the absence of rewards) to volunteering measures (willingness to return for another session) to self-report measures of task interest and enjoyment. Conflicting results of previous studies may be due to different operational definitions of intrinsic motivation.

Condry (1977) draws a distinction between dependent measures that focus on the products of the rewarded session, for example, subsequent interest, and measures that focus on the process of task involvement, and he suggests that the promise of extrinsic rewards may affect the manner in which subjects engage in a task, as well as subsequent motivation. Products and processes may be differentially affected by rewards, however, and should be considered separately.

Objectives

The primary aim of the present study was to examine the effects of performance- and task-contingent rewards on intrinsic motivation, compared to conditions of no reward. To accomplish this, two kinds of performance-contingent rewards were created. A performance-contingent reward with performance norms supplied at the outset of the task would convey no additional information concerning competence beyond that inherent in the norms. A performance-contingent reward with no prior norms established would provide nonredundant information concerning competence. These two performance-contingent rewards were compared with task-contingent rewards and control conditions of no reward. In addition, the effects of positive performance feedback, relative to no feedback, were examined in conditions of task-contingent and no reward in order that results might be comparable to both those of Karniol and Ross and those of other studies that have demonstrated the undermining effects of task-contingent rewards in the absence of positive performance feedback.

To facilitate comparison with other studies, several behavioral, volunteering, and self-report product measures of intrinsic motivation were utilized. Interest was measured at several points in time in order to examine the endurance of reward effects. The task in this study was constructed in such a way that attention to aspects of it not relevant to performance could be measured with an incidental recall measure, intended to assess the process of intrinsic motivation.

It was predicted that positive feedback would enhance intrinsic motivation, that task-contingent rewards would undermine it, and that these two effects would be independent. The effects of performance-contingent rewards were of particular interest because of conflicting predictions: an enhancement of motivation mediated by perceived self-efficacy (Karniol & Ross, 1977) or a detrimental effect because of the perceived controlling nature of the reward (Deci, 1975). Controlling the informational value of the reward permits closer analysis of performance-contingent rewards. Karniol and Ross predict that the more informational performance-contingent reward with no prior norms established should produce the most enhancement of motivation. Although Deci does not make specific predictions regarding informational value, a speculative control prediction was generated. Because the two rewards are both contingent on performance quality, equal decrements might be predicted, but this negative effect could be differentially mitigated by the two types of positive feedback. When norms are supplied at the outset, the feedback is received earlier in the motivational process and might have a

greater positive effect than would feedback received after completion of the task. According to this analysis, the control hypothesis would produce a larger decrement in interest in the performance-contingent condition where norms are not supplied. It was hypothesized that the control prediction would be supported in this study because of the greater similarity of the subject population (high school students) to Deci's college students than to the children used as subjects by Karniol and Ross.

Method

Overview

In this study six groups of subjects received various combinations of reward and positive feedback with respect to a challenging but enjoyable task. Interest in this task was assessed both before and after the experimental manipulations. The six experimental conditions are summarized in Table 1. The no reward – no feedback (NRNF) and task-contingent reward – no feedback (TCRNF) cells replicate the conditions employed in overjustification studies, whereas the no reward – positive feedback (NRPF), task-contingent reward – positive feedback (TCRPF) and performance-contingent reward (norms supplied) – positive feedback (PCRNSPF) cells are conceptual replications of the central features of the Karniol and Ross (1977) design. Subjects in the PCRNSPF condition were offered a reward contingent on performing above average and were informed of the performance standard before beginning the task, so that receipt of the reward only confirmed that they had performed above average. In the performance-contingent reward (no norms) – positive feedback (PCRNNPF) condition, however, the reward itself was the only source of information available to the subject concerning level of performance. Thus the reward received in the PCRNNPF condition was more informational than that received in the PCRNSPF condition, although subjects in both conditions received the same feedback concerning performance. The standard of performance was set at a relatively low level so that all subjects would be eligible for positive feedback.

Subjects

The participants in this study were 64 male and 29 female students in a suburban Massachusetts high school, with a mean age of 16.4 years. The subjects were tested during their English class.

Procedure

The study was conducted in three phases. The first and third sessions were group administered in the classroom, whereas subjects were tested individually in a separate room for the second session. This experimental session began 1 month after the first session and continued for 2 weeks, and the third session took place 1 month after the end of the experimental sessions.

The experimental activity in this study was a series of hidden-figure puzzles—cartoon-style drawings by Al Hirschfeld, in which the name *Nina* is hidden sev-

Table 1
Schematic Design of Experiment and Contrast Weights for Main Hypothesis

Condition	Reward	Feedback	Contrast weight
No reward–no feedback (NRNF)	None mentioned	None	1
No reward–positive feedback (NRPF)	None mentioned	Positive[a]	3
Task-contingent reward–no feedback (TCRNF)	Offered for doing puzzles	None	−1
Task-contingent reward–positive feedback (TGRPF)	Offered for doing puzzles	Positive[a]	1
Performance-contingent reward (norms supplied)–positive feedback (PCRNSPF)	Offered for meeting performance criterion, norms supplied before task	Positive[a]	−1
Performance-contingent reward (no norms)–positive feedback (PCRNNPF)	Offered for meeting performance criterion, no norms supplied	Positive[a]	−3

[a] Positive feedback was "We've found that the average student usually finds 4 Ninas, so you did better than the average high school student on these puzzles."

Figure 1. A Nina puzzle (Hirschfeld's drawing for the play *Anna Christie,* © courtesy of the Margo Feiden Galleries, New York, New York).

eral times. These drawings appear in the Arts and Leisure section of the Sunday *New York Times* and are usually caricatures of actors and scenes from current films and plays. Each drawing has a title and a caption describing the scene and naming the actors depicted. The drawings used in this study contained three to seven hidden Ninas. An example of a Nina puzzle (in which there are seven hidden Ninas) is shown in Figure 1.

Pretest. The students were given booklets containing three kinds of puzzles. The word puzzles consisted of several anagrams; the face and body recognition puzzle was the booklet form of the Profile of Nonverbal Sensitivity (PONS; Rosenthal, Hall, DiMatteo, Rogers, & Archer, 1979), which consists of 40 pictures of a face or body and two brief situation labels from which the subject selects the more appropriate description of the picture; there were also three Nina puzzles. The order in which students did the three puzzle sets was counterbalanced. After doing each puzzle set, students completed a puzzle questionnaire that consisted of 13 7-point bipolar

rating scales (e.g., good–bad, easy–hard, boring–exciting) and eight questions concerning the subject's involvement with the puzzle (e.g., "How much did you enjoy the puzzle?"), also answered on 7-point scales. After completing all three types of puzzles, the students rank ordered them according to enjoyment.

Experimental session. The experimenter informed each subject that he or she would be doing more Nina puzzles and that the instructions were tape recorded. After turning on the tape recorder she left the room in order to remain blind to the subject's condition, which was randomly determined. All subjects were instructed to find as many Ninas as possible within the 2 minutes of inspection time allowed per drawing. Subjects in the two task-contingent reward conditions (TCRNF and TCRPF) were promised a reward for doing the puzzles, whereas subjects in the two performance-contingent reward conditions (PCRNSPF and PCRNNPF) were promised a reward for exceeding the performance of the average high school student. Those in the performance-contingent reward (norms supplied) – positive feedback group were then informed of the exact performance standard (detecting four or more Ninas in the three puzzles combined). All subjects who had been promised a reward (TCRNF, TCRPF, PCRNSPF, and PCRNNPF) were told that it consisted of two felt-tip pens and a notebook, in their choice of colors, and were reminded of the conditions of the reward. Finally, the subjects completed an identification form and answered a series of questions concerning their understanding of the directions, the reward contingencies, their current mood state, expected puzzle difficulty and performance level, and attitude toward the reward.

The experimenter returned to the room and administered the three Nina puzzles. Subjects in the NRPF, TCRPF, PCRNSPF, and PCRNNPF groups then received positive performance feedback ("We've found that the average high school student usually finds four Ninas, so you did better than the average student on these puzzles"), and those who were promised a reward (TCRNF, TCRPF, PCRNSPF, and PCRNNPF) indicated their color choices on forms that again reminded them of the conditions under which the reward had been promised. The experimenter was blind to this latter information. While the experimenter filed away the puzzles and order forms, the subject was permitted to read a current popular magazine, work on another Nina puzzle, or simply sit quietly. The experimenter unobtrusively observed the subject's behavior and recorded the amount of time spent on the puzzle, if any.

After 2 minutes the subject completed several questionnaires, including a Nina puzzle questionnaire identical to the one administered at the pretest. A volunteering questionnaire stated that although there were no definite plans for future sessions, the researcher wanted a general idea of student willingness to return. It was composed of four 7-point rating scales asking the subject how willing he or she would be to return for a further, unrewarded session of Nina puzzles if it were scheduled during their English class, during a study hall, after school, or on a weekend. The extra puzzles request form informed subjects that copies of different Nina puzzles would be available after the conclusion of the research and permitted students to request up to four additional puzzles. The recall questionnaire was composed of a list of actors, in which were embedded the names of those depicted in the Nina puzzles from the experimental session, and a similar list of play and film titles; students were asked to circle any names that they remembered seeing in the puzzles.

If the subject had not previously been offered a reward, he or she was informed at this time that there was a gift for having participated in psychological research. These subjects then indicated their color choice for the pens and notebook on an order form that reminded the subject of the rationale for the gift.

Finally, all subjects were assured that the experiment would be explained to them in detail when the rewards/gifts were delivered at its conclusion. They were also cautioned not to discuss the details of the experiment with other students.

Posttest. One month after all students had participated in the experimental session, their teacher distributed a letter from the experimenter saying that she would be visiting their classroom soon with deliveries and to answer questions about the research project. At this time they completed yet a third Nina puzzle questionnaire.

The experimenter visited the classroom a few days after the students received the letter, distributed the rewards/gifts (and extra puzzles if requested), and debriefed the students.

Dependent Measures

In addition to various manipulation checks concerned with the informational effects of positive feedback and postmanipulation motivation, several dependent variables were of interest in this study. Performance was indexed by the number of Ninas actually found on the three puzzles administered during the experimental session; all subjects found at least 4 Ninas, and no subject found all 15 contained in the drawings. Initial levels of intrinsic motivation were indexed by a pretest enjoyment scale derived from the pretest Nina puzzle questionnaire in the manner described below.

There were five outcome measures of intrinsic motivation taken after the experimental manipulation: (a) *experimental enjoyment*—a scale identical to the pretest enjoyment scale, but administered after the experimental task; (b) *time*—the amount of time that the subject spent looking at the extra Nina puzzle during the free-choice period; (c) *volunteering*—score on the volunteering questionnaire; (d) *extra puzzles*—the number of puzzles requested by the subject on the extra puzzles request form; and (e) *posttest enjoyment*—a scale identical to the pretest and experimental enjoyment scales, but administered dur-

ing the posttest.[2] A further measure was a process variable intended to index the level of task orientation during the experimental session: (f) *incidental recall*—the number of names and titles correctly recognized on the recall questionnaire, corrected for overcircling by subtracting one half the number of incorrect answers.

Results

Pretest Analyses

A principal-components factor analysis with varimax rotations was performed on the 21 items of the Nina puzzle questionnaire, and the first factor extracted accounted for 50.6% of the variance. Seven items had factor loadings higher than .70 and were considered to contribute to the definition of the factor: Enjoyable, Good, Interesting, Exciting, Enjoyment of Puzzle, Interest in Puzzle, and Worthwhile. This factor was interpreted independently by two judges to measure task enjoyment, and scores on each of the 7 items were combined using unit weighting to form an enjoyment scale. Pretest enjoyment was used as a covariate in subsequent analyses of covariance.

Since the overjustification effect is predicted to occur only when intrinsic motivation for an activity is initially high (Lepper & Greene, 1976) it is necessary to establish that the Nina puzzles were in fact intrinsically interesting. The mean score for pretest enjoyment was 33.98 ($SD = 9.20$); this was significantly higher than 25, the midpoint of the enjoyment scale, $t(92) = 9.41$, $p < .001$, $d = 1.95$, suggesting that subjects rated the Nina puzzles more enjoyable than not. Examination of the rank ordering of puzzles indicated that 50% of the subjects rated the Nina puzzles most enjoyable of the three. This observed frequency was significantly higher than the chance expectation of 33.33%, $\chi^2(1) = 10.88$, $p < .002$, $\phi = .34$, offering additional evidence for the high level of initial enjoyment of the Nina puzzles.

Manipulation Checks

The effectiveness of the positive feedback manipulation was examined in two Treatment × Sex (6 × 2) analyses of covariance[3] covarying pretest enjoyment on the perceived performance item ("How well do you think you did on these puzzles?") from the Nina puzzle questionnaire for the experimental session and posttest. A planned contrast tested whether subjects in the four conditions receiving positive feedback (NRPF, TCRPF, PCRNNPF, and PCRNSPF) were higher in perceived performance than subjects in the two no-feedback conditions (NRNF, TCRNF). In this contrast the first four groups were assigned weights of +1 and the latter two groups were given weights of −2. Subjects who received positive feedback showed higher levels of perceived performance both immediately following the experimental session, $F(1, 86) = 7.08$, $p < .01$, $\eta = .28$, and at the posttest 1 month later, $F(1, 80) = 7.03$, $p < .01$, $\eta = .28$.

A Treatment × Sex (6 × 2) analysis of variance was performed on the seven postmanipulation scales. No significant effects were found for alert, comfortable, relaxed, motivated, excited, anticipated puzzle difficulty, or expected performance level, suggesting that there were no treatment differences in motivation as measured by these scales after the students heard the instruction tape but before they actually did the puzzles.

In order to examine possible differences between the two groups that received performance-contingent rewards, post hoc protected t tests were used to compare the PCRNSPF and PCRNNPF groups on the seven postmanipulation scales and on experimental and posttest perceived performance. PCRNSPF subjects were

---

[2] Six subjects graduated from high school before the administration of the posttest, so posttest enjoyment scores were not obtained for them.

[3] Sex of subject was included as a factor in all analyses of variance, although it was not relevant to the main hypotheses of this study. A significant sex effect for posttest perceived performance was found, $F(1, 74) = 4.33$, $p < .05$, $\eta = .24$, indicating that males thought they had performed better than females. A significant sex effect for expected performance indicated that males thought they would do better than females on the puzzles, $F(1, 80) = 6.78$, $p < .05$, $\eta = .27$. A significant main effect for sex on recall, $F(1, 80) = 6.62$, $p < .05$, $\eta = .28$, showed that males were higher in amount of incidental recall than females, suggesting that the females may have been more task oriented. The sex differences found in this study were unanticipated, and interpretation of them would require replication.

significantly higher than PCRNNPF subjects on expected performance level ("How well do you think you will do on these puzzles?"), alert, and posttest perceived performance (PCRNSPF Ms = 5.13, 4.93, and 5.84 vs. PCRNNPF Ms = 4.16, 3.69, and 4.72 on 7-point scales), ts(87) = 2.36, 2.46, and 2.32, respectively, all ps < .05. The two groups did not differ significantly on comfortable, relaxed, motivated, excited, anticipated puzzle difficulty, or experimental perceived performance. This pattern of significant differences suggests that the groups differed primarily in expectations for and perceptions of performance level.

Intercorrelation of the Dependent Measures

Table 2 presents the correlation matrix for the six measures of intrinsic motivation (pretest, experimental, and posttest enjoyment, time, volunteering, and extra puzzles), performance, recall, and experimental and posttest perceived performance. The average intercorrelation of the six measures was .51 (p < .001), and each individual correlation was significant at the p < .001 level, indicating that all of the operationalizations of intrinsic motivation were positively interrelated.

Performance was positively related to pretest and experimental enjoyment, $r(91)$ = .20 and .22; p < .05, but was not significantly related to any other measures of intrinsic motivation. Performance (actual) did correlate positively with experimental perceived performance, $r(91)$ = .33, p < .001. Recall did not correlate significantly with any measures of intrinsic motivation or performance.

Because the four intrinsic motivation variables measured during the experimental session were positively related, a principal-components factor analysis was performed on them in order to create one measure of motivation for the experimental session. Volunteering, extra puzzles, time, and experimental enjoyment had loadings of .57, .75, .60, and .73, respectively, on the first factor extracted, and standardized factor scores were computed for the resulting intrinsic motivation measure, which was used as the primary dependent variable in the main analyses.

Effects of Reward and Feedback on Intrinsic Motivation

A Treatment × Sex (6 × 2) analysis of covariance with pretest enjoyment as a covariate was performed on intrinsic motivation

Table 2
Intercorrelations of Dependent Measures

Measure	1	2	3	4	5	6	7	8	9
1. Pretest enjoyment									
2. Experimental enjoyment	.69***								
3. Posttest enjoyment	.55***	.69***							
4. Time	.33***	.44***	.42***						
5. Volunteering	.37***	.50***	.40***	.33***					
6. Extra puzzles	.46***	.54***	.41**	.53***	.43***				
7. Performance	.20*	.22*	.14	.10	.09	.03			
8. Recall	.05	.04	.01	.14	.06	.17	.11		
9. Experimental perceived performance	.04	.11	.00	.02	.02	−.07	.30**	−.13	
10. Posttest perceived performance	.12	.19	.28**	.29**	−.02	.03	.33**	.08	.51***

* p < .05. ** p < .01. *** p < .001.

Table 3
Reward Condition Means for the Dependent Variables

Measure	Reward condition/contrast weight					
	NRPF/ +3	NRNF/ +1	TCRPF/ +1	TCRNF/ −1	PCRNSPF/ −1	PCRNNPF/ −3
Intrinsic motivation	1.05	−.16	.39	−.64	.07	−.96
Posttest enjoyment	38.47	33.09	35.36	31.96	35.97	29.96
Recall	2.56	2.44	3.13	2.30	1.49	1.22
Performance	8.87	7.81	8.56	8.27	9.00	8.62

Note. For the intrinsic motivation measure, standardized factor scores were created from volunteering, extra puzzles, time, and experimental enjoyment variables. NRPF = no reward – positive feedback. NRNF = no reward – no feedback. TCRPF = task-contingent reward – positive feedback. TCRNF = task-contingent reward – no feedback. PCRNSPF = performance-contingent reward (norms supplied) – positive feedback. PCRNNPF = performance-contingent reward (no norms) – positive feedback.

and posttest enjoyment. The effect of the pretest enjoyment covariate was highly significant for both intrinsic motivation, $F(1, 86) = 56.98$, $p < .001$, $\eta = .63$, and posttest enjoyment, $F(1, 80) = 39.99$, $p < .001$, $\eta = .58$, indicating that, as expected, intrinsic interest following the experimental manipulation was linearly related to initial task interest. No effects involving sex were significant for either dependent measure.

A planned contrast was used to test the main effect of treatment. Weights for the contrast were operationalized according to the control hypothesis as follows: no reward – no feedback—+1; no reward – positive feedback—+3; task-contingent reward – no feedback—−1; task-contingent reward – positive feedback—+1; performance-contingent reward (norms supplied) – postive feedback —−1; and performance-contingent reward (no norms) – positive feedback—−3. These equal interval weights are based on the assumption that the predicted effects are all equal in magnitude. This contrast tests the prediction that task-contingent rewards will produce negative effects relative to control conditions of no reward, that positive feedback will produce positive effects relative to conditions of no feedback, and that these two effects do not interact. This contrast also tests the prediction that performance-contingent rewards will decrease intrinsic motivation more than task-contingent ones and that performance-contingent rewards with no norms supplied will decrease intrinsic motivation more than performance-contingent rewards with norms supplied.

Means for the two dependent measures of intrinsic motivation are presented in Table 3. As predicted, the treatment planned contrast was highly significant for both measures: intrinsic motivation, $F(1, 86) = 12.72$, $p < .001$, $\eta = .36$, and posttest enjoyment, $F(1, 80) = 9.58$, $p < .005$, $\eta = .33$.[4] Thus intrinsic motivation was enhanced by positive feedback and undermined by rewards, particularly performance-contingent ones. This treatment planned contrast tests the overall patterning of the six cell means with respect to the predictions specified by the weights. Because this overall test incorporates several specific predictions, and since particular cell comparisons are relevant with respect to comparison with other studies, several additional nonorthogonal but theoretically relevant planned comparisons were computed for the primary dependent measure, intrinsic motivation.

In the four cells in which task-contingent reward or no reward was crossed with positive feedback or no feedback (TCRNF, TCRPF, NRNF, and NRPF), task-contingent rewards were found to reduce intrinsic motivation relative to control conditions of no reward, replicating the overjustification effect: TCRNF and

[4] This planned contrast was also significant for each of the four dependent measures combined to form the intrinsic motivation measure, in four separate analyses of variance on time, extra puzzles, volunteering, and experimental enjoyment.

TCRPF versus NRNF and NRPF, $t(87) = 1.78$, $p < .05$, one-tailed,[5] $d = .38$; and positive feedback increased interest relative to no feedback: NRPF and TCRPF versus NRNF and TCRNF, $t(87) = 2.74$, $p < .05$, $d = .59$. The interaction between task-contingent reward and feedback was not significant. Comparison of the two effect sizes (*d*s) suggests that the positive feedback effect was stronger than the overjustification effect.

Planned comparisons were also computed for the four cells with positive feedback (NRPF, TCRPF, PCRNSPF, and PCRNNPF), which represent the conceptual replication and extension of the Karniol and Ross (1977) design. The TCRPF and PCRNSPF means were not significantly different, but both were significantly lower than the control group: TCRPF and PCRNSPF versus NRPF, $t(87) = 2.01$, $p < .05$, $d = .43$. Thus the Karniol and Ross results were not replicated: Performance-contingent rewards, as well as task-contingent ones, reduced interest relative to control conditions. Furthermore, the Karniol and Ross prediction for the more informational PCRNNPF group was not supported; the PCRNNPF group was significantly lower than the PCRNSPF group in interest, $t(87) = 1.76$, $p < .05$, one-tailed, $d = .37$. Additionally, although the PCRNSPF group was not significantly lower than the TCRPF group, PCRNNPF rewards were significantly more undermining than TCRPF ones, $t(87) = 2.34$, $p < .05$, $d = .50$, and the two performance-contingent rewards considered together were significantly more undermining than the task-contingent one: PCRNNPF and PCRNSPF versus TCRPF, $t(87) = 1.67$, $p < .05$, one-tailed, $d = .36$, replicating the results of Deci (1972b).

Effects of Treatment on Performance and Recall

Treatment × Sex (6 × 2) analyses of covariance covarying pretest enjoyment were performed on incidental recall and performance. There were no significant effects for performance.

It was predicted that subjects offered performance-contingent rewards would be more task oriented and would recall fewer irrelevant details than subjects offered no rewards or task-contingent ones. A planned contrast comparing the two performance-contingent conditions (PCRNSPF and PCRNNPF: −2) with the other four conditions (NRNF, NRPF, TCRNF, and TCRPF: +1) was significant, $F(1, 86) = 8.95$, $p < .005$, $\eta = .32$, indicating that recall was lower in the two performance-contingent conditions, as predicted. The effect of the pretest enjoyment covariate was not significant ($F < 1$). Amount of recall, then, did not vary with the subjects' initial enjoyment of the puzzles. No other effects, except those involving sex, were significant for this measure.

Discussion

The control hypothesis received very strong support in this study. Performance-contingent rewards, particularly informational ones, were found to undermine subsequent intrinsic motivation more than task-contingent ones, which produced decrements in intrinsic motivation relative to control conditions of no reward. This pattern of results was found across five different operational definitions of intrinsic motivation, some of which were behavioral and some of which were self-report, attitudinal measures. Most striking, perhaps, is the fact that this pattern of results was still found 1 month after the experimental manipulation had taken place, on the task enjoyment measure.

The findings of this study lend support to the generalizability of the overjustification effect, which was replicated using a different subject population (high school students as opposed to college sophomores or preschoolers), a different experimental activity, and a different reward (notebook and pens, as opposed to the cash usually offered to adult subjects). The enhancing effect of positive feedback was also replicated, and this effect was found to be independent of the overjustification effect. This finding has important implications for the comparison of previous studies and for the identification of appropriate control groups.

---

[5] All reported *p* levels are two-tailed unless otherwise stated.

Since the receipt of a performance-contingent reward means that a subject has received both a reward *and* positive feedback concerning performance, the appropriate control group for examination of the effects of rewards may be a group that receives the same positive feedback but no reward. The comparison of performance-contingent and task-contingent rewards requires that the task-contingent group receive the same positive feedback, assuming that the feedback does not interact with the offer of a reward. On the other hand, it can be argued that the feedback inherent in a performance-contingent reward is an integral part of the reward itself, in which case it would not be unreasonable to compare performance-contingent rewards with task-contingent rewards or controls without positive feedback added. Choice of an appropriate comparison group is therefore somewhat arbitrary but of course affects the nature of the conclusions drawn.

Both kinds of performance-contingent rewards were found to reduce intrinsic motivation relative to control conditions of no reward. The informational reward (no norms) decreased intrinsic motivation more than the noninformational (norms supplied) reward, as predicted. Although the reward itself contained more information, subjects in both conditions received the same amount of information about their performance. In the noninformational condition, subjects knew how many Ninas they had to find and knew that they were doing well as soon as they found four Ninas. Subjects in the informational condition, on the other hand, had no idea about their performance level until after they finished the puzzles and received their reward. Intrinsic motivation may have been lower in the informational condition because the reward was more controlling (students had to perform as well as possible for the whole session, whereas those in the noninformational condition were less controlled because of the lesser demand of having to find four Ninas), or intrinsic motivation may have been higher in the noninformational condition because of the greater amount of feedback available to subjects during the actual process of doing the puzzles. Positive feedback may be more effective if it is supplied (or implied by preestablished norms) during the process of task involvement (so that subjects feel that they are doing well) rather than after task completion.

The informational hypothesis of Karniol and Ross (1977) received no support in this study. The means for intrinsic motivation fell in almost the exact opposite ordering of the pattern predicted by their hypothesis. It was originally predicted that their hypothesis would not be supported in this study because of differences in subject populations, but post-experimental conversations with students revealed that the nature of the experimental activity might be an important determinant of the relative salience of controlling versus informational aspects of performance-contingent rewards. Although students enjoyed doing the Nina puzzles, their enjoyment did not seem related to competence concerns. Enjoyment of the task was only slightly related to actual performance level, $r(91) = .22$, $p < .05$. Neither performance nor perceived performance was significantly related to any of the other measures of intrinsic motivation, suggesting that actual or perceived competence is not always an integral part of intrinsic motivation.

Intrinsically motivated activities may be engaged in for their own sake (e.g., drawing with magic markers and reading novels) or for the feelings of competence they produce (e.g., games, puzzles, and sports). This is, of course, an unrealistic dichotomy, because most activities involve both kinds of intrinsic satisfaction, but the extent to which an activity is relatively more competence relevant or fun relevant may determine whether informational or controlling aspects of a performance-contingent reward will be more salient to an individual.

The results of the analyses of recall differences are particularly interesting. The fact that recall was unrelated to the measures of intrinsic motivation and unrelated to performance, yet was affected by reward condition, suggests that this measure tapped an important dimension of task involvement that has not previously been examined in overjustification studies. The finding that contingently rewarded subjects did not remember

as much about the puzzles as did other subjects, but did not find any more Ninas either, suggests that they may have been using a more answer-oriented strategy that did not prove to be effective in terms of performance on these particular puzzles. Exactly what sort of processing strategy is gauged by the recall measure employed in this study is not clear, but the results of the analyses lend support to Condry's (1977) contention that more attention needs to be directed toward an exploration and understanding of the process of task involvement.

References

Anderson, R., Manoogian, S., & Reznick, J. The undermining and enhancing of intrinsic motivation in preschool children. *Journal of Personality and Social Psychology*, 1976, *34*, 915–922.

Condry, J. Enemies of exploration: Self-initiated versus other-initiated learning. *Journal of Personality and Social Psychology*, 1977, *35*, 459–477.

Deci, E. Effects of externally mediated rewards on intrinsic motivation. *Journal of Personality and Social Psychology*, 1971, *18*, 105–111.

Deci, E. The Effects of contingent and noncontingent rewards and controls on intrinsic motivation. *Organizational Behavior and Human Performance*, 1972, *8*, 217–229. (a)

Deci, E. Intrinsic motivation, extrinsic reinforcement and inequity. *Journal of Personality and Social Psychology*, 1972, *22*, 113–120. (b)

Deci, E. *Intrinsic motivation*. New York: Plenum, 1975.

Greene, D., & Lepper, M. Effects of extrinsic rewards on children's subsequent intrinsic interest. *Child Development*, 1974, *45*, 1141–1145.

Karniol, R., & Ross, M. The effects of performance relevant and performance irrelevant rewards on children's intrinsic motivation. *Child Development*, 1977, *48*, 482–487.

Lepper, M., & Greene, D. On understanding "overjustification": A reply to Reiss and Sushinsky. *Journal of Personality and Social Psychology*, 1976, *33*, 25–35.

Lepper, M., Greene, D., & Nisbett, R. Undermining children's interest with extrinsic rewards: A test of the "overjustification hypothesis." *Journal of Personality and Social Psychology*, 1973, *28*, 129–137.

Pinder, C. Additivity versus nonadditivity of intrinsic and extrinsic incentives: Implications for work motivation, performance, and attitudes. *Journal of Applied Psychology*, 1976, *61*, 693–700.

Pittman, T., Cooper, E., & Smith, T. Attribution of causality and the overjustification effect. *Personality and Social Psychology Bulletin*, 1977, *3*, 280–283.

Pritchard, R. D., Campbell, K., & Campbell, P. Effects of extrinsic financial rewards on intrinsic motivation. *Journal of Applied Psychology*, 1977, *62*, 9–15.

Rosenthal, R., Hall, J., DiMatteo, M. R., Rogers, P. L., & Archer, D. L. *Sensitivity to nonverbal communication: the PONS Test*. Baltimore, Md.: Johns Hopkins University Press, 1979.

Received July 31, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 8, 1364-1376

Variables That Moderate the Attitude–Behavior Relation: Results of a Longitudinal Survey

Andrew R. Davidson
University of Washington

James J. Jaccard
Purdue University

The following factors were hypothesized to moderate the attitude–behavior relation: (a) the behavioral sequence that must be successfully completed prior to the occurrence of the behavior, (b) the time interval between the measurement of attitudes and behavior, (c) attitude change, (d) the respondent's educational level, and (e) the degree of correspondence between attitudinal and behavioral variables. The behaviors investigated were having a child and using oral contraceptives. A stratified random sample of 244 married women in a midwestern urban area was studied during a three-wave, 2-year longitudinal study. Selection of attitudinal and belief measures was guided by the Fishbein model of behavioral intentions. Consistent with the hypotheses, the relations between behavior and both intention and the model's attitudinal and normative components were substantially attenuated by (a) events in the behavioral sequence not under the volitional control of the actor, (b) an increase in the time interval between the measurement of attitudes and behavior from 1 to 2 years, and (c) changes in the model's attitudinal and normative components during the first year. The respondent's educational level did not affect attitude–behavior consistency. Finally, the attitude–behavior correlation increased significantly as the degree of correspondence between the two variables increased.

Wicker (1969) concluded his comprehensive review of the attitude–behavior relation with the suggestion, "It is considerably more likely that attitudes will be unrelated or only slightly related to overt behaviors than that attitudes will be closely related to actions" (p. 76). Rather than signaling a decrease in research on this topic, pessimistic reviews by Wicker and others (e.g., Deutscher, 1966, 1969; Ehrlich, 1969; McGuire, 1969) appear to have prompted a renewal of interest in the relationship between attitude and action. One encouraging line of inquiry has focused on methodological refinements in the attitudinal and behavioral measures. By assessing both variables at corresponding levels of specificity, that is, measuring attitude toward the act for the prediction of a specific behavior or measuring a global attitude toward an object for the prediction of a multiple-act behavioral criterion, a reasonable degree of predictive accuracy can be obtained (Ajzen & Fishbein, 1973, 1977; Fishbein & Ajzen, 1974; Heberlein & Black, 1976; Weigel & Newman, 1976; Weigel, Vernon, & Tognacci, 1974; Weinstein, 1972; Wicker & Pomazal, 1971).

A second line of inquiry, encouraged by Kelman (1974), has attempted to specify the personal and situational variables influencing the extent to which people act in accord with their stated attitudes and beliefs. Individual difference variables that have been found to moderate attitude–behavior correspondence include the degree of affective-cognitive consistency (Norman, 1975), the tendency to ascribe responsibility to the self (Schwartz, 1973), and the degree of self-monitoring

This research was supported by Research Grant HD 07024-01,02,03 from the National Institute of Child Health and Human Development. We are grateful to Marilyn Cohen for her substantial aid both in supervising the interviewers and data coders and in locating respondents who moved.

Requests for reprints should be sent to Andrew R. Davidson, Department of Psychology, NI-25, University of Washington, Seattle, Washington 98195.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3708-1364$00.75

(Snyder & Tanke, 1976). Similarly, a number of situational variables appear to moderate the attitude–behavior relation. For example, Warner and DeFleur (1969) found that the public versus private nature of the behavior interacts with attitude in influencing attitude–behavior consistency, and Snyder and Swann (1976) demonstrated that situations that increase the relevance of salient attitudes as guides to actions increase the correspondence between attitude and action. A recent series of studies (Fazio & Zanna, 1978; Fazio, Zanna, & Cooper, 1978; Regan & Fazio, 1977) provide support for the notion that attitudes formed through direct behavioral experience with the attitude object are more predictive of later behavior toward that object than are attitudes based upon indirect, nonbehavioral experience. In summary, several personal and situational factors have been found to affect the attitude–behavior relation. As Regan and Fazio (1977) commented, "The question facing researchers is, therefore, no longer *whether* an individual's attitudes can be used to predict his overt behavior, but *when*" (p. 30, italics in the original).

The present research continues the study of factors that moderate the degree of attitude–behavior consistency and investigates the following variables:

1. Sequence of prior events. As Fishbein and Jaccard (1973) have observed, the occurrence of a behavior under investigation is frequently dependent upon the successful completion of a sequence of prior events. For example, in order to have a child a woman must (a) have intercourse, (b) not use birth control, (c) conceive, and (d) not have a spontaneous or induced abortion. While some of the sequences are potentially under the control of the actor (e.g., not using birth control), others are not (e.g., conception). If performance of the final behavior requires the successful completion of sequences not under the volitional control of the actor, it is hypothesized that attitude toward the behavior will provide a better estimate of the likelihood that the actor will initiate the sequence than that the actor will complete the sequence and perform the behavior in question.

2. Attitude change. In a recent review of the attitude–behavior literature, Schuman and Johnson (1976) noted that little attention has been paid to the possibility that attitude change may occur between the measurement of attitude and behavior. They commented, "The extent to which A–B correlations are reduced because of real changes in underlying attitude needs to be identified as far as possible" (p. 178). Such identification is impossible in most studies because attitude is assessed at only one time. In the present research, attitude was measured twice prior to the performance of behavior in order to detect attenuation attributable to attitude change. Respondents whose attitudes remain stable are hypothesized to exhibit greater attitude–behavior consistency than respondents who experience attitude change.

3. Time interval. A third variable hypothesized to influence the degree of relation between attitude and behavior is the time interval between the measurement of attitude and the performance of behavior. As Fishbein and Jaccard (1973) noted, it is probably the processes occurring during that time interval —most notably exposure to new information —and not the passage of time per se that moderate the relation. This hypothesis is conceptually similar to the one above; the longer the time interval between the measurement of attitude and behavior, the higher the probability of exposure to new information and, in turn, attitude change.

Three previous investigations have reported data relevant to the influence of temporal instability on attitude–behavior consistency. Norman (1975) regressed behavior (volunteering to be a subject) on attitudes toward acting as a subject, assessed both 3 and 6 weeks prior to the measurement of behavior. For each of three attitudinal measures, the second administration was more highly correlated with behavior than the first, but the differences were not statistically significant. It has been suggested (Schwartz, 1978) that for this behavior the 3-week interval between attitude assessments was too brief for an unambiguous effect to be detected. In studies of voting behavior, Kelley and Mirer (1974)

have reported that for some subgroups the time interval between the measurement of voting behavior and attitudes can moderate the attitude–behavior correlation. Analyses of data from the 28% of the subjects judged most likely to change their minds found that a substantial proportion of variance in errors of prediction was accounted for by the number of days that intervened between the interview and the election. In the final and most recent investigation, Schwartz (1978) regressed stated willingness to tutor blind children on perceived moral obligation to tutor blind children, measured both 3 and 6 months prior to the criterion measure. The correlation with the criterion was significantly higher over the shorter time interval. However, as Schwartz noted, it is uncertain whether the results obtained with perceived moral obligation can be generalized to more traditional attitude scales measuring evaluation. In sum, although specific aspects of each of these studies—interval between attitude assessments, nature of the subsample investigated, nature of the predictive variable—have precluded a clear demonstration of the influence of time interval on the consistency between behavior and the evaluation of the behavior, these studies together provide a sound basis for positing the present hypothesis.

4. Respondent's education. Consistent with formulations proposed by Peak (1955), Rosenberg (1956), and Fishbein (1963), attitude toward a behavior is viewed as the sum of one's beliefs about the consequences of performing the behavior multiplied by the evaluation of those consequences. Such formulations, with their emphasis on cognitive-affective consistency, are frequently criticized as having greater validity for college students and intellectuals than for the majority of the population (cf. Bem, 1970). If such critiques are correct, education should interact with attitudes in the prediction of behavior, and degree of attitude–behavior consistency should correlate positively with the respondent's educational level. The findings of Rokeach (1973) also lead to the hypothesis that greater consistency will be observed for respondents with more education. In a national sample of American adults, the instrumental value *logical* defined as *consistent, rational* was more positively valued by highly educated respondents than by those with less education.

5. Correspondence between attitudinal and behavioral variables. As noted earlier, if attitude and behavior are both measured at a similar level of specificity, a reasonable relation is generally observed. In an extension of the specificity hypothesis, Ajzen and Fishbein (1977) maintain that attitudinal and behavioral variables are defined by four elements: an *action*, the *target* the action is directed toward, the *context* in which the action is to be performed, and the *time* at which the action occurs. They argue that the correlation between attitude and behavior is determined in part by the degree of correspondence or match between the elements comprising the two variables. They classified past studies in terms of the correspondence between attitude and behavior on two of the four elements, target and action. Consistently significant attitude–behavior correlations were obtained only if there was high correspondence between target and action elements. Their investigation, however, examined the moderating effects of target and action correspondence by comparing attitude–behavior correlations obtained from different studies. In the present research, the effects of target, action and, in addition, time correspondence on attitude–behavior consistency are tested within one study. It is hypothesized that as the degree of correspondence increases, the obtained attitude–behavior correlation will increase.

Attitude Model

The selection of attitudinal measures was guided by the Fishbein (1967) model of behavioral intention. The model is predicated on the assumption that most human behavior of concern to social scientists is to some degree volitional in nature and hence guided by the behavioral intent of the individual. Algebraically, the model is expressed as follows:

$$B \sim BI = [\sum_{i=1}^{n} B_i E_i] W_1 + [\sum_{i=1}^{m} NB_i MC_i] W_2, \quad (1)$$

where B = overt behavior, BI = the behavioral intention to perform the behavior, B_i = the belief (perceived probability) that performing the behavior will lead to consequence X_i, E_i = the evaluation of X_i, NB_i = the perceived expectation of Referent i, MC_i = the motivation to comply with Referent i, n = the number of salient consequences, m = the number of salient normative beliefs, and W_1 and W_2 = empirically determined regression weights.

As Peak (1955), Rosenberg (1956), and Fishbein (1963) have theorized, the ΣB_iE_i component is viewed as an index of the person's attitude toward performing the behavior (*Aact*). Fishbein and Ajzen (1975) maintain that the second predictive component (ΣNB_i-MC_i) assesses the influence of the social environment or the general subjective norm (*SN*) on behavior. Empirical support for the equivalence of ΣB_iE_i and *Aact* and of ΣNB_i-MC_i and *SN* is discussed in Fishbein and Ajzen (1975).

According to the model, any variable other than the attitudinal and normative components can influence *BI* only indirectly. Social, demographic, and personality characteristics of the respondent can affect intentions only if they influence ΣB_iE_i or ΣNB_iMC_i or their relative weights. Thus if this framework is correct, such variables as education should not have a direct effect on the magnitude of the relation between the components of the model and intention, and hence behavior.

Fishbein and Ajzen (1975) argue that the relation between *BI* and *B* should be very strong, provided that intention and behavior are measured at correspondent levels of specificity and that nothing intervenes to alter intention. Whenever a strong intention–behavior relation is observed, the behavior in question should also be predictable from the attitudinal and normative components. Assuming a significant *BI–B* correlation, it is hypothesized that behavior can be predicted from a linear combination of the attitudinal and normative components. In addition, it is hypothesized that the variables moderating the intention–behavior relation will also moderate the relation between attitudes, subjective norms, and behavior.

Behavioral Domain

The behaviors chosen for study were childbearing and contraceptive use. While social scientists have met with success in documenting fertility differentials on the basis of social and demographic variables, (cf. Rindfuss & Sweet, 1977) there has been a marked lack of success in explaining fertility behavior with psychological constructs. Two major fertility surveys, the Indianapolis study (Whelpton & Kiser, 1946–1958) and the Princeton study (Westoff, Potter, Sagi, & Mishler, 1961) failed to find any noteworthy relationships between psychological variables (primarily personality measures) and fertility and family planning behaviors.

Recently, research findings have indicated that a few psychological models that can be categorized as expectancy models are of some utility in the prediction of fertility decision making (see, for example, Beach, Townes, Campbell, & Keating, 1976; Davidson & Jaccard, 1975; Werner, Middlestadt-Carter, & Crawford, 1975). Working with the Fishbein model, Davidson and Jaccard (1975) investigated attitudes concerning three family planning behaviors in a sample of married women. In support of the model, ΣB_iE_i and ΣNB_iMC_i explained an average of 60% of the variance in behavioral intentions.

The data reported in Davidson and Jaccard (1975) were obtained during the first wave of a three-wave longitudinal survey. The sample was reinterviewed both 1 year and 2 years after the initial interview. The present study examines the relation between intentions, at the first interview, of having a child during the next 2 years and using oral contraceptives during the next 2 years and the corresponding self-reports of behavior during the 2-year period.

Method

Sample

Respondents in the initial survey were 270 white married women, age 18–38 years, residing in a mid-

western urban area. The sample was stratified by three levels of socioeconomic status and two levels of religious affiliation (Catholic and Protestant). The sampling procedure was designed to select randomly 45 women for each cell of the 2×3 design. For a detailed description of the sampling strategy, see Davidson and Jaccard (1975).

Loss to Follow-Up

The respondents were reinterviewed both 1 and 2 years after the initial survey. Of the original 270 respondents, 244 completed all three interviews.[1] Approximately one half of the respondents lost to follow-up refused to be reinterviewed and the other half could not be located. Loss to follow-up did not significantly vary among the six cells of the sampling design.

Interviews

Approximately 1 hour was required for the respondent to complete each questionnaire. Although the interviews were self-administered, the interviewer remained in the house while the questionnaire was being completed to answer any questions the respondent might have. The respondent received $10 for each interview.

Measurement Procedures

To insure that relevant beliefs and referents for the attitudinal and normative components of the model were included in the precoded questionnaire, initial elicitation interviews were conducted with an independent sample of 55 women. These women were selected from the same population as the women in the longitudinal sample. The interviews identified four referents (e.g., husband, close friends) and nine beliefs (e.g., the effect of having a child on the husband–wife relationship, the woman's freedom to pursue other activities, the family budget, etc.) salient to the population. A list of the referents and beliefs included in the questionnaire is presented in Davidson and Jaccard (1976).

All predictive components were measured on 7-point adjective scales. Examples of the measures are

1. Behavioral intention (*BI*) (*I intend to have a child within the next 2 years*) was measured on a *likely–unlikely* scale.

2. Belief about the act (B_i) (*For me, having a child within the next two years would make my marriage stronger*) was measured on a *likely–unlikely* scale.

3. Evaluation of the consequence (E_i) (*Making my marriage stronger would be . . .*) was measured on a *good–bad* scale.

4. Attitude toward the act (*Aact*) (*For me, having a child within the next two years would be . . .*) was obtained by summing the responses to three evaluative scales: *good–bad, nice–awful, pleasant–unpleasant.*

5. Normative belief (NB_i) (*My parents think I should have a child within the next two years*) was measured on a *likely–unlikely* scale.

6. Motivation to comply (MC_i) was measured on a scale anchored by *generally speaking, I want to do what Referent* i *thinks I should do* and *generally speaking, I do not want to do what Referent* i *thinks I should do.*

7. General subjective norm (*SN*) (*People who are important to me and whose opinions I value think I should have a child within the next two years*) was measured on a *likely–unlikely* scale.

Scales assessing evaluations of consequences and attitude toward the act were scored −3 (bad) to +3 (good). All other scales were scored from 1 (unlikely; not motivated to comply) to 7 (likely; motivated to comply).[2] ΣB_iE_i and ΣNB_iMC_i were obtained by multiplying the score on each belief statement by the score on the corresponding evaluation or motivation to comply and then summing these products for all beliefs. In addition to the above items, the questionnaire assessed a number of fertility-related attitudes and contained measures of education, scored from 1 (sixth-grade education or less) to 8 (PhD or other professional degree), and other demographic variables.

A dichotomous criterion variable was created for each behavior. For childbearing, women reporting a birth between the first and third interviews were assigned a score of 1, and the remainder of respondents were assigned a score of 0. In families reporting a birth, the child was typically observed by the interviewer. For oral contraceptive use, women reporting that they used oral contraceptives at any time during the 2-year interval were scored 1, and women never using the pill were scored 0.

The ΣB_iE_i and ΣNB_iMC_i measures were obtained for only one of the behaviors—having a child. Due to time constraints, only the direct measures of the attitudinal and normative components, *Aact* and *SN*, respectively, were obtained for the second behavior—using oral contraceptives. In the first of the three

---

[1] Due to missing data and failure to follow instructions, the sample sizes in the following analyses range from 242 to 244.

[2] In our previous studies all scales were scored −3 to +3. However, we now recognize that with the exception of the evaluation measures, the remainder of scales are assessing variables that are theoretically unipolar (e.g., subjective probability). Hence, the unipolar scales are scored 1 to 7 in the present analysis. A comparison of variables formed from both the present and previous scoring rules indicated that the two were substantially intercorrelated; for ΣB_iE_i, $r = .93$, and for ΣNB_iMC_i, $r = .76$. For a discussion of unipolar versus bipolar scoring of attitude and belief measures, see Fishbein and Ajzen (1975, pp. 82–86).

interviews, questions referred to performing each behavior during the next 2 years. In the second interview, the questionnaire was adjusted to query the respondents about having a child and using oral contraceptives during the next year.

Results

The first half of this section focuses on intention–behavior consistency and the factors hypothesized to influence the degree of consistency. The second half investigates the prediction of behavior from the model's attitudinal and normative components and examines the variables that moderate that relation.

Prediction of Behavior from Behavioral Intention

The intention–behavior relation. For each behavior a point-biserial correlation was computed between behavioral intention, measured during the initial interview, and behavior during the 2-year period. The *BI–B* correlation was .526 ($p < .01$) for childbearing and .678 ($p < .01$) for contraceptive use. The relative magnitude of these two correlations is probably influenced by the extent to which each behavior is under volitional control. While contraceptive use–nonuse is to a large degree under the control of the actor, childbearing is determined by an interaction of biological and volitional events.

The sequence of prior events. As discussed above, in order to bear a child a woman must have intercourse without using contraception, conceive, and not have a spontaneous or induced abortion. It was hypothesized that intention would provide a better estimate of the probability that the person would initiate the sequence than of the likelihood that the person would complete the sequence and perform the behavior in question (bear a child). Women were classified as initiating the sequence if, in the final survey, they responded that by having intercourse while not using contraception they had been attempting for at least 9 months to become pregnant or if they actually gave birth during the 2-year period. The point-biserial correlation of intention with birth or attempted conception was .615 ($p < .01$). As hypothesized, this correlation was significantly larger than the correlation of intention with birth, .615 versus .526, $t(241) = 4.45$, $p < .01$.

The time interval between the assessment of intention and behavior. It was hypothesized that as the time interval between the measurement of intention and behavior increases, the *BI–B* correlation decreases. This hypothesis was tested by comparing the *BI–B* correlation for the 2-year period (Year 1 to Year 3) with the correlation for the 1-year period (Year 2 to Year 3). In support of the hypothesis, for both behaviors, the correlation for the 1-year period was significantly higher than for the 2-year period (oral contraceptive use: .853 versus .678, $z = 5.21$, $p < .01$; birth or attempted conception: .829 vs. .615, $z = 4.95$, $p < .01$).[3]

Change in intention. The effect of time interval on intention–behavior consistency can be attributed to processes occurring during the time interval (e.g., exposure to new information and attitude change). To demonstrate directly the effect of change on the *BI–B* correlation it was necessary to identify respondents who changed intentions between the first two waves of the three-wave survey. However, as Kiesler (1977) has shown, a respondent might change her intention to make it consistent with her behavior during the prior year. For example, if a woman did not intend to have a child at the first interview but became pregnant during the first year of the study, she would probably indicate an intention to have a child at the second interview. Such a case would more accurately be classified as a missed prediction rather than as a change in intention. Hence, only respondents not performing the behavior (becoming pregnant, bearing a child, using oral contracep-

[3] As these correlations are themselves intercorrelated, it is incorrect to use the traditional Fisher's z transformation (Hayes, 1963, pp. 529–532) to test for the significance of differences between them. Rather, the correlations were appropriately compared using a test initially suggested by Pearson and Filon and reported in Peters and Van Voorhis (1940, p. 185).

tives) during the first year were eligible for the classification of change in intention.

Eligible respondents were classified into one of three groups based on their responses at the first and second interviews to the relevant intention items. They were classified as intenders if they indicated it was slightly, moderately, or very likely that they would perform the behavior. Respondents indicating that performance of the behavior was slightly, moderately, or very unlikely were classified as nonintenders. Women checking the midpoint of the scale (neither likely nor unlikely) were classified as uncertain. Any eligible respondent who changed classification from Time 1 to Time 2 was identified as having changed her intention.

Based on this procedure, 30 women were identified as having changed their childbearing intentions, and the remaining 214 women were classified as nonchangers. For oral contraceptive intentions, 36 women were classified as changers and 206 as nonchangers. Each *BI–B* correlation was then computed separately for the change and nonchange respondents. For both childbearing and contraceptive use, the intention–behavior correlation was substantially lower for women who changed intentions than for nonchangers (birth or attempted conception: $-.015$ vs. .742, $z = 4.65$, $p < .01$; oral contraceptive use: $-.397$ vs. .880, $z = 9.57$, $p < .01$). Thus, in support of the hypothesis, change in intention attenuated the intention–behavior relation over the 2-year period.

Respondent's educational level. It was hypothesized that the respondent's educational level would be positively related to the degree of intention–behavior consistency. Hence intention should interact with education in the prediction of behavior. To test this hypothesis, behavior was regressed hierarchically (Allison, 1977; Cohen, 1978) against intention, education and the multiplicative interaction term of Intention × Education. For neither behavior did the inclusion of the interaction term lead to a significant increment in R^2; both Fs < 1. Therefore the hypothesis that educational level moderates the intention–behavior relation was not supported.

In summary, there was a reasonably strong correlation between intentions and behavior during the subsequent 2 years. Consistent with the hypotheses, however, this relation was attenuated by (a) events in the behavioral sequence not under the volitional control of the actor, (b) changes in intention, and (c) the time interval between the assessment of intention and behavior. The respondent's education level did not moderate the intention–behavior relation.

A demonstrated relation between what people say they are going to do (intention) and what they actually do (behavior) is probably of greater applied than theoretical significance. Once strong intention–behavior consistency has been observed, however, the relation between the attitudinal and normative components of the model and behavior is of considerable theoretical interest. The subsequent analyses examine this relation.

Prediction of Behavior From the Model's Attitudinal and Normative Components

The regression of behavior on attitudes and norms. For both contraceptive use and childbearing, the dichotomous index of behavior during the 2-year period was regressed on attitudes and norms, measured during the initial interview. As indicated in Rows 1 and 3 of Table 1, for each behavior both of the model's components received significant regression weights in the prediction of behavior. For the prediction of birth the multiple correlation was .508 ($p < .01$), and R^2 adjusted for shrinkage was .252. For the prediction of oral contraceptive use, the multiple correlation was .606 ($p < .01$) and R^2 adjusted for shrinkage was .362.

The sequence of prior events. It was hypothesized that events in the behavioral sequence not under the volitional control of the actor would attenuate the prediction of behavior from the model's components. The data presented in row 2 of Table 1 support this hypothesis. The multiple correlation with birth or attempted conception is significantly higher than the correlation with birth, .595 versus .508; $t(241) = 4.33$, $p < .01$.

Table 1
Regression of Fertility and Contraceptive Behavior on Attitudes and Norms

Behavior	Standardized regression coefficients		R	R^2	R^2 adjusted for shrinkage
	$\sum B_iE_i$	$\sum NB_iMC_i$			
Birth[a,b]	.187	.361	.508	.258	.252
Birth or attempted conception[a,c]	.203	.438	.595	.354	.349
	Aact	*SN*			
Use of oral contraceptives[d,e]	.392	.268	.606	.367	.362

Note. All regression coefficients and multiple correlations are significant ($p < .01$). Abbreviations are as follows: $\sum B_iE_i$ = one's beliefs about the consequences of performing the behavior multiplied by the evaluation of those consequences; $\sum NB_iMC_i$ = one's normative beliefs about performing the behavior multiplied by one's motivation to comply with those perceived norms; *Aact* = one's attitude toward performing the behavior; *SN* = one's subjective norm (generalized normative belief) about performing the behavior.
[a] $N = 244$.
[b] Coded: birth during the 2-year period = 1; no birth = 0.
[c] Coded: birth or attempted conception during the 2-year period = 1; no birth or attempted conception = 0.
[d] $N = 242$.
[e] Coded: use of oral contraceptives during the 2-year period = 1; nonuse = 0.

Time interval between the assessment of the model's predictive components and behavior. For both behaviors, the multiple correlation with the model's predictive components was higher for the 1-year period (Year 2 to Year 3) than for the 2-year period (Year 1 to Year 3). However, for birth or attempted conception the difference between the *R*s failed to achieve the traditional significance level (oral contraceptive use: .704 vs. .606, $z = 2.48$, $p < .01$; birth or attempted conception: .649 vs. .595, $z = 1.39$, $p < .09$).

Attitudinal and normative change. As previously noted, the effect of time interval on attitude–behavior consistency is attributable, in part, to attitude and belief change occurring during the interval. To demonstrate the effect of change on the correlation between behavior and the model's components, it was necessary to identify respondents whose attitudes or norms changed from Interview 1 to Interview 2. The procedure used here is similar to the one described above for assessing change in *BI*. First, to differentiate norm and attitude change from post hoc justification of prior behavior, only respondents not performing the behavior (becoming pregnant, bearing a child, using oral contraceptives) during the first year were eligible for the classification of change in attitudes or norms. Next, separately for each behavior and each of the model's two predictive components, eligible respondents who changed from Time 1 to Time 2 were identified as changers, and the remaining respondents were classified as nonchangers. The correlation between the relevant component and behavior was then computed separately in each group. Consistent with the hypothesis, lower correlations were expected in the change than in the nonchange group.

Attitude toward the act of using oral contraceptives (*Aact*) was assessed by summing the responses to three evaluative scales, each scored +3 to −3. A score of 0 was categorized as a neutral attitude, a score greater than 0 was a positive attitude, and less than 0 constituted a negative attitude. Sixty-three eligible respondents changed attitude categories between Interview 1 and Interview 2. As predicted, the *Aact–B* correlation was substantially higher in the nonchange than in the change group (.733 vs. −.238; $z = 7.88$, $p < .01$).

Responses to the general subjective norm concerning the use of oral contraceptives

(*SN*) were also divided into three categories: uncertain (a score of 4, the midpoint of the scale), positive norm (a score greater than 4), and negative norm (a score less than 4). Seventy-four eligible respondents changed categories from Interview 1 to Interview 2. In support of the hypothesis, the *SN–B* correlation was significantly higher in the nonchange than in the change group (.682 vs. −.039; $z = 6.14$, $p < .01$).

For childbearing behavior, the midpoints of the scales for the $\Sigma B_i E_i$ and $\Sigma NB_i MC_i$ components had to be empirically identified. Using data from the first interview, the mean scores on $\Sigma B_i E_i$ and $\Sigma NB_i MC_i$ were determined separately for those intending and not intending to have a child. The score midway between the means was designated the neutral point on the scale. In the interval between Interview 1 and Interview 2, 34 eligible respondents moved across the neutral point of the $\Sigma B_i E_i$ measure, and 25 eligible respondents moved across the neutral point of the $\Sigma NB_i MC_i$ measure. Consistent with the prediction, the correlation of both $\Sigma B_i E_i$ and $\Sigma NB_i MC_i$ with birth or attempted conception was higher in the nonchange than in the change group ($\Sigma B_i E_i$: .580 vs. .011, $z = 3.38$, $p < .01$; $\Sigma NB_i MC_i$: .648 vs. −.061, $z = 3.71$).

Respondent's educational level. It was hypothesized that the respondent's educational level would interact with attitudes and norms in the prediction of behavior. To test this hypothesis, behavior was regressed hierarchically against the model's attitudinal and normative components, education, and the multiplicative interactions of Education × Each Component. For neither behavior did the inclusion of the interaction terms lead to a significant increment in R^2; both Fs < 1. Hence, the hypothesis was not supported.

In summary, there was a reasonably strong correlation between the model's attitudinal and normative components and behavior during the subsequent 2 years. Similar to the results obtained with behavioral intention, the prediction of behavior from attitudes and norms was attenuated by (a) events in the behavioral sequence not under the volitional control of the actor, (b) the time interval between the measurement of the model's predictive components and behavior (for childbearing there was only a trend in the hypothesized direction), and (c) changes in attitudes and norms. The respondent's educational level did not moderate the strength of the relation.

Table 2
Correlations Between Selected Attitudinal Variables and Behavior

Attitudinal variable	*r*
Use of birth control pills during the 2-year period[a]	
Attitude toward birth control	.083
Attitude toward birth control pills	.323*
Attitude toward using birth control pills	.525*
Attitude toward using birth control pills during the next 2 years	.572*
Birth or attempted conception during the 2-year period[b]	
Attitude toward children	−.007
Attitude toward having children	.187*
Attitude toward having a child in the next 2 years	.535*

[a] $N = 244$.
[b] $N = 242$.
* $p < .01$.

Correspondence Between Attitudinal and Behavioral Variables

It was hypothesized that the degree of correspondence between the elements of the attitudinal and behavioral variables would be positively related to attitude–behavior consistency. The present analysis investigated the effect of three elements on the attitude–behavior correlation: *action*, the *target* the action is directed toward, and the *time* at which the action occurs. As indicated in Table 2, four attitudinal variables, measured during the first interview, were correlated with the dichotomous behavioral measure of birth control pill use during the 2-year period. The attitudinal measures systematically varied with regard to the number of elements they had in correspondence with the behavioral measure, as follows: zero elements in correspondence (attitude toward birth control, $r = .083$, *ns*); one element in correspondence—target (attitude toward birth control pills, $r = .323$, $p < .01$); two elements in correspondence—target

and action—(attitude toward using birth control pills, $r = .525$, $p < .01$); three elements in correspondence—target, action, and time—(attitude toward using birth control pills during the next 2 years, $r = .572$, $p < .01$). In support of the hypothesis, for each successive increase in correspondence between attitude and behavior variables there was a significant increase (ranging from $p < .01$ to $p < .06$) in the magnitude of the attitude–behavior correlation.

Following a similar pattern, the dichotomous behavioral measure of birth or attempted conception during the 2-year period correlated −.007 (*ns*) with attitude toward children, .187 ($p < .01$) with attitude toward having children, and .535 ($p < .01$) with attitude toward having a child in the next 2 years. For each successive increase in correspondence there was a significant increase (all ps $< .01$) in the attitude–behavior correlation.

Discussion

The primary purpose of the present research was to investigate five factors hypothesized to influence the extent to which people act in accord with their stated attitudes. The investigation of the first of these factors, the sequence of prior events, directed attention to the behavioral variable. The majority of previous attitude–behavior research has focused exclusively on problems associated with attitudinal variables during attempts to account for low attitude–behavior correlations. However, the performance of a behavior is often dependent upon the successful completion of a series of behavioral events. As the present research has demonstrated, if all of these steps are not under the volitional control of the actor, intention will provide a better estimate of the probability that the actor will initiate the sequence than of the likelihood that the actor will complete the sequence and perform the behavior in question. These results suggest that future studies of attitude–behavior consistency should seek to identify and quantify the behavioral sequences that must be completed prior to the performance of the behavior of interest. Analyses can then isolate the sequences at which serious disruptions occur between attitudes and actions. From an applied perspective, such knowledge would be of considerable use in the selection of behavior-change strategies. This approach would lead to the identification of situations in which the most appropriate strategy is *not* to change attitudes but rather to institute policy changes making it easier for people to act in accord with their existing attitudes.

The second and third factors, attitude change and the time interval between the measurement of attitude and behavior, focused on instability in the attitude and belief variables. As the time interval decreased from 2 years to 1 year, there was a 58.3% increase in the amount of variance in oral contraceptive use accounted for by *BI*. The corresponding increase for childbearing was 81.7%. Substantial improvements in the prediction of behavior were also achieved when analyses were limited to the sample of respondents who did not exhibit attitude change. The magnitude of these effects directs attention to the problem of using static models in the analysis of a dynamic process. When stating attitudes and intentions about future behaviors, a respondent is probably assuming either that the environment will not change or that it will change in an anticipated manner. Unexpected changes (e.g., loss of employment, exposure to new information) can lead to a change in attitude and, if such changes are not monitored, to an apparent lack of consistency between attitudes and behavior. Future research should attempt to specify the personal and situational variables influencing attitudinal stability. From the studies reviewed above, some of the factors that might initially be hypothesized to contribute to stability include the extent to which the attitude was formed through direct behavioral experience with the attitude object (Fazio & Zanna, 1978; Regan & Fazio, 1977) and the individual difference variables of self-monitoring (Snyder & Tanke, 1976) and tendency to ascribe responsibility to the self (Schwartz, 1973).

The fourth factor hypothesized to moderate the attitude–behavior relation was the respondent's educational level. It was hypothesized that as completed years of schooling

increased, so would the magnitude of the attitude–behavior correlation. However, an interaction between education and attitude–behavior consistency was not observed. Hence, these results lend no support to the contention that consistency is especially prevalent among the highly educated. It appears unlikely that nonsupport for this hypothesis is attributable to lack of variance on the education variable; years of education ranged from 8 to 16. In addition, attitudinal models similar to the present one have provided good prediction of behavioral intentions and behavior in cross-cultural samples with considerably lower educational levels than the present sample (Davidson, Jaccard, Triandis, Morales, & Diaz-Guerrero, 1976; Davidson & Thomson, 1980). It should be noted that the present finding provides indirect support for the theoretical framework of Fishbein. As noted above, if the framework is correct, social, demographic, and personality characteristics of the respondent should not have a direct effect on the magnitude of the relation between the components of the model and intentions, and hence behavior.

The fifth factor hypothesized to moderate attitude–behavior consistency is the degree of correspondence between the elements of the attitudinal and behavioral variables. As the attitudinal measures moved along a continuum of correspondence in terms of the number of elements they had in agreement with the behavioral measure (from zero to three elements), their correlations with behavior significantly increased. The present results concerning the impact of action and target correspondence are consistent with the conclusions of Ajzen and Fishbein (1977). In addition, the findings highlight the impact of a third element, not previously investigated, the time during which the action occurs. To obtain reasonable predictions of behavior from attitudinal variables, it appears important to ensure correspondence in target, action, and time elements.

The findings from the present study also provide support for the predictive validity of the components of the model of Fishbein. Even prior to the adjustments made for attitude change and the sequence of events, both intention and the attitudinal and normative components measured at the first interview provided reasonable correlations with behavior during the subsequent 2 years (the lowest validity coefficient was .508).

A weakness in the present research is that the data on oral contraceptive use were obtained solely through self-reports. Hence it is not possible to determine if the respondents misrepresented their behavior in order to make it consistent with their previously stated attitudes and intentions. It appears doubtful that there was extensive misrepresentation, because the findings concerning oral contraceptive use were very similar to those for childbearing in terms of both the percentage of subjects correctly classified and the variables found to moderate the degree of attitude–behavior consistency. There was also a reliance on self-report data concerning the attempt to become pregnant. Respondents might have falsely indicated that they had been unsuccessfully trying for at least 9 months to become pregnant in an attempt to appear consistent with previously stated intentions. However, this rival hypothesis requires that the respondents admit something that is socially undesirable about themselves for the purpose of appearing consistent.

In conclusion, this research indicates that the contraceptive and fertility behavior of married women is quite predictable from behavioral intentions, attitudes, and normative beliefs. However, the degree of the relation is attenuated by events in the behavioral sequence not under the volitional control of the actor, the time interval between the measurement of behavior and the model's predictive components, changes in attitudes and beliefs, and the degree of correspondence between the attitudinal and behavioral variables.

References

Ajzen, I., & Fishbein, M. Attitudinal and normative variables as predictors of specific behaviors. *Journal of Personality and Social Psychology*, 1973, *27*, 41–57.

Ajzen, I., & Fishbein, M. Attitude–behavior relations: A theoretical analysis and review of empirical research. *Psychological Bulletin*, 1977, *84*, 888–919.

Allison, D. Testing for interaction in multiple regression. *American Journal of Sociology*, 1977, *83*, 144–153.

Beach, L., Townes, B., Campbell, F., & Keating, G. Developing and testing a decision aid for birth planning decisions. *Organizational Behavior and Human Performance*, 1976, *15*, 99–116.

Bem, D. *Beliefs, attitudes, and human affairs*. Belmont, Calif.: Brooks/Cole, 1970.

Cohen, J. Partialed products are interactions: Partialed powers are curve components. *Psychological Bulletin*, 1978, *85*, 858–866.

Davidson, A., & Jaccard, J. Population psychology: A new look at an old problem. *Journal of Personality and Social Psychology*, 1975, *31*, 1073–1082.

Davidson, A., and Jaccard, J. Determinants of fertility intentions. In S. Newman & V. Thompson (Eds.), *Population psychology: Research and educational issues*. Washington, D.C.: U.S. Government Printing Office, 1976.

Davidson, A., Jaccard, J., Triandis, H., Morales, M. L., & Diaz-Guerrero, R. Cross-cultural model testing: Toward a solution of the etic-emic dilemma. *International Journal of Psychology*, 1976, *11*, 1–13.

Davidson, A., & Thomson, E. Cross-cultural studies of attitudes and beliefs. In H. Triandis & R. Brislin (Eds.), *Handbook of cross-cultural psychology* (Vol. 5: Social Psychology). New York: Allyn & Bacon, 1980.

Deutscher, I. Words and deeds. *Social Problems*, 1966, *13*, 235–254.

Deutscher, I. Looking backward: Case studies on the progress of methodology in sociological research. *American Sociologist*, 1969, *4*, 35–41.

Ehrlich, H. Attitudes, behavior and the intervening variables. *American Sociologist*, 1969, *4*, 29–34.

Fazio, R., & Zanna, M. Attitudinal qualities relating to the strength of the attitude–behavior relationship. *Journal of Experimental Social Psychology*, 1978, *14*, 398–408.

Fazio, R., Zanna, M., & Cooper, J. Direct experience and attitude–behavior consistency: An information processing analysis. *Personality and Social Psychology Bulletin*, 1978, *4*, 48–51.

Fishbein, M. An investigation of the relationship between beliefs about an object and attitude toward that object. *Human Relations*, 1963, *16*, 233–240.

Fishbein, M. Attitude and the prediction of behavior. In M. Fishbein (Ed.), *Readings in attitude theory and measurement*. New York: Wiley, 1967.

Fishbein, M., & Ajzen, I. Attitudes toward objects as predictors of single and multiple behavioral criteria. *Psychological Review*, 1974, *81*, 59–74.

Fishbein, M., & Ajzen, I. *Belief, attitude, intention and behavior*. Reading, Mass.: Addison-Wesley, 1975.

Fishbein, M., & Jaccard, J. Theoretical and methodological considerations in the prediction of family intentions and behavior. *Representative Research in Social Psychology*, 1973, *4*, 37–51.

Hayes, W. *Statistics*. New York: Holt, Rinehart & Winston, 1963.

Heberlein, T., & Black, J. Attitudinal specificity and the prediction of behavior in a field setting. *Journal of Personality and Social Psychology*, 1976, *33*, 474–479.

Kelley, S., & Mirer, P. W. The simple act of voting. *American Political Science Review*, 1974, *68*, 572–591.

Kelman, H. Attitudes are alive and well and gainfully employed in the sphere of action. *American Psychologist*, 1974, *29*, 310–324.

Kiesler, S. Post-hoc justification of family size. *Sociometry*, 1977, *40*, 59–67.

McGuire, W. The nature of attitudes and attitude change. In G. Lindzey & E. Aronson (Eds.), *The handbook of social psychology* (2nd ed., Vol. 3). Reading, Mass.: Addison-Wesley, 1969.

Norman, R. Affective-cognitive consistency, attitudes, conformity, and behavior. *Journal of Personality and Social Psychology*, 1975, *32*, 83–91.

Peak, H. Attitude and motivation. *Nebraska Symposium on Motivation* (Vol. 3). Lincoln: University of Nebraska Press, 1955.

Peters, C., & Van Voorhis, W. *Statistical procedures and their mathematical bases*. New York: McGraw-Hill, 1940.

Regan, D., & Fazio, R. On the consistency between attitudes and behavior: Look to the method of attitude formation. *Journal of Experimental Social Psychology*, 1977, *13*, 28–45.

Rindfuss, R., & Sweet, J. *Postwar fertility trends and differentials in the United States*. New York: Academic Press, 1977.

Rokeach, M. *The nature of human values*. New York: Free Press, 1973.

Rosenberg, M. Cognitive structure and attitudinal affect. *Journal of Abnormal and Social Psychology*, 1956, *53*, 367–372.

Schuman, H., & Johnson, M. Attitudes and behavior. *Annual Review of Sociology*, 1976, *2*, 161–207.

Schwartz, S. Normative explanations of helping behavior: A critique, proposal and empirical test. *Journal of Experimental Social Psychology*, 1973, *9*, 349–364.

Schwartz, S. Temporal instability as a moderator of the attitude–behavior relationship. *Journal of Personality and Social Psychology*, 1978, *36*, 715–724.

Snyder, M., & Swann, W. When actions reflect attitudes: The politics of impression management. *Journal of Personality and Social Psychology*, 1976, *34*, 1034–1042.

Snyder, M., & Tanke, E. Behavior and attitudes: Some people are more consistent than others. *Journal of Personality*, 1976, *44*, 501–517.

Warner, L., & DeFleur, M. Attitude as an interactional concept: Social constraint and social distance as intervening variables between attitudes and action. *American Sociological Review*, 1969, *34*, 153–169.

Weigel, R., & Newman, L. Increasing attitude–behavior correspondence by broadening the scope of the behavioral measure. *Journal of Personality and Social Psychology,* 1976, *33,* 793–802.

Weigel, R., Vernon, D., & Tognacci, L. The specificity of the attitude as a determinant of attitude–behavior congruence. *Journal of Personality and Social Psychology,* 1974, *30,* 724–728.

Weinstein, A. Predicting behavior from attitudes. *Public Opinion Quarterly,* 1972, *36,* 355–360.

Werner, P., Middlestadt-Carter, S., & Crawford, T. Having a third child: Predicting behavioral intentions. *Journal of Marriage and the Family,* 1975, *37,* 348–358.

Westoff, C., Potter, R., Sagi, P., & Mishler, E. *Family growth in metropolitan America.* Princeton, N.J.: Princeton University Press, 1961.

Whelpton, P. K., & Kiser, C. V. (Eds.). *Social and psychological factors affecting fertility* (5 vols.). New York: Milbank Memorial Fund, 1946–1958.

Wicker, A. Attitudes versus actions: The relationship of verbal and overt behavioral responses to attitude objects. *Journal of Social Issues,* 1969, *25,* 41–78.

Wicker, A., & Pomazal, R. The relationship between attitudes and behavior as a function of specificity of attitude object and presence of a significant person during assessment conditions. *Representative Research in Social Psychology,* 1971, *2,* 26–31.

Received August 1, 1978 ▪

Journal of Personality and Social Psychology
1979, Vol. 37, No. 8, 1377–1386

Mind Over Matter: Perceived Success at Psychokinesis

Victor A. Benassi, Paul D. Sweeney, and Gregg E. Drevno
California State University, Long Beach

Four experiments showed that college students' estimates of success on a psychokinetic (PK) task were affected independent of actual performance. In Experiment 1, subjects given a positive introductory set or no set about PK evidenced more illusory control than subjects given a negative set. In Experiment 2, both degree of general belief in psychic phenomena and the number of practice trials that subjects received influenced performance estimates, with high believers who received 10 practice trials providing the highest estimates and low believers who received 1 practice trial the lowest. In Experiment 3, subjects actively involved with the PK task judged their performance more positively than passively involved subjects. Experiment 4 showed that when they were actively involved in the task, subjects with an internal locus of control gave higher estimates of their success than subjects with an external locus of control. When passively involved, internals and externals did not reliably differ in their estimates, but their estimates were lower than those of active/internals. These results support Langer's illusion-of-control theory and also highlight the importance of general psychic belief and locus-of-control orientation in affecting perceived success at a psychic task.

How do people come to believe in the existence of psychic phenomena? Although research on the determinants of psychic belief is sparse, we can presume that such beliefs are determined by a number of factors, including media inputs, impactful personal experiences, inherent biases in human reasoning, and so forth (Sweeney, Benassi, & Drevno, Note 1). The present studies focused on several factors affecting estimates of personal performance on a psychokinetic (PK) task. Our intention was to provide a partial answer to the question of how people come to believe in their personal efficacy. We suggest that the same variables affecting perceived control in general also influence judgments of perceived psychokinetic control.

Using Langer's (1975) illusion-of-control model, Ayeroff and Abelson (1976) reasoned that belief in success on a mental telepathy task could be influenced by the introduction of skill-related but irrelevant variables into the situation. They found that subjects permitted to choose the symbols to be transmitted in a standard mental telepathy task and/or to engage in a warm-up session thought they did much better on the task than subjects who were not provided with these experiences. Objectively, the subjects performed at chance levels.

Layton and Turnbull (1975) found that attitude toward and evaluation of ESP were systematically related to *performance* on an ESP task. Two experiments were conducted, and an ESP effect was found in the first study, but not in the second. In both studies the authors found that subjects with a positive attitude toward ESP and a positive evaluation of it *thought* they did better on the ESP task

We wish to thank Rhoda Lindner, David Dowell, Lyle Creamer, Barry Singer, Raphael Hanson, and Peggy Murray for their helpful comments. Thanks are extended also to Robert Root for his technical assistance.

Paul D. Sweeney is now at the University of Pittsburgh.

Requests for reprints should be addressed to Victor A. Benassi, Department of Psychology, California State University, Long Beach, California 90840.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3708-1377$00.75

than their negative counterparts. In an early study, James and Rotter (1958) found that verbalized expectancies for success on an ESP task during extinction depended on whether subjects were told that ESP was a skill or chance task and whether they were given 100% or 50% reinforcement (positive feedback) during training.

In Experiment 1 we predicted that estimates of PK performance would vary with the instructional set. We hypothesized that subjects given a positive introductory statement about the existence of PK would rate their performance more positively than subjects given a negative set (cf. Layton & Turnbull, 1975). Additionally, we predicted that subjects given no specific set about PK would give higher estimates of success than those given a negative set, but subjects given no specific set would not differ in perceived success from positive-set subjects. This latter prediction derives from the finding that college students as a group are generally positive in their attitudes toward psychic phenomena (e.g., Jones, Russell, & Nickel, 1977; Polyella, Popp, & Hinsman, 1975). We hypothesized that our positive set merely provides support for what subjects already tend to believe.

Experiment 1: Instructional Set

Method

Subjects. Thirteen male and 24 female introductory psychology students met a course requirement at California State University, Long Beach, by participating in the study.

Apparatus. A device called a *die funnel*[1] was constructed in such a way that subjects could not see the outcome of their die tosses. A subject sat at one end of the funnel and tossed the die into the chute. The experimenter then opened a box at the opposite end, recorded the outcome, and returned the die to the subject.

Procedure. Individual subjects reported to the experimental room and were told that they were to be part of a study dealing with psychokinesis. Subjects were then randomly assigned to one of the following instructional set conditions:

1. (*Pro set*). Research studies have indicated that people have varying degrees of psychokinetic abilities, for example, they are able to bend spoons and repair watches without physically touching them. These abilities have been documented by scientists from both the USA and the USSR. The following study will attempt to verify and replicate these findings. Specifically, this study will concern itself with measuring the effect that one's psychokinetic abilities have on the throwing of dice.

2. (*Con set*). Much research has been done concerning people's psychokinetic abilities, that is, the alleged ability to move objects or alter their shapes without physically touching them, for example, bending spoons and repairing watches. The research demonstrates little or no evidence that people actually possess such abilities. The experiments by both United States and Soviet scientists that have tested these purported abilities have to date failed to establish evidence for PK. Our experiment is another in a long series in search of PK abilities.

3. (*Neutral set*). This will be an experiment investigating psychokinesis, that is, the ability to move objects or alter their shapes without physically touching them, for example, bending spoons or repairing watches.

After completing one of the above sets, Experimenter A left the room as Experimenter B entered. Blind to the set provided by Experimenter A, Experimenter B described the procedure for conducting the PK test as follows:

> We have constructed this apparatus called a *die funnel*. It acts as a funnel into which you will throw this die. The die will be put into this cup and then thrown into the funnel. It will tumble down to the end, whereupon I will record the outcome of the toss. As you will notice, the die has three sides painted green and three sides painted red [subject is shown the die]. For each of 20 trials we will run, you will choose a color and then I will ask you to concentrate your full mental energies for a period of 10 seconds on affecting the outcome of the toss. For example, you might choose the target color red; next you will concentrate for 10 seconds; then, when I say "begin," you will immediately toss the die into the funnel. I will then record the outcome of your toss on the sheet I am holding. Further, I should like you to indicate on this sheet [subject is given data sheet] whether you are confident or not confident that the color you chose actually comes up.

Upon hearing the above instructions, a subject completed 20 trials, after which he or she filled out a short questionnaire. In introducing the questionnaire, Experimenter B briefly described chance performance (i.e., about 10 hits) on the task the subject had just completed.

In this experiment and in the subsequent three experiments, subjects were debriefed through written feedback after all subjects had participated.

---

[1] Photographs of this apparatus may be obtained from the first author.

Table 1
Estimated Success as a Function of Instructional Set

Group	Number of confident trials (0–20)	Self-influence	Estimates on 20 more trials	Perceived correct
Pro (n = 13)				
M	11.85***	5.00	12.20**	11.90*
SD	2.79	2.90	2.19	2.39
Neutral (n = 10)				
M	12.60***	4.40	12.00**	11.50*
SD	1.27	1.43	2.40	1.51
Con (n = 14)				
M	7.57	3.36	9.80	9.90
SD	3.18	2.17	1.89	1.83

Note. For each dependent variable, a Fisher's least significant difference test was performed following the analysis of variance. None of the differences between the means of the pro- and neutral-set groups exceeded the critical values required at a p value of .05. However, for each measure, the p values adjacent to the means of the pro- and neutral-set groups indicate the level at which they are significantly different from the means of the con-set group.
* $p < .05$.
** $p < .025$.
*** $p < .001$.

Results

We found no evidence of sex differences on any of the measures taken in this study or in any of the subsequent three studies. Therefore all data have been collapsed across sex.

The mean success rate did not differ across groups ($F < 1$): pro = 11.6, neutral = 10.7, and con = 11.0. Binomial tests showed that no group performed significantly above or below chance levels. We also performed binomial tests to determine whether people deviated from chance performance on those trials on which they were confident of being correct, and we found the tests for each group to be clearly nonsignificant.[2]

Table 1 shows that there were significant differences on the confidence measure between the treatment groups, $F(2, 34) = 13.31$, $p < .001$, $\omega^2 = .280$. While the means for the pro- and neutral-set subjects did not systematically differ, both were significantly different from the mean for the con-set participants ($p < .001$ with Fisher's least significant difference [LSD] test).

Table 1 also shows the mean scores on the questions asked of subjects upon completion of the PK task. In response to the Likert format question "How much do you feel that your concentration influenced what colors came up? That is, how much PK influence do you feel you have exerted?" we found nonsignificant mean differences among groups, $F(2, 34) = 2.31$, $p > .05$. The results of this measure were in the same directional pattern as those on the other dependent measures. To the question "If you were given another 20 trials, on how many do you think you would influence what colors would come up?" we found a significant effect, $F(2, 34) = 5.04$, $p < .05$, $\omega^2 = .175$. We also found a significant difference on the question "On how many of the 20 trials do you think you correctly predicted what colors came up?":

[2] Combining the PK data from the three experimental groups resulted in a mean hit rate of 55.5%, which is nearly three binomial standard deviations from the expected mean of 50%. There are, of course, any number of factors that might have been responsible for this effect. The combined hit rate cannot account, however, for the significant estimated success differences among the experimental groups on the dependent measures. We may recall that the hit rates for groups did not separately deviate from chance. The combined group PK effect, while perhaps interesting, does not discount the result that belief measures were affected by the instructional set and not by actual performance.

$F(2, 34) = 4.06$, $p < .02$, $\omega^2 = .139$. Post hoc comparisons done on all of the significant analyses revealed that although the pro- and neutral- group means did not differ from each other, they were both significantly higher than the means for the con group ($ps < .05$ with Fisher's LSD test).

Discussion

The results support the notion that a variation in set can lead subjects to make differing estimates of performance on a PK task, independent of actual performance (cf. Layton & Turnbull, 1975). Subjects in the pro- and neutral-set conditions judged their performance to be more successful than subjects in the con-set condition. However, estimates of the pro and neutral groups did not reliably differ. Whether our manipulations actually influenced belief or only influenced verbal report remains an open question. A strong believer in psychic abilities, for example, may have been reluctant to tell the experimenter that she or he performed well when the experimenter had just stated that there is no evidence to support the existence of PK. In other words, the results could be an artifact of demand characteristics of the experimental situation. The results of the subsequent three experiments cannot easily be attributed to demand characteristics.

Subjects threw the die and chose the target color before each trial; the ratings of the neutral- and pro-set groups might have been lower if the experimenter had performed these functions. Additionally, the negative set seems to have overridden the effects of the subjects' active involvement with the task. Had these subjects been passively involved, we might expect their ratings to have been even lower than they were. Because involvement was not manipulated in this study, we cannot say to what extent active involvement influenced the judgments of our subjects separate from the effects of the instructional set. This orthogonal variation of set and involvement might clarify the contributions of these variables in determining perceived success.

Experiment 2: Practice and Belief

Based on the work of Langer (1975) we predicted that subjects given more prior exposure to a PK task would judge their performance during an actual test series more positively than subjects provided with less prior exposure. To the extent that skill-related variables are introduced into a chance situation, Langer found, people will judge their performance to be skillful. For example, practice in flipping a coin does not enhance subsequent performance, but the subject may mistakenly form the impression that it does. We also predicted that subjects with relatively strong general beliefs in psychic phenomena would give higher ratings of their performance on the PK task than more skeptical subjects would. Jones et al. (1977) found that subjects who scored higher on a paranormal belief scale gave higher estimates of their performance on a PK-type task than subjects who scored lower on the belief scale.

Method

Subjects. Fifty-five students, 30 female and 25 males, at California State University, Long Beach fulfilled an introductory psychology course requirement by participating.

Apparatus and materials. An apparatus called the die tumbler (see Footnote 1) allowed for electronic release of a die. The die, which sat on a platform, was released 10 seconds after a timer switch had been activated by the experimenter. During the experiment subjects sat at one end of the tumbler, from which they could see two lights: a red light that stayed on for the duration of the 10-second concentration period and a white light that signified the end of this period and the release of the die. The subjects could not see the outcome of the die tosses.

A portion of the Belief in the Paranormal Scale (BPS) was used to assess subject's general belief (see Table 1 of Jones et al., 1977, for the complete scale). The nine questions used were Numbers 1, 11, 15, 16, 19, 21, 23, 24, and 25.

Procedure. Same-sex subjects reported to the laboratory in groups of three or four. Subjects were given a definition of PK similar to the one in Experiment 1. They were then asked to inspect the die tumbler. The procedure was demonstrated, and subjects were asked to sit at predetermined spots located about 3 feet (1 m) from the tumbler. They sat at a desk separated by partitions, which prevented them from seeing each other's responses but allowed them to see the die tumbler. Subjects were

next told that when the red light on the die tumbler came on, they were to "concentrate your full mental energies on making the die land face up." The die used in Experiment 1 was then presented for all subjects to inspect.

According to procedure, the experimenter called out a target color and then started the 10-second concentration period. After the die was released, subjects completed the same confident/not confident checklist described in Experiment 1.

Subjects were assigned to a condition providing 1 or 10 practice trials in which they were given experience with the procedure to be used in the test series. After the practice trial(s), Experimenter A left the room, and Experimenter B entered and began the series of 20 trials. The sequence of target colors during practice and testing sessions was constant across conditions. Experimenter B was blind to the practice-trial condition of the subjects. Upon completion of the 20 trials, subjects filled out two Likert-format questionnaires. The first contained one question that asked, "How much do you feel that your concentration influenced what colors came up on the 20 trials?" The second questionnaire was the BPS.

Results

We found no evidence of PK influence. The mean hit rate was 9.2 for the group that performed 1 practice trial and 10.4 for the group that performed 10 practice trials ($t < 1$, *ns*). Also, as determined by binomial tests, neither group performed above or below chance. As in Experiment 1, subjects did no better or worse on the trials on which they were confident.

Within both trial groups a median split was done on the BPS questionnaire results. Subjects in the 1- and 10-trial conditions who scored above the median values of 32 and 31, respectively, were placed in high-belief groups and those below the median in low-belief groups. The above procedure yielded a 2 × 2 analysis of variance (ANOVA) design with practice trials and belief in the paranormal as independent variables.

Table 2 (column 1) shows that the number of trials on which subjects were confident of success varied with both the number of practice trials, $F(1, 51) = 16.24$, $p < .001$, $\omega^2 = .194$, and with degree of general psychic belief, $F(1, 51) = 10.1$, $p < .005$, $\omega^2 = .116$. The interaction was nonsignificant ($F < 1$).

Table 2 (column 2) shows ratings on the posttask question, "How much do you feel that your concentration influenced what colors came up on the 20 trials?" We found main effects for the practice trial variable, $F(1, 51) = 4.33$, $p < .05$, $\omega^2 = .05$; and for the general belief variable, $F(1, 51) = 7.72$, $p < .01$, $\omega^2 = .101$. The interaction was nonsignificant, $F(1, 51)$ 2.29, $p > .05$.

Table 2
Estimated Success as a Function of Practice Trials and Prior Belief

Group	Number of confident trials (0–20)	Self-influence estimates (0–9)
1 Practice trial		
Low belief		
M	6.83	1.88
SD	3.74	2.15
High belief		
M	10.14	4.57
SD	3.01	2.21
10 Practice trials		
Low belief		
M	10.87	4.13
SD	3.62	2.53
High belief		
M	13.00	4.93
SD	2.04	2.34

Discussion

This experiment demonstrates the effects of practice and prior belief on judgments of success on a psychic task, without any apparent accompanying psychic ability. These results replicate Ayeroff and Abelson (1976) and support Langer's (1975) illusion-of-control model.

Experiment 3: Involvement

We sought further confirmation of the hypothesis that people given a skill orientation would judge their performance on a PK task more positively than those not given this orientation. Subjects run in pairs were either actively or passively involved in the PK task. We predicted that active subjects would perceive more control over the task than passive subjects. We also attempted to replicate the effect of general paranormal belief on estimated PK performance.

Method

Subjects. Thirty female and 30 male introductory psychology students at California State University, Long Beach, fulfilled a course requirement by participating.

Apparatus. The apparatus was the same as in Experiment 1.

Procedure. Subjects came to the experiment room in same-sex pairs, were given a definition of PK similar to the one given in the previous two studies, and were told that the present investigators were beginning a research program on PK. Subjects were then asked to inspect the die funnel, after which they were told that the experiment called for both members of each pair to explore at the same time the possible effects that two persons have on the throw of a die. They were also told that although they both were to try to affect the outcome of each die toss, only one of them would actually toss the die. Experimenter A randomly chose one of the subjects to throw the die and then asked both to sit at predetermined spots at one end of the die funnel. Subjects were separated by a partition so that they could not see each other's responses on the confidence checklist.

The painted die used in Experiments 1 and 2 was again produced for the subjects to inspect. During this process, subjects were read instructions similar to those given in Experiments 1 and 2, asking them to center and concentrate their full mental energies for a period of 10 seconds on affecting the color called. Also, Experimenter A instructed subjects on how to complete the confident/not confident checklists. The sequence of target colors was constant across conditions.

On each of 14 trials, both subjects concentrated on making the target color land face up, one subject then threw the die, and finally both responded to the confident/not confident checklist. Following the last trial, Experimenter A left the room after indicating that a coexperimenter would arrive shortly. Experimenter B entered, blind to the condition experienced by the subjects, and administered two questionnaires. Questionnaire 1 consisted of two questions constructed on a Likert-format scale. The first asked, "How much do you feel that *your* concentration influenced what colors came up on the 14 trials? That is, how much psychokinetic control do you feel you had?" The second question asked, "How much do you feel the *other* person's concentration influenced what colors came up on the 14 trials? That is, how much psychokinetic control do you feel the other person had?" The order of presentation of the questions was reversed for different subjects. Questionnaire 2 asked subjects to judge on a Likert-format scale the extent to which they believed in psychic phenomena: "Do you believe that people possess psychic abilities, such as ESP and PK?"

Results

The mean success rate for subjects was 7.1. The results of all analyses attempting to show PK effects may be summarized succinctly—no PK effects were found.

As in Experiment 2, we generated high- and low-psychic-belief groups within the active and passive conditions. The median scores on the general belief question for the active and passive conditions were 6 and 7, respectively. Within each condition, subjects scoring above the median were placed in a high-believer group and those scoring below the median were placed in a low-believer group.

Judgments of success varied across the involvement conditions. Table 3 (column 1) shows that active subjects were confident on more trials than were passive subjects, and an ANOVA on these data shows this difference to be significant, $F(1, 59) = 11.10$, $p < .005$, $\omega^2 = .147$. No other significant differences were found (Fs < 1).

An ANOVA was performed with high/low belief and active/passive involvement as between-subjects variables and the self/other perceived-influence questions as a within-subject variable, the dependent measure being the rating on the self/other questions. There was no significant main effect for belief ($F < 1$). There was a significant main effect for the self/other variable such that subjects rated their own control over the die to be greater than that of their partners, $F(1, 56) = 11.21$, $p < .005$, $\omega^2 = .09$. Similarly, active subjects gave higher ratings than passive subjects, $F(1, 56) = 4.23$, $p < .05$, $\omega^2 = .04$. The above main effects must be qualified in light of a significant Active/Passive × Self/Other interaction, $F(1, 56) = 10.49$, $p < .005$, $\omega^2 =$ differ, although they were lower than the *self* PK influence ratings relative to their *other* PK influence ratings (see Figure 1). The *self/other* ratings for passive subjects did not differ, although they were lower than the *self* ratings of active subjects. No other significant interactions were found.

Discussion

Active subjects were confident on significantly more trials than passive subjects (cf.

Table 3
Estimated Success as a Function of Involvement and Prior Belief

Group	Number of confident trials (0–14)	Self-influence estimates (0–9)	Other influence estimates (0–9)
Passive involvement			
Low belief			
M	7.06	2.59	2.53
SD	2.46	2.15	2.15
High belief			
M	7.62	3.58	3.38
SD	1.19	2.07	1.56
Active involvement			
Low belief			
M	8.73	4.73	3.13
SD	1.58	2.34	2.00
High belief			
M	8.93	4.87	2.93
SD	1.16	1.46	2.12

Golin, Terrell, & Johnson, 1977; Langer, 1975; Wortman, 1975; Benassi, Roher, & Reynolds, Note 2). The second finding was the significant Self/Other × Active/Passive interaction. In general, when two people are jointly involved with a task, one may feel more responsible for producing an outcome than the partner is only when one is the more active participant in the situation. The fact that our active subjects gave higher *self* than *other* ratings in a chance situation provides further support for Langer's (1975) illusion-of-control model.

Experiment 4: Involvement and Locus of Control

The major purpose of the present study was to determine how persons with an internal or external locus of control respond to a PK task. Rotter (1966) proposed that persons differ in the extent to which they feel control over events, with internals perceiving more control than do externals. Some research is consistent with the notion that internals perceive control over events only when the events are seen as potentially controllable (see Geen, 1976, pp. 242–243). In general terms, internals should perceive control in skill but not chance situations (cf. Miller & Seligman, 1973; Rotter, 1966). Conversely, externals should perceive little control in chance or skill situations. We predicted that active/internals would perceive more control in a situation than active/externals, passive/internals, and passive/externals, who would not reliably differ in their estimated control. In our experiment we induced the perception of skill by making one of a pair of subjects actively involved with the task; the perception of chance was induced by making the other partner a passive participant in the task.

Method

Subjects. Thirty male and 30 female introductory psychology students at California State University, Long Beach, completed a course requirement by participating.

Apparatus and materials. The die funnel was used. Responses to each of the 46 statements from Rotter's (1966) locus-of-control scale were given on a *disagree/agree* format (based on Collins, 1974). Scoring was accomplished by summing subject's ratings to each statement, with higher sums reflecting externality. Scoring was reversed on external items, that is, a rating of 1 was changed to 5, and so forth.

Procedure. Subjects reported to the experimental room in groups of five. The experimenter gave a general introduction similar to that given in Experiment 3. Subjects were told that the experimenter could only work with one subject at a time, and they were asked to complete an attitude questionnaire (the locus-of-control scale) while waiting for

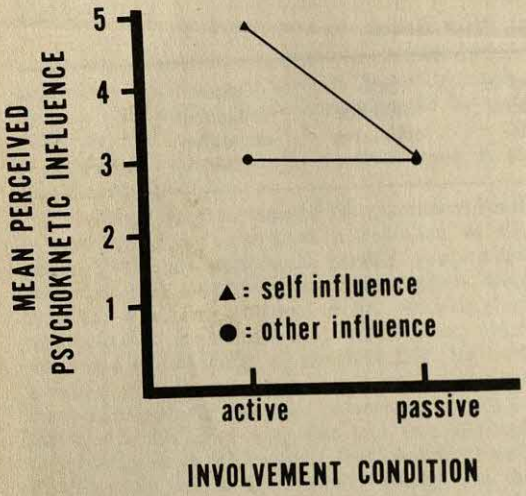


Figure 1. Active/Passive × Self/Other interaction.

their turns. Before the questionnaire was completed, a "lottery" chose one subject to serve as the experimenter's helper; the lottery was manipulated so that a confederate of the experimenter was always selected.

The four real subjects were given time to complete the questionnaire. Next, they were individually put through the following experimental procedure, similar to that of Experiment 3. The subject was given a target color; he or she concentrated on producing the target for 10 seconds; the die was thrown; the subject verbally indicated on a 0-to-10 scale his or her degree of confidence that the target color had come up. The confederate recorded the confidence ratings and the outcome of each die toss. In one condition, the subject threw the die (active involvement); in the other condition, the experimenter threw the die (passive involvement). Each subject received 20 trials with a constant sequence of target colors. Neither the experimenter nor the confederate was informed about the purpose of the study until all subjects had been run.

Results

The data of the 10 most internal and 10 most external subjects from the active and passive conditions were analyzed. The rationale for this decision was that we wished to create more extreme groups on the locus-of-control dimension than would probably have resulted if we had created groups by doing a median split on the locus-of-control scores. Table 4 (row 1) shows the mean locus-of-control scores for each of the Locus of Control × Involvement groups.

The hit rates for the experimental groups did not reliably differ from each other ($F < 1$). Also, no group reliably deviated from chance performance (Table 4, row 2).

Each subject's confidence rating was obtained by taking the mean rating for the 20 test trials. Table 4 (row 3) shows that active subjects were more confident than passive subjects, $F(1, 36) = 4.02$, $p < .06$, $\omega^2 = .06$. Internals gave nonsignificantly higher ratings than externals, $F(1, 36) = 2.11$, $p > .10$. The predicted interaction did not reach significance, $F(1, 36) = 2.53$, $p > .10$.

The ratings on the posttask measures provided consistent and moderate support for the hypothesized interaction. On the Likert-format question "How much control do you feel you had over what colors came up from the die tosses?" active subjects gave higher estimates than passive subjects, $F(1, 36) = 4.23$, $p < .05$, $\omega^2 = .07$ (Table 4, row 4). Internals gave nonsignificantly higher ratings than externals, $F(1, 36) = 2.75$, $p > .10$. The interaction was marginally significant, $F(1, 36) = 3.70$, $p < .08$, $\omega^2 = .06$. To the Likert-format question "Do you feel you would get better on the task if you were given another 20 trials?" active subjects gave higher ratings than passive subjects, $F(1, 36) = 7.93$, $p < .01$, $\omega^2 = .11$ (Table 4, row 5). Internals were significantly more confident than externals, $F(1, 36) = 8.73$, $p < .01$, $\omega^2 = .13$. The interaction was significant, $F(1, 36) = 7.18$, $p < .02$, $\omega^2 = .10$. To the question "If you were given another 20 trials, how many do you think you would get?" active subjects gave nonsignificantly higher ratings than passive subjects, $F(1, 36) = 3.14$, $p > .1$ (Table 4, row 6). Internals gave higher ratings than externals, $F(1, 36) = 5.99$, $p < .02$, $\omega^2 = .09$. The interaction was significant, $F(1, 36) = 8.72$, $p < .01$, $\omega^2 = .14$.

Discussion

Active subjects gave higher estimates of their performance, replicating Experiment 3 as well as previously cited research. Internals gave higher ratings than externals on two of the four dependent measures. These results should be qualified, however, in light of the

Table 4
Summary of Data From Experiment 4

Measure	Active		Passive	
	Internals	Externals	Internals	Externals
Locus of control				
M	120.80	150.80	118.40	151.40
SD	10.95	4.38	12.69	6.83
Hit rate				
M	9.80	10.10	10.00	9.80
SD	2.80	2.28	1.83	2.40
Confidence (0–10)				
M	6.20	4.00	3.60	3.70
SD	1.61	2.16	2.87	2.31
Perceived control (0–10)				
M	4.90	2.20	1.90	2.10
SD	2.86	2.25	2.28	2.07
Improvement (0–10)				
M	6.10	2.00	2.10	1.90
SD	2.72	1.86	2.76	1.60
Another 20 trials				
M	13.10	9.90	10.30	10.60
SD	2.51	1.79	1.63	1.34

Note. For each dependent variable, a one-way ANOVA was done on the results of the active/external, passive/internal, and passive/external groups; all *F*s were less than 1. Next, for each measure, the data from these groups were collapsed and compared with the data of the active/internal subjects by way of *t* tests for independent means. Each comparison showed a difference between means at the $p < .001$ level of significance for a two-tailed test.

significant Involvement × Locus of Control interaction obtained on three of the four dependent measures. Comparison of the mean values in Table 4 shows that active/internals gave consistently higher estimates of their performance than subjects in the other three experimental groups, whose estimates did not reliably differ from one another. Internals seem to detect the presence or absence of skill cues and then attempt to exert control if skill is perceived as possible. In our experiment, skill was merely illusory. While an internal locus of control is viewed as a positive characteristic in many contexts, internals who threw the die manifested the highest degree of superstitious behavior among our subjects.

In a study similar to ours, Miller and Seligman (1973) failed to find a main effect for locus of control or a Locus of Control × Task (chance vs. skill) interaction. Differences between the studies make it difficult to account for these discrepancies. Miller and Seligman, for example, assigned subjects to internal or external groups based on a median split done on responses to Rotter's (1966) forced-choice scale. Also, they gave subjects bogus feedback after each trial of their skill and chance tasks. Lastly, they used different tasks for their skill and chance conditions.

General Discussion

In the present studies the situational variables of instructional set, practice, and involvement affected estimated success on a PK task independent of actual success. Similarly, general belief in paranormal phenomena and locus of control orientation was significantly related to estimated performance. The locus-of-control variable was found to interact with the involvement variable in Experiment 4. The results of Experiments 1–4 are consistent with Langer's (1975) skill theory of nonveridical perceived control. The fact that the theory was supported in a psychic context

using a variety of independent and dependent variables bolsters the theory's generalizability.

The consistent pattern of interaction found in Experiment 4 demonstrates the relevance of the locus-of-control construct for a theory of nonveridical perceived control. Externals actively involved with the task gave no higher estimates of their performance than either internals or externals passively involved with the task. Thus, being actively involved in a chance task may evince no greater estimate of performance than would be made after being passively involved with the task, *unless* the person has an internal locus of control. Future work might determine whether the effects of locus-of-control orientation could be overriden in contexts different from ours.

Langer's (1975) skill theory is a cogent account of how illusions of control can be induced. At the same time, the phenomenon of illusory control is sufficiently complex that it cannot as yet be accounted for by a single model. Future research might profitably focus on the following factors, which may be shown to be related to perceived illusory control: human biases in information processing (Kahneman & Tversky, 1973)[3]; coincidences, particularly dramatic (psychic?) ones; the human propensity to recall successes and to forget or play down failures (Ward & Jenkins, 1965); personality variables such as locus of control, need for certainty, and so forth. We are inclined to speculate that the same variables that affect one's judgment of personal control also affect one's judgment of another's control. Also, we suspect that the emotional concomitants associated with perceived psychic control are noticeably greater than those correlated with more mundane perceived control. For whatever reasons, the realm of the paranormal seems to conjure deep-seated emotional reactions that far exceed those warranted by the phenomena on which they are based. We plan to examine these speculations in our future work.

[3] In a manuscript in preparation, the first author found that in judging another's success on a psychic task, students fell prey to the availability and representation fallacies described by Kahneman and Tversky (1973).

Reference Notes

1. Sweeney, P. D., Benassi, V. A., & Drevno, G. E. *Perceived success at ESP & PK: A review.* Paper presented at the meeting of the Western Psychological Association, San Francisco, California, April 1978.
2. Benassi, V. A., Roher, C., & Reynolds, C. *Perceived control over chance events.* Paper presented at the meeting of the Western Psychological Association, Seattle, Washington, April 1977.

References

Ayeroff, F., & Abelson, R. P. ESP and ESB: Belief in personal success at mental telepathy. *Journal of Personality and Social Psychology,* 1976, *34,* 240–247.

Collins, B. E. Four components of the Rotter Internal–External Scale: Belief in a difficult world, a just world, a predictable world, and a politically responsive world. *Journal of Personality and Social Psychology,* 1974, *29,* 381–391.

Geen, R. G. *Personality: The skein of behavior.* St. Louis, Mo. C. V. Mosby, 1976.

Golin, S., Terrell, F., & Johnson, B. Depression and the illusion of control. *Journal of Abnormal Psychology,* 1977, *86,* 440–442.

James, W., & Rotter, J. B. Partial and 100% reinforcement under chance and skill conditions. *Journal of Experimental Psychology,* 1958, *55,* 397–403.

Jones, W. H., Russell, D. W., & Nickel, T. W. Belief in the paranormal scale: An objective instrument to measure belief in magical phenomena and causes. JSAS *Catalog of Selected Documents in Psychology,* 1977, *7,* 100. (Ms. No. 1577)

Kahneman, D., & Tversky, A. On the psychology of prediction. *Psychological Review,* 1973, *80,* 237–251.

Langer, E. J. The illusion of control. *Journal of Personality and Social Psychology,* 1975, *32,* 311–328.

Layton, B. D., & Turnbull, B. Belief, evaluation, and performance on an ESP task. *Journal of Experimental Social Psychology,* 1975, *11,* 166–179.

Miller, W. R., & Seligman, M. E. P. Depression and the perception of reinforcement. *Journal of Abnormal Psychology,* 1973, *82,* 62–73.

Polyella, D. J., Popp, R. J., & Hinsman, M. C. ESP? JSAS *Catalog of Selected Documents in Psychology,* 1975, *5,* 1. (Ms. No. 1087)

Rotter, J. B. Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs,* 1966, *80*(1, Whole No. 609).

Ward, W. C., & Jenkins, H. M. The display of information and the judgment of contingency. *Canadian Journal of Psychology,* 1965, *19,* 231–241.

Wortman, C. B. Some determinants of perceived control. *Journal of Personality and Social Psychology,* 1975, *31,* 282–294.

Received August 17, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 8, 1387-1397

Communicator Physical Attractiveness and Persuasion

Shelly Chaiken
University of Toronto, Toronto, Canada

In a field setting, physically attractive or unattractive male and female communicator-subjects delivered a persuasive message to target-subjects of each sex. Results indicated that attractive (vs. unattractive) communicators induced significantly greater persuasion on both a verbal and behavioral measure of target agreement. In addition, female targets indicated greater agreement than did male targets. Data gathered from communicator-subjects during an earlier laboratory session indicated that physically attractive and unattractive communicators differed with respect to several communication skills and other attributes relevant to communicator persuasiveness, including grade point average, Scholastic Aptitude Test scores, and several measures of self-evaluation. These findings suggest that attractive individuals may be more persuasive than unattractive persons partly because they possess characteristics that dispose them to be more effective communicators.

Experimental evidence regarding the effect of communicator physical attractiveness on persuasion is equivocal. Although two studies have demonstrated that attractiveness can significantly enhance a male communicator's persuasiveness with both male and female message recipients (Horai, Naccari, & Fatoullah, 1974; Snyder & Rothbart, 1971), the majority of published experiments have failed to obtain significant attractiveness effects or have obtained interactions between attractiveness and other variables (Chaiken, Eagly, Sejwacz, Gregory, & Christensen, 1978; Mills & Aronson, 1965; Blass, Alperstein, & Block, Note 1). For example, Mills and Aronson (1965), using a female communicator and male recipients, found no overall effect of communicator attractiveness on persuasion. However, on a marginally significant basis, they did find that the communicator's expression of a desire to influence recipients enhanced her persuasiveness when she was attractive but not when she was unattractive. Blass, Alperstein, and Block (Note 1), who also studied a female communicating to male recipients, found that communicator attractiveness and race interacted to affect opinions. A white communicator was more persuasive if attractive than if unattractive, whereas a black communicator was more persuasive if unattractive. Finally, one experiment reported by Chaiken, Eagly, Sejwacz, Gregory, and Christensen (1978) indicated that the persuasive impact of attractive communicators depended both on the sexual composition of the communicator–recipient dyad and on whether recipients anticipated interacting with the communicator. Their second experiment, which employed the identical stimulus materials but utilized a somewhat different cover story, yielded no significant persuasion findings involving the attractiveness variable, however.

All of these experiments were conducted in laboratory settings and all employed experimental manipulations of communicator attractiveness. The majority of studies have

The author is grateful to Alice H. Eagly, Jonathan L. Freedman, and an anonymous reviewer for their comments on an earlier draft of this manuscript, and to John W. Fee III, Herschl Forman, Gerry Johnson, John Keenan, and Aaron Moscoe for their assistance during various phases of the research. A preliminary version of this article was presented at the meeting of the Eastern Psychological Association, Washington, D.C., 1978.

Requests for reprints should be sent to Shelly Chaiken, Department of Psychology, University of Toronto, Toronto, Ontario, Canada M5S 1A1.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3708-1387$00.75

used pictured communicators, varying appearance by means of photographs scaled for physical attractiveness (Chaiken et al., 1978; Horai et al., 1974; Snyder & Rothbart, 1971). Two experiments have utilized either live communicators (Mills & Aronson, 1965) or videotaped communicators (Blass et al., Note 1) whose appearance was varied by means of makeup, dress, and grooming.

The present research addressed two general concerns raised by previous empirical work. The first deals with the importance of the attractiveness cue outside the laboratory. The lack of consistent findings across previous experiments—particularly the paucity of attractiveness main effects and the occurrence of unpredicted and often theoretically ambiguous interactions (Chaiken et al., 1978; Blass et al., Note 1)—suggests that the persuasive impact of physical attractiveness may be far from robust. This concern is heightened by the possibility that the highly controlled settings of previous experiments made the attractiveness cue more salient than it would typically be in more naturalistic settings, and the context of these experiments may therefore have inflated the importance of this variable as a determinant of persuasion. In this regard, it is interesting to note that both experiments demonstrating an attractiveness main effect on opinion change (Horai, Naccari, & Fatoullah, 1974; Snyder & Rothbart, 1971) used photographic stimuli and employed relatively simple experimental procedures. In contrast, studies that have not demonstrated this main effect have used live or videotaped communicators (Mills & Aronson, 1965; Blass et al., Note 1) or have employed relatively more complex experimental procedures (Chaiken et al., 1978, Experiment 1).

Conversely, it might be that previous empirical findings underestimate the importance of the attractiveness variable in persuasion. The implicit demands of the laboratory may encourage subjects to adopt a highly logical mode of cognitive functioning. At the same time, such demands may discourage subjects from utilizing information such as physical attractiveness, which popular wisdom would hold to be irrelevant. Thus, although attractiveness may well influence opinion change in naturalistic settings, the psychology laboratory may provide a particularly unsuitable setting for demonstrating such an effect.

In sum, one concern of the experiment dealt with the importance of communicator attractiveness as a determinant of persuasion. To address this concern, the experiment utilized live, physically attractive and unattractive communicators who delivered a persuasive message to targets in a field setting. It was predicted that attractive communicators would be more persuasive than unattractive ones. No hypotheses were formed with respect to whether *attractiveness* might interact with *communicator* or *target sex* to affect opinions, although some previous findings (Chaiken et al., 1978, Experiment 1) suggest this possibility.

The second concern of the present research, one that presupposes that attractiveness does influence persuasion, deals with the explanation of attractiveness effects. In previous studies, the style in which the persuasive message is presented to subjects has remained constant across experimental conditions. In addition, excluding trait inferences that subjects make, the (typically hypothetical) communicators in these studies are equated on attributes other than attractiveness. Such standardization is of obvious value, since it allows one to conclude that observed differences in treatment are a product of the attractiveness cue rather than the result of some factor that is unintentionally confounded with the attractiveness manipulation. This type of design has been and should be used to evaluate a variety of explanations for the facilitating effect of attractiveness on persuasion, including psychological mechanisms such as credibility enhancement, identification (Kelman, 1961), social reinforcement, cognitive balance, and classical conditioning. However, such designs preclude from consideration the possibility that attractive and unattractive individuals are differentially persuasive because attractiveness is correlated with other attributes that affect communicator persuasiveness.

While little is currently known about actual differences between physically attractive and unattractive persons, researchers (Berscheid & Walster, 1974) have argued that because of differing socialization experiences (e.g., Clifford & Walster, 1973; Dion, 1972) attractiveness may be confounded with other attributes (e.g., intelligence, status, self-concept, personality). Consistent with this argument, Goldman and Lewis (1977) recently reported that attractive individuals may possess greater social skill than unattractive persons. The idea that attractiveness covaries naturally with other factors suggests that attractive individuals may be more persuasive than unattractive persons partly because they possess communication skills or other attributes that dispose them to be particularly effective communicators. To explore this possibility, the present study selected 17 different *communicator-subjects* to represent each of four possible combinations of communicator attractiveness and sex. Videotapes of these communicators delivering the persuasive message were examined to determine possible individual differences in communication skills. The skills measured in the experiment included those identified by previous research (speech rate, Miller, Maruyama, Beaber, & Valone, 1976; nonfluencies, McCroskey & Mehrley, 1969; vocal confidence, London, 1973; eye gaze, Mehrabian & Williams, 1969) or suggested by intuition (smiling) as correlates of communicator persuasiveness. To examine differences with respect to other attributes relevant to persuasiveness, communicator-subjects also completed the Rotter (1966) Internal–External Locus of Control Scale, rated themselves on a series of evaluative scales, and reported their grade point averages and Scholastic Aptitude Test (SAT) scores.

Method

Subjects and Overview of Design

A total of 110 male and female University of Massachusetts undergraduate psychology students participated for extra course credit as communicator-subjects. In a laboratory session immediately preceding their field participation, these individuals were trained to deliver a persuasive message, and their final practice performances were videotaped. In addition, communicator-subjects completed a questionnaire and were photographed. In the field, each communicator delivered the persuasive message to 2 University of Massachusetts undergraduates of each sex. Subsequently, an independent group of judges ($n = 56$) rated communicator-subjects' photographs on physical attractiveness (15-point scale). Communicator-subjects were rank ordered according to their mean attractiveness ratings, and those in the top or bottom third of the distribution for their sex were selected for inclusion in the design. This procedure resulted in the analysis of data from 68 communicators (17 per sex/attractiveness level) and 272 target-subjects (2 male and 2 female targets per communicator). Mean attractiveness scores for the four communicator groups were 8.80 (male/attractive, 9.09 (female/attractive), 6.10 (male/unattractive), and 6.34 (female/unattractive).

Laboratory Procedure

Communicator-subjects were recruited for an experiment entitled "social influence" and were scheduled individually for their laboratory sessions. The (male) experimenter informed communicators that they would be trained to deliver a persuasive message and would then attempt to persuade students whom they would approach on campus. Communicators were told that "in order to explore correlates of communicator persuasiveness," they would complete a questionnaire, be photographed, and be videotaped delivering the persuasive message.

After this introduction subjects completed the first part of a questionnaire on which they indicated their sex, age, grade point average, SAT scores, and agreement (7-point scale) with the position advocated in the persuasive message. Subjects completed the remaining three sections of the questionnaire in one of six possible orders. One section consisted of the 23-item (plus fillers) Internal–External Locus of Control Scale (Rotter, 1966). In a second section, subjects described themselves on 7-point bipolar adjective scales. Positive poles of the 17 scales used were *intelligent, interesting, assertive, confident, likable, knowledgeable, competent, sincere, physically attractive, modest, moral, persuasive, trustworthy, warm, attractive, friendly,* and *sensitive.* Finally, a third section asked subjects to "speculate about their future" by responding on 7-point scales to the following items: *to have an excellent* (vs. *poor*) *job, have a happy* (vs. *drab*) *family life, be regarded as a successful* (vs. *unsuccessful*) *person, be a contented* (vs. *discontented*) *person, be well-off financially* (vs. *in bad financial straits*), *be enjoying life* (vs. *depressed about things*), *be famous for something* (vs. *a very ordinary person*), *be a highly* (vs. *only moderately*) *educated person, be a confident* (vs. *a worried*) *person.*

Subjects next received a script containing the persuasive message as well as all procedural details pertaining to their field participation. To standardize communicator-subjects' training as much as possible,

the following procedure was employed. The experimenter twice went through the field procedure with each subject, once playing the role of communicator and once playing that of target-subject. Communicator-subjects were then allotted 10 minutes to practice the script on their own. Afterwards, the experimenter again went through the script twice with the communicator-subject, both times playing the role of target-subject.

Next, communicator-subjects were asked to deliver the persuasive message "just as you would to a real target-subject." This performance was delivered to an experimental assistant (male) and was videotaped. Subsequently the videotapes were scored by two independent judges (blind to the experimental hypotheses). Exposed only to the audio component of the videotape, each judge rated the communicator's vocal confidence (5-point scale). Using a digital stop-clock, judges recorded the time that the communicator gazed directly at his or her target (experimental assistant) and, to assess speech rate (seconds per word), recorded the time each communicator took to deliver the 111-word persuasive message. A smiling index (possible range 0 to 5) was obtained by having judges record the presence (1) or absence (0) of smiling in each of five segments of the message. Judges also counted the number of nonfluencies (vocal pauses within sentences, repetitions, "umms," "ers," stuttering, etc.) made during delivery of the message. Finally, judges rated the communicator's physical attractiveness (5-point scale) while exposed only to the video component of the tape. For each of the above measures, the two judges' ratings were averaged to obtain a score for each communicator-subject. Interrater reliability coefficients were speech rate, $r = .99$; vocal confidence, $r = 79$; gaze, $r = .99$; smiling, $r = 83$; nonfluencies, $r = .90$; physical attractiveness, $r = .67$.

Just prior to leaving the laboratory for the field, communicators were photographed in a standard pose. The $3\frac{1}{2} \times 5$ inch black and white photographs showing a head and shoulder view of each communicator were subsequently scaled for physical attractiveness by an independent group of judges (see above).

Field Procedure

Each communicator was randomly assigned to one of five campus locations and was then taken to that location by the experimenter. To prevent communicators from selecting their own target-subjects, each communicator was required to approach *every* passer-by until he or she had completed the entire field procedure with 2 university students of each sex. This rule was subject to the following constraints: (a) sex of person approached was required to alternate, (b) only persons walking alone were to be approached, and (3) an approach could not be initiated while the communicator was occupied with another target-subject.

The communicator introduced himself or herself to the target-subject and requested that the target complete an opinion survey. If the target agreed, the communicator stated that he or she was in a campus group favoring the proposition that "the University should stop serving meat at breakfast and lunch at all dining commons," and the communicator supported this position with two brief arguments. After restating the position, the communicator gave the target a confidential questionnaire to complete. On this questionnaire, targets indicated their agreement (7-point scale) with the message's overall position and rated the communicator's friendliness, knowledgeability, and attractiveness (15-point scales). Targets also indicated their sex and age and responded to some filler items designed to make the questionnaire appear to be an opinion survey. Afterwards, targets placed the completed questionnaire in a box provided by the communicator. The communicator next stated that he or she was also circulating a petition demanding that "the University stop serving meat at breakfast and lunch at all dining commons," and asked the target to sign it. After the target had signed (or refused to sign), he or she was thanked for participating in the survey.

Throughout the entire field procedure, the experimenter stood nearby and monitored all communicator–target interactions. Data from 6 (of the initial 110) communicators were excluded from consideration because they deviated greatly from the experimental script (4) or informed one or more of their targets that they were participating in an experiment (2). The experimenter also recorded the total number of approaches required by each communicator in order to complete the field procedure successfully with two male and two female targets.

Results and Discussion

Data provided by target-subjects were analyzed by four-way multivariate and univariate analyses of variance with 2 levels each of communicator sex, attractiveness, and target sex, and 17 levels of communicator (nested within communicator sex and attractiveness). Data from communicator-subjects were analyzed by two-way multivariate and univariate analyses of variance with 2 levels each of communicator sex and attractiveness.

Check on Experimental Design

The attractive communicator-subjects of the present study were clearly perceived as more physically attractive than their unattractive counterparts. An analysis of attractiveness ratings made by photograph judges revealed only an attractiveness main effect ($p < .0001$), as did an analysis of attractive-

Table 1
Mean Target Agreement and Proportion of Targets Signing Petition as a Function of Communicator Attractiveness, Communicator Sex, and Target Sex

Variable	Attractive communicator				Unattractive communicator			
	Male		Female		Male		Female	
	Male target	Female target	Male target	Female target	Male target	Female target	Male target	Female target
Target agreement	4.50	3.32	4.24	3.53	4.91	4.00	4.53	4.03
Petition signing	.29	.53	.35	.47	.35	.38	.24	.29

Note. On the agreement measure, lower numbers indicate greater agreement with the position advocated in the persuasive message. Cell $n = 34$.

ness ratings made by videotape judges' ($p < .001$). Further, targets-subjects perceived physically attractive (vs. unattractive) communicators as more attractive ($p < .05$). It should also be noted that communicator-subjects did not differ significantly as a function of attractiveness or sex with respect to their age or opinions on the persuasive message topic.

Field Data

The five target-subject variables (agreement, petition signing, perceptions of communicator friendliness, knowledgeability, and attractiveness) were first submitted to a four-way multivariate analysis of variance and were then analyzed on a univariate basis.[1] The multivariate analysis yielded significant effects due to communicator sex, $F(5, 60) = 2.96$, $p < .02$, and target sex, $F(5, 60) = 3.16$, $p < .02$, and a marginal effect due to communicator attractiveness, $F(5, 60) = 1.74$, $p < .14$.[2] The univariate analyses revealed that the communicator-sex, target-sex, and communicator-attractiveness main effects attained significance on one (attractiveness), two (agreement, petition signing), and three (agreement, petition signing, attractiveness) of the five target variables, respectively. The remainder of this section presents the results of these more informative univariate analyses.

Persuasion measures. Analysis of targets' agreement with the communicator's overall position (see Table 1) revealed main effects due to communicator attractiveness and target sex. Attractive communicators elicited greater agreement from targets than did unattractive communicators ($M = 3.89$ vs. $M = 4.37$), $F(1, 64) = 4.23$, $p < .05$, and female targets expressed greater agreement than male targets ($M = 3.72$ vs. $M = 4.54$), $F(1, 64) = 11.22$, $p < .005$. No other effects were significant on this measure. Dunnett's test (cf. Myers, 1972) indicated that all experimental groups (in the Communicator Sex × Communicator Attractiveness × Target Sex design) differed significantly from an opinion-only group of pilot subjects[3] in the direction of greater

[1] Following Lunney (1970) and the advice of an anonymous reviewer, the dichotomous petition-signing measure (coded 0,1) was included in the multivariate analysis and was also treated by univariate analysis of variance. The latter analysis yielded main effects due to communicator attractiveness, $F(1, 64) = 3.95$, $p < .06$, and target sex, $F(1, 64) = 5.07$, $p < .05$. These univariate results parallel those presented in the text, which are based on a more conventional treatment of dichotomous data.

[2] Since the nested communicator factor accounted for no significant effects in the four-way analysis, it was ignored in a subsequent three-way multivariate analysis on the five field measures. The three-way analysis, which employs an error term with greater degrees of freedom than the error terms utilized in the four-way hierarchical design analysis, yielded comparable findings: significant effects due to communicator sex, $F(5, 260) = 3.80$, $p < .005$, and target sex, $F(5, 260) = 2.93$, $p < .02$, and a marginal effect due to communicator attractiveness, $F(5, 260) = 1.83$, $p < .11$.

[3] Pilot subjects ($n = 18$) were approached by the experimenter at various campus locations and were

agreement with the position advocated in the message ($ps < .05$ or smaller).

The proportion of targets in each experimental condition who signed the communicator's petition was computed (see Table 1), and z tests were performed on the arcsin transformations of these proportions (cf. Langer & Abelson, 1972). Paralleling the agreement results, this analysis revealed that a greater proportion of targets signed the petition when the communicator was attractive ($p = .41$) rather than unattractive ($p = .32$), $z = 1.56$, $p < .06$, one-tailed, and a greater proportion of female (vs. male) targets signed ($p = .42$ vs. $p = .31$), $z = 1.91$, $p < .06$, two-tailed. No other effects were significant on petition signing.

It should be noted that communicators approached an average of 6.04 individuals in their efforts to deliver the persuasive message to the requisite number of targets. An analysis of the number of approaches that each communicator made yielded no significant effects. Thus male and female and attractive and unattractive communicators were equally successful in gaining the audience of targets.[4]

Other measures. In addition to the significant attractiveness effect on targets' attractiveness ratings, attractive communicators were perceived as somewhat friendlier than unattractive ones ($p = .07$). Finally, targets (unlike either videotape or photograph judges) rated female (vs. male) communicators as more attractive ($p < .001$). No other effects were significant on these measures, and no effects were obtained on targets' ratings of communicator knowledgeability.

The field data generally substantiate the physical attractiveness – persuasion relationship. On both a verbal and behavioral measure, target-subjects expressed significantly greater agreement with attractive, rather than unattractive, communicators. Although the study was not designed to evaluate possible psychological mechanisms underlying this effect, the fact that that targets' perceptions of communicator knowledgeability were unaffected by attractiveness suggests that attractiveness effects in persuasion are not typically mediated by differential perceptions of credibility (also see Norman, 1976; Snyder & Rothbart, 1971). On the other hand, the finding that targets' perceptions of communicator friendliness were marginally affected by attractiveness is consistent with psychological mechanisms such as identification (Kelman, 1961), social reinforcement, cognitive balance, and classical conditioning.

Target sex also had a significant impact on persuasion. Females expressed greater agreement with the communicator's message than did males. While not anticipated, this finding is consistent with two hypotheses recently proposed by Eagly (1978). One hypothesis is that heightened female influenceability may be observed when the message topic is biased against the expertise or interests of women. In view of pilot-subjects' ratings (see Footnote 3), it seems unlikely that the sexes were differentially expert on the meat topic. However, while this study did not assess involvement with the issue, recent findings obtained by Eagly (Note 2) indicate that males may be more interested in or involved with the meat question, perhaps because they are greater consumers of beef than are females. Since greater involvement typically decreases persuasion (e.g., Miller, 1965; Sherif & Hovland, 1961), the present findings are consistent with the idea that male targets' greater involvement with the message topic made them less willing than females to yield to the persuasive message. Eagly's second hypothesis suggests that heightened female influenceability may be observed in situations characterized by the communicator's physical presence

asked to indicate their agreement (7-point scale) with a variety of opinion statements and to judge whether males or females were more expert with respect to each opinion issue (7-point scale). The meat topic was chosen because pilot subjects disagreed strongly with the position advocated in the persuasive message ($M = 6.50$) and considered males and females equally expert on the issue ($M = 4.08$). Further, male and female pilot subjects did not differ significantly with respect to either their agreement or their expertise ratings, $ts(16) < 1.0$, on this issue.

[4] Of the 411 individuals approached by the study's 68 communicators, 139 (34%) refused the communicators' initial requests that they "complete an opinion survey."

and surveillance, and may be a product of females' greater concern with maintaining social harmony and insuring smooth interpersonal relations. The fact that a sex difference occurred in this experiment where communicators presented their messages to targets during face-to-face encounters is also compatible with this hypothesis.

As Eagly (1978) has noted, differences in interpersonal orientation as an explanation of sex differences in influenceability imply that the greater persuasibility of women may represent surface agreement rather than genuine belief change. Yet, if this explanation is entertained, the meaning of the experiment's persuasion measures and of the observed attractiveness effects is called into question. Did the greater "persuasion" induced by attractive (vs. unattractive) communicators represent genuine belief change or mere compliance? Although a compliance interpretation cannot be completely ruled out, at least two aspects of the field data favor the idea that communicator attractiveness did actually effect genuine changes in targets' opinions. First, unlike the petition-signing measure, communicators did not see targets' verbal agreement responses. Targets were told that their opinions were confidential, and targets themselves inserted their completed questionnaires in a ballot box provided by the communicator. Yet this measure, which should have been less susceptible to compliance pressures, showed an attractiveness effect of the same (and slightly higher) magnitude as that shown on petition signing. Second, a compliance interpretation suggests that attractive (vs. unattractive) communicators should have been more successful in inducing compliance with their initial requests that potential targets "complete an opinion survey." However, analysis on the number of approaches made by each communicator revealed no attractiveness effect ($p > .65$).

Individual Differences Between Communicator-Subjects

A second focus of the experiment was to investigate whether the greater persuasiveness of physically attractive communicators might be understood in terms of differences between attractive and unattractive individuals on dimensions relevant to communicator persuasiveness. The individual difference measures obtained during communicators' laboratory participation were first submitted to a two-way (Attractiveness × Sex) multivariate analysis of variance and were then analyzed on a univariate basis. The multivariate analysis yielded only an attractiveness main effect, $F(36, 29) = 9.52$, $p < .001$. However, the univariate analyses revealed that this effect attained significance or marginal significance on only a small proportion of variables. The remainder of this section presents the findings of these univariate tests.

Communication skills. Analyses on the communication skills measured in this study revealed that compared to unattractive communicators, attractive communicators were significantly more fluent speakers ($p < .05$) and had a marginally faster rate of speech ($p = .07$). Attractive and unattractive communicators did not differ with respect to vocal confidence, gaze, or smiling. The only other effects on these variables were due to communicator sex. Male (vs. female) communicators spoke faster ($p < .05$) and smiled less ($p < .001$) while delivering the message.

Other attributes. Phares (1965) found that students who were more internal (vs. external) in their locus-of-control orientation (Rotter, 1966) were more successful in influencing other students' attitudes. And Miller (1970) has shown that attractive (vs. unattractive) persons are perceived by others as being more internal. In the present study, an analysis of communicators' locus-of-control scores yielded no significant effects. It should also be noted that locus of control was not significantly associated with an index of communicator persuasiveness (sum of targets' agreement scores), $r = -.15$, $p < .20$, although the direction of the relationship is consistent with Phares' (1965) results.

On communicator-subjects' self-reported SAT scores, both an attractiveness main effect ($p < .05$) and a Communicator Attractiveness × Communicator Sex interaction ($p < .05$) were obtained. Overall, attractive communicators reported higher SAT scores

than did unattractive ones. However, this difference was most evident for female (vs. male) subjects. Attractive (vs. unattractive) communicators also reported marginally higher grade point averages ($p = .11$).

Although attractive communicators tended to regard both themselves and their future more positively than unattractive communicators did on many (but not all) of the self-descriptive items, the attractiveness main effect reached significance or marginal significance on only a very few scales. Physically attractive (vs. unattractive) communicators tended to regard themselves as more persuasive ($p < .025$), attractive ($p < .10$), and interesting ($p < .10$), and were somewhat more optimistic about obtaining an excellent job ($p < .10$). These analyses also revealed that female (vs. male) communicators viewed themselves as more moral (vs. immoral; $p < .01$) and indicated more optimism with respect to obtaining an excellent job ($p < .05$) and being a successful person ($p < .05$). Finally, a Communicator Sex × Communicator Attractiveness interaction was obtained on communicators' speculations regarding their future contentedness ($p < .05$). Attractive females and unattractive males speculated that they would be more contented than discontented whereas unattractive females and attractive males speculated that they would be more discontented.

These results indicate that attractive and unattractive individuals do differ on dimensions other than physical appearance. Attractive communicators were more fluent speakers and faster speakers than their unattractive counterparts. Further, attractive communicators tended to report higher scores on two indices of educational accomplishment (grade point average, SAT scores) and described themselves somewhat more favorably along several dimensions (persuasiveness, attractiveness, interestingness, optimism about getting an excellent job) that may tap aspects of self-concept.

The remaining question is whether the differential persuasiveness of physically attractive and unattractive communicators can be understood in terms of the observed differences with respect to communication skills, educational accomplishment, and components of self-concept. Correlations between communicators' scores on these measures and their persuasive effectiveness were generally low and nonsignificant (ps = .07 or higher). Although the relationship between communicators' attractiveness scores and persuasive effectiveness was low to begin with ($r = .20$, $p = .05$), it did attenuate somewhat when the influence of these individual difference measures was removed ($r = .14$, $p = .14$). These results suggest that while the observed individual differences do not provide a full explanation for the observed effect of attractiveness on persuasion, they may have contributed at least partially to this effect. In any case, it would seem premature to discount such individual difference variables as correlates of communicator persuasiveness or, by extension, to discount the role of these variables in understanding the attractiveness–persuasion relationship. Research has demonstrated that both fluent speech (McCroskey & Mehrley, 1969) and speech rate (Miller et al., 1976) relate positively to persuasion. And although little is known concerning the psychology of the persuasive communicator (in contrast to the psychology of the message recipient), it seems reasonable that factors such as self-concept and educational achievement, a frequently used indicant of intelligence, should contribute to one's effectiveness as a social influence agent.

Conclusions and Implications

The present research indicates that physical attractiveness can significantly enhance communicator persuasiveness. A number of aspects of the findings and the experimental design are worth noting, since they suggest that the attractiveness–persuasion relationship may be fairly general. First, the attractiveness effect on persuasion was obtained on both a verbal and behavioral measure, and attractiveness did not interact with either communicator or target sex to affect agreement. Second, the design featured multiple-stimulus persons at every level of communicator attractiveness and sex, and this nested communicator factor accounted for no significant effects in the

analyses of variance. Third, communicator-subjects were not judged to be extreme in physical attractiveness. Thus, the persuasion results were obtained using attractive and unattractive communicators who are probably quite representative of those whom individuals might encounter in their everyday lives. Last, the findings were obtained in a field setting where the salience of information about physical attractiveness should approximate its salience in genuine interpersonal situations and where any implicit experimental demands on the subject to overutilize or underutilize attractiveness information should be minimal.

Despite the study's apparent success in demonstrating the attractiveness–persuasion relationship, a number of issues remain unresolved and require further investigation. First, the interpersonal-orientation explanation of the observed sex differences in persuasibility raises the possibility that the enhanced persuasion induced by attractive communicators may have represented heightened compliance rather than genuine belief change. While the field data seem more compatible with a belief-change interpretation, compliance cannot be completely eliminated as an explanation. Research that manipulated communicator presence and/or surveillance in conjunction with physical attractiveness would be quite useful in resolving this issue. A second issue concerns the psychological mechanisms underlying the attractiveness–persuasion relationship. The results of this experiment as well as those of some previous studies (Norman, 1976; Snyder & Rothbart, 1971) suggest that credibility enhancement does not typically mediate the attractiveness–persuasion relationship. However, mechanisms such as identification (Kelman, 1961), social reinforcement, cognitive balance, classical conditioning, and perhaps others remain possible and require further evaluation.

Finally, there is the question of why most previous experiments have failed to demonstrate the attractiveness effects on persuasion observed here. One answer is suggested by the fact that the attractiveness effects obtained in this study, while statistically reliable, were not very strong. If, as the magnitude of these effects imply, the persuasive impact of attractiveness is not particularly robust, the failure of previous efforts to document such a relationship becomes more understandable. One obvious implication of this explanation is that while physical attractiveness can be a significant determinant of persuasion, it may not be a particularly important determinant. A second possible explanation stems from the laboratory nature of previous research. As suggested earlier, the implicit demands of the psychology laboratory may often lead subjects to adopt highly logical modes of cognitive functioning and, as a consequence, to underutilize attractiveness information in forming their opinions. The ecological validity of the laboratory for studying the attractiveness–persuasion relationship requires further investigation, perhaps through research that explicitly incorporates social-influence setting (field vs. laboratory) as a feature of its experimental design.

Individual Differences Between Attractive and Unattractive Communicators

The experiment also revealed differences between attractive and unattractive communicator-subjects with respect to characteristics relevant to persuasive effectiveness (communication skills, educational accomplishment, components of self-concept). These results provide preliminary evidence that in genuine interpersonal situations, attractive individuals may be more persuasive than unattractive persons because they possess characteristics or skills that dispose them to be particularly effective communicators. In this regard, it should be noted that in a recent investigation of manipulative social influence among children, Dion and Stein (1978) found that attractive and unattractive children were both differentially successful and employed markedly different interaction styles in attempting to influence peers.

Previous investigators have studied physical attractiveness in isolation from, rather than together with, its concomitant variables. The present findings and those of Dion and Stein (1978) underscore the utility of the latter

approach. Future research should address itself to a more systematic investigation of individual differences between the physically attractive and unattractive with respect to variables relevant to social influence. For example, research on the physical attractiveness stereotype indicates that attractive (vs. unattractive) individuals are perceived by others as more likable, friendly, interesting, and poised (cf. Berscheid & Walster, 1974). Such attributes would seem of obvious advantage to someone wishing to influence others. Further research might attempt to assess whether physically attractive people actually behave in a more likable, friendly, interesting, and poised fashion with others—either because they possess these traits or, as Snyder, Tanke, and Berscheid (1977) have recently argued, because the stereotype-based expectations of others tend to elicit and maintain such behaviors—and whether such behaviors can help to explain their greater effectiveness as agents of social influence. In addition to furthering our understanding of the relationship between attractiveness and social influence, such research would provide valuable information regarding the validity of the physical attractiveness stereotype and would also increase our knowledge about communicator characteristics and behavioral styles and their role in social influence.

Reference Notes

1. Blass, T., Alperstein, L., & Block, S. H. Effects of communicator's race and beauty and of receiver's objectivity–subjectivity on attitude change. In C. Ward (Chair), *Attitude change*. Symposium presented at the meeting of the American Psychological Association, New Orleans, 1974.
2. Eagly, A. H. Personal communication, January 5, 1979.

References

Berscheid, E., & Walster, E. Physical attractiveness. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 7). New York: Academic Press, 1974.

Chaiken, S., Eagly, A. H., Sejwacz, D., Gregory, W. L., & Christensen, D. Communicator physical attractiveness as a determinant of opinion change. JSAS *Catalog of Selected Documents in Psychology*, 1978, *8*. (Ms. No. 1639)

Clifford, M. M., & Walster, E. The effect of physical attractiveness on teacher expectation. *Sociology of Education*, 1973, *46* (Spring), 248–258.

Dion, K. K. Physical attractiveness and evaluations of children's transgressions. *Journal of Personality and Social Psychology*, 1972, *24*, 207–213.

Dion, K. K., & Stein, S. Physical attractiveness and interpersonal influence. *Journal of Experimental Social Psychology*, 1978, *14*, 97–108.

Eagly, A. H. Sex differences in influenceability. *Psychological Bulletin*, 1978, *85*, 86–116.

Goldman, W., & Lewis, P. Beautiful is good: Evidence that the physically attractive are more socially skillful. *Journal of Experimental Social Psychology*, 1977, *13*, 125–130.

Horai, J., Naccari, N., & Fatoullah, E. The effects of expertise and physical attractiveness upon opinion agreement and liking. *Sociometry*, 1974, *37*, 601–606.

Kelman, H. C. Processes of opinion change. *Public Opinion Quarterly*, 1961, *25*, 57–78.

Langer, E. J., & Abelson, R. P. The semantics of asking a favor: How to succeed in getting help without really dying. *Journal of Personality and Social Psychology*, 1972, *24*, 26–32.

London, H. *Psychology of the persuader*. Morristown, N.J.: General Learning Press, 1973.

Lunney, G. H. Using the analysis of variance with a dichotomous dependent variable: An empirical study. *Journal of Educational Measurement*, 1970, *7*, 263–269.

McCroskey, J. C., & Mehrley, R. J. The effects of disorganization and nonfluency on attitude change and source credibility. *Speech Monographs*, 1969, *36*, 13–21.

Mehrabian, A., & Williams, M. Nonverbal concomitants of perceived and intended persuasiveness. *Journal of Personality and Social Psychology*, 1969, *13*, 37–58.

Miller, A. G. Social perception of internal–external control. *Perceptual and Motor Skills*, 1970, *30*, 103–109.

Miller, N. Involvement and dogmatism as inhibitors of attitude change. *Journal of Experimental Social Psychology*, 1965, *1*, 121–132.

Miller, N., Maruyama, G., Beaber, R. J., & Valone, K. Speed of speech and persuasion. *Journal of Personality and Social Psychology*, 1976, *34*, 615–624.

Mills, J., & Aronson, E. Opinion change as a function of the communicator's attractiveness and desire to influence. *Journal of Personality and Social Psychology*, 1965, *1*, 173–177.

Myers, J. L. *Fundamentals of experimental design* (2nd ed.). Boston: Allyn & Bacon, 1972.

Norman, R. When what is said is important: A comparison of expert and attractive sources. *Journal of Experimental Social Psychology*, 1976, *12*, 294–300.

Phares, E. J. Internal–external control as a determinant of amount of social influence exerted. *Journal of Personality and Social Psychology,* 1965, *2,* 642–647.

Rotter, J. B. Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs,* 1966, *80*(1, Whole No. 609).

Sherif, M., & Hovland, C. I. *Social judgment.* New Haven, Conn.: Yale University Press, 1961.

Snyder, M., & Rothbart, M. Communicator attractiveness and opinion change. *Canadian Journal of Behavioral Science,* 1971, *3,* 377–387.

Snyder, M., Tanke, E. D., & Berscheid, E. Social perception and interpersonal behavior: On the self-fulfilling nature of social stereotypes. *Journal of Personality and Social Psychology,* 1977, *35,* 656–666.

Received August 25, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 8, 1398-1405

A Creative Personality Scale for the Adjective Check List

Harrison G. Gough
Institute of Personality Assessment and Research
University of California, Berkeley

The Adjective Check List was administered to seven male and five female samples comprising 1,701 subjects. Direct or inferred ratings of creativity were available for all individuals. The samples covered a wide range of ages and kinds of work; criteria of creativity were also varied, including ratings by expert judges, faculty members, personality assessment staff observers, and life history interviewers. The creativity scales of Domino and Schaefer were scored on all protocols, as were Welsh's A-1, A-2, A-3, and A-4 scales for different combinations of "origence" and "intellectence." From item analyses a new 30-item Creative Personality Scale was developed. It is positively and significantly ($p < .01$) related to all six of the prior measures but surpasses them in its correlations with the criterion evaluations.

Creativity is a valued commodity in every kind of human endeavor. Since the publication of Guilford's (1950) influential presidential address to the American Psychological Association, an enormous amount of effort has been invested in the study of creativity and its determinants. One line of investigation within the larger domain of inquiry has been the search for methods of assessment that can identify creative talent and potential within the individual. Many of these studies have addressed cognitive issues and problem solving. For example, Guilford and his colleagues (Guilford, Wilson, Christensen, & Lewis, 1951) developed a series of tests stressing ingenuity, the ability to overcome constraining sets, and fluency in ideation. Mednick (1962) proposed a method of assessment requiring the generation of remote associations for the solution of analogies.

In regard to intellectual functioning, it should be noted that most studies have found intellectual ability as usually measured to be unrelated to criteria of originality. MacKinnon and Hall (1972) obtained Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1958) protocols from 88 architects, 37 research scientists, 33 male mathematicians, and 27 female mathematicians who had also been rated on creativity. Within each sample, subjects were dichotomized into those with higher or lower ratings. The higher rated subgroup, taken from all four samples, had a mean full-scale IQ of 133, and the lower rated subgroup had a mean IQ of 131. The difference between the means was not statistically significant, nor were there any noteworthy differences in range or dispersal of scores. In their study of gifted students, Getzels and Jackson (1962) found intelligence to play a smaller role than personality in determining creativity, and Taylor (1960) also found general intellectual ability to be less important than special kinds of thinking and motivational factors.

Ordinary observations have long suggested that artistic temperament and aesthetic dispositions are related to creative potential. A landmark study of this hypothesis was that of Barron and Welsh (1952) in which a nonverbal figure-preference scale was introduced. The original Barron-Welsh Art Scale and the Revised Art Scale (Welsh, 1969, 1975), in which *like* and *dislike* responses are balanced, have repeatedly been shown to differentiate between more and less creative persons in various scientific, literary, and artistic fields (Barron, 1972; Welsh, 1977). Another ex-

Requests for reprints should be sent to Harrison G. Gough, Institute of Personality Assessment and Research, University of California, Berkeley, California 94720.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3708-1398$00.75

ample of measurement in the aesthetic domain is the Hall Mosaic Construction Test (Hall, 1972).

Personal traits and dispositions have also been examined in regard to creativity, using both standard personality inventories and specially developed scales and questionnaires. Examples of this kind of work may be found in the writings of Barron (1957, 1958), Domino (1974), Gough (1956/1962, 1976), Helson (1977), Kanner (1976), MacKinnon (1962, 1965), and Stein and Heinze (1960). An overall analysis of intellectual, aesthetic, motivational, and other kinds of tests for creativity has been published by Barron (1965).

A particular topic within the realm of studies of personality is that dealing with the self-concept. The Adjective Check List (ACL; Gough & Heilbrun, 1965) is an assessment device intended for appraising views of the self, and as might be expected it has frequently been employed in investigations of creativity (Cashdan & Welsh, 1966; MacKinnon, 1963; Schaefer, 1969). Several attempts, in fact, have been made to develop creativity scales for the ACL. Schaefer (Smith & Schaefer, 1969) identified 27 items that differentiated between the responses of high school boys rated as more and less creative. Only one of these items (cooperative) was checked more often by those with lower ratings; the other 26 items were more often endorsed by higher rated respondents. Follow-up studies (Schaefer, 1972, 1973) indicated that the scale retained its validity over time.

Domino (1970) asked faculty members of a liberal arts college to identify all male freshmen who had manifested creative ability. Ninety-six students were selected; these students were then matched against 96 unnominated controls on age, IQ, personal adjustment as estimated from the Minnesota Multiphasic Personality Inventory (MMPI), and declared major. In the next year, faculty members were asked to make a special effort to observe the performance of all of these students and to judge which ones had shown evidence of creative ability; this procedure was repeated in the third year. These steps yielded a final sample of 59 creative males and a control sample of 82. For each student, at least three faculty raters completed descriptive ACLs. Composites were formed for the 141 students by assigning a 1 to items checked by two or more raters, and a 0 to items checked by none or by only one observer. An item analysis of the 300 adjectives in the list identified 68 that differentiated between the two subsamples at the .05 level of probability or beyond. Fifty-nine of these were more often used to describe the creative subsample and 9 were used more often to describe the controls. To simplify scoring and analysis, Domino decided to base his scale solely on the 59 items associated with greater creativity. Cross-validation of the scale on the self-report protocols of new samples of males and females produced correlations ranging from .24 to .45 with criterion classifications of creativity.

A third study in which item analyses were made of the ACL was that of Welsh (1975). Two dimensions were first defined, one deriving from intellectual functioning and behavior and termed *intellectence,* and the other deriving from originality and aesthetic sophistication and termed *origence*. The interactive grid for these two dimensions permitted the specification of four types of cognitive functioning (Welsh, 1977). Type 1, high on origence but low on intellectence, was characterized by diffuse, global, and imprecise integration with little or no differentiation. Type 2, high on both origence and intellectence, was characterized by synthesis, organization, and the cathexis of metaphor. Type 3, low on both origence and intellectence, was characterized by fragmentation of elements and overattention to details. Type 4, low on origence but high on intellectence, was characterized by analytic and logical preferences. The four types could be very briefly described as imaginative, intuitive, conventional, and analytic in their cognitive styles. Adjective Check List scales were developed for each quadrant by contrasting the responses of individuals in that cell with those of individuals in the other three. The A-1 scale for respondents high on origence but low on intellectence contained 21 items. The A-2 scale for respondents high on both axes contained 25 items. The A-3 and

A-4 scales contained 17 and 24 items, respectively. Scoring on all four scales was by endorsement only; no points were given for nonendorsement of contraindicative adjectives.

During the late 1950s, the writer directed a study of creativity among research scientists (Gough & Woodworth, 1960). When the six ACL scales described above were applied to the 45 scientists in this study, rather disappointing results were obtained. The correlations with the criterion ratings of creativity were .01 for Domino's scale, .08 for Schaefer's, and −.06, −.01, −.08, and .08 for the four measures developed by Welsh. Similar results were obtained when the six scales were scored on the self-report ACLs of the 57 male mathematicians studied by Helson and Crutchfield (1970). Correlations with criterion ratings of creativity were −.11, −.03, .05, −.08, .05, and −.08 for the six ACL scales, in the same sequence. Because of these inconclusive findings, it was decided to undertake a new analysis of the ACL, using larger samples and a broader range of criteria, to see if a stronger measure could be developed.

Method

Samples

Six male samples were available in which the ACL had been administered and criterion ratings of creativity had been gathered. The first was composed of 124 architects studied by MacKinnon and Hall (Hall & MacKinnon, 1969). The second was comprised of the 57 male mathematicians just mentioned (Helson & Crutchfield, 1970). The third included the 45 research scientists also just mentioned (Gough & Woodworth, 1960). The fourth was a previously unreported sample of 530 graduate students in psychology at Berkeley, tested during the first week of entry and rated 3 or 4 years later. The fifth was a sample of 66 seniors in engineering (Gough, 1976). The sixth was a composite sample of 256 men who had been intensively assessed at the Institute of Personality Assessment and Research and rated by staff observers on creativity. This composite sample included 100 Air Force officers (MacKinnon, Crutchfield, Barron, Block, Gough, & Harris, 1958), 70 medical school applicants (Gough & Hall, 1973), 20 college sophomores from an unpublished study of vocational and career planning, 41 males from a study of population psychology (Gough, 1973), and 25 males from a study of environmental preferences (Craik, Note 1).

There were four samples of women for which ACL protocols and ratings of creativity were available. The first was comprised of the 41 women mathematicians studied by Helson (1971). The second included 335 graduate students in psychology at Berkeley, tested at the time of entry and rated between 3 and 4 years later by faculty members. The third was composed of the 51 college seniors reported on earlier by Helson (1967). The fourth was a composite sample of 126 women, including 20 college sophomores from the study of career planning, 41 women from the study of population psychology, 25 women from the study of environmental preferences, and 40 first-year students of law (LaRussa, 1977).

In addition to these 1,631 subjects, there were 35 males and 35 females for whom ACL protocols were available and who had been interviewed by two psychologists and described by them on the ACL and on Block's (1961) 100-item California Q-set. Although direct ratings of creativity were not obtained, indirect estimates were derived from the ACL and Q-sort portraits furnished by the interviewers, in a way to be described below.

Tests and Criteria

The ACL self-report protocols of all subjects were scored for the creativity scales of Domino and Schaefer and for Welsh's four scales. Four kinds of criterion evaluations were utilized. For the architects, male and female mathematicians, and research scientists, creativity was specified by ratings furnished by expert judges. The validity and reliability of these ratings are discussed in the papers already cited. Ratings by faculty members supplied the criterion for the engineering students, psychology graduate students, and college seniors. The psychology graduate students had entered over a 24-year period. Faculty ratings were obtained every 3 or 4 years, covering students who had entered in the preceding interval. Corrected interjudge reliabilities for these ratings of creativity ranged from .73 to .87, with a median of .77. Equivalent coefficients were found for the engineers and college seniors.

The 256 males and 126 females seen at the institute in intensive programs of assessment were rated on creativity by panels of 10 or more observers. Corrected interjudge reliabilities for these ratings were typically very high, ranging from .80 to .98.

To develop a criterion for the 70 interviewed subjects, the interviewers' checks on five adjectives and placement of five Q-sort items were summed. The five adjectives were *imaginative, insightful, intelligent, original,* and *resourceful.* There were four Q-sort items given positive weighting: *Has a wide range of interests, Appears to have a high degree of intellectual capacity, Thinks and associates to ideas in unusual ways; has unconventional thought processes,* and *Able to see to the heart of important problems.* One Q-sort item was assigned a negative weight: *Is uncomfortable with uncertainty and complexities.* The total score based on these 10 components had a standardized item alpha reliability coefficient of .79.

Analyses

For purposes of item analysis, four subgroups were defined: (a) 558 males from the samples of architects, mathematicians, scientists, engineers, and other assessed males; (b) 530 male graduate students in psychology; (c) 218 females from the samples of mathematicians, college seniors, and other assessed females; and (d) 335 psychology graduate students. Ratings of creativity were converted to standard scores by subsample. Point-biserial coefficients of correlation were calculated between each of the 300 items in the ACL and the criterion standard scores.

Results

Thirty items were selected for inclusion in the Creative Personality Scale (CPS) on the basis of item analysis. Several examples may be given. The item *egotistical* yielded correlations with the criterion ratings of creativity of .17, .06, .08, and .08 in the four subgroups. Although only the first of these coefficients is statistically significant ($p < .01$), all are positive, and the item itself is consonant with prior conceptualizations of the creative personality. The item also appears on Domino's scale. The item *original* produced correlations of .13, .06, .17, and .11. Two of these coefficients are statistically significant ($p < .05$), and the word appears on both the Schaefer and Domino scales. It is also quite clearly consonant with prior conceptualizations of the creative personality.

The item *conservative* had correlations with the criterion of −.16, −.08, −.13, and −.10. Two of these coefficients are significant at the .05 level of probability, and one is significant at $p = .06$. Because of these findings and because absence of the attribute appears to be consonant with previous conceptualizations of the creative personality, the item was retained with a negative weighting. In this way, 18 positive and 12 negative items were selected for the final scale. The positive items were *capable, clever, confident, egotistical, humorous, individualistic, informal, insightful, intelligent, interests wide, inventive, original, reflective, resourceful, self-confident, sexy, snobbish,* and *unconventional.* The negatively weighted items were *affected, cautious, commonplace, conservative, conventional, dissatisfied, honest, interests narrow, mannerly, sincere, submissive,* and *suspicious.* In scoring a protocol, 1 point is given each time one of the 18 positive items is checked, and 1 point is subtracted each time one of the 12 negative items is checked. The theoretical range of scores is therefore from −12 to +18.

The 30 items were then compared with those in the previous six scales. Seven CPS items were found in Schaefer's 27-item scale, all scored in the same direction. Fifteen CPS items were found in Domino's 59-item scale; 14 were scored in the same direction, the exception being *dissatisfied,* which received a negative weight on CPS and a positive weight on Domino's measure. Two CPS items were found to overlap with Welsh's A-1 scale, 5 with his A-2 scale, 1 with his A-3 scale, and 2 with his A-4 measure. Except for *dissatisfied* on A-2, all of these common items were scored in the same direction. Because *dissatisfied* showed these two reversals, attention was redirected to the item analytic data. On the four subgroups the item had correlations of −.02, −.07, −.10, and −.02. Additional analyses were then computed on the male mathematicians, the research scientists, and the college seniors, where the coefficients were −.17, −.36, and −.10. Because of these findings it was decided to retain the item with a negative scoring weight.

The 30-item Creative Personality Scale (CPS) was then scored on all of the samples included in the present study. Alpha coefficient reliabilities were computed on the four subgroups defined for the item analysis. The coefficients were .77 for the male composite group, .73 for the male graduate students, .81 for the female composite group, and .73 for the female graduate students. Means and standard deviations for each sample are given in Table 1. Among the male samples, the highest mean was attained by the research scientists, followed closely by the psychology graduate students. The lowest mean was that for the 35 males in the sample seen in interviews only. Among the females, the highest mean was that for the female graduate students, and the lowest was that for the women seen only in interviews. Where comparisons seem appropriate, as between the two samples of graduate students and the two samples of

interviewed subjects, CPS scores were significantly ($p < .05$), albeit only slightly, higher for males than for females.[1]

Table 2 gives the correlations among the seven ACL scales, computed on the samples of male and female graduate students in psychology. The median coefficients between CPS and each of the other measures were .68 for Domino's scale for creativity, .74 for Schaefer's scale, .42 Welsh's A-1, .51 for Welsh's A-2, .32 for Welsh's A-3, and .32 for Welsh's A-4. For the other measures, the highest median correlations were those of .89 between the Domino and Schaefer scales, .64 between A-1 and the Domino and Schaefer scales, .82 between A-2 and the Domino and Schaefer scales, .64 between A-3 and A-4, and .64 between A-4 and the Domino and Schaefer scales. The presence of the 12 negatively weighted items in CPS appears to be the reason why the other six scales correlate more highly among themselves than they do with CPS.

Table 3 gives the correlations of the seven ACL scales with criterion ratings for creativity. For all seven of the male samples, the highest coefficient each time was that for CPS, and in six of the seven instances the coefficients for CPS were significant at or beyond the .05 level of probability. The exception was the coefficient of .25 for the 45 research scientists, where $p = .097$ in a two-tailed test. The smallest coefficient for CPS was that of .15 for the psychology graduate students, and the largest was that of .42 for the 256 males rated by observers in the various assessments. On

Table 1

Means and Standard Deviations for the Samples Indicated on the Adjective Check List Creative Personality Scale

Sample	*n*	*M*	*SD*
Male			
Architects	124	5.28	3.86
Mathematicians	57	4.44	4.20
Research scientists	45	5.98	3.71
Psychology graduate students	530	5.96	3.86
Engineering students	66	3.88	3.94
Assessed males	256	3.57	3.99
Interviewed males	35	2.00	3.01
Female			
Mathematicians	41	3.34	4.45
Psychology graduate students	335	5.43	3.88
College seniors	51	5.10	4.24
Assessed females	126	4.40	4.07
Interviewed females	35	0.00	3.25

Table 2

Intercorrelations Among Seven Adjective Check List Scales in Samples of 530 Male and 335 Female Graduate Students in Psychology

Scale	Correlation					
	2	3	4	5	6	7
1. Domino creativity						
Male	.89	.64	.82	.51	.63	.69
Female	.89	.64	.81	.54	.65	.67
2. Schaefer creativity						
Male	—	.63	.76	.47	.55	.73
Female	—	.61	.73	.46	.48	.76
3. Welsh A-1						
Male		—	.62	.51	.44	.45
Female		—	.62	.46	.35	.40
4. Welsh A-2						
Male			—	.31	.43	.53
Female			—	.36	.42	.49
5. Welsh A-3						
Male				—	.69	.32
Female				—	.61	.33
6. Welsh A-4						
Male					—	.34
Female					—	.30
7. Creative Personality						
Male						—
Female						—

Note. All coefficients are statistically significant beyond the .01 level of probability. A-1 = Welsh high origence, low intellectence; A-2 = Welsh high origence, high intellectence; A-3 = Welsh low origence, low intellectence; A-4 = Welsh low origence, high intellectence.

[1] For purely normative purposes, Adjective Check List protocols for 1,121 college females and 760 college males from the author's research files were scored for the new 30-item scale. The mean of 5.03 ($SD = 4.01$) for males was significantly higher ($p < .01$) than that of 3.97 ($SD = 4.34$) for females. This difference suggests that where raw scores on the new scale are used, analyses should be conducted separately for males and females. It should be noted that most studies with the ACL use scale scores that have been standardized by sex.

Table 3
Correlations of Selected Scales From the Adjective Check List With Criterion Ratings of Creativity

Sample	*n*	Scale and correlation						
		D	S	A-1	A-2	A-3	A-4	CPS
Males								
Architects[a]	124	.35**	.36**	.16	.34**	−.13	−.10	.39**
Mathematicians[a]	57	−.11	−.03	.05	−.08	.05	−.08	.31*
Research scientists[a]	45	.01	.08	−.06	−.01	−.08	.08	.25
Psychology graduate students[b]	530	.01	.01	−.06	.01	−.09*	−.05	.15**
Engineering students[b]	66	.27*	.33*	.03	.26*	−.04	.09	.28*
Assessed males[c]	256	.34**	.34**	.16*	.25**	.11	.23**	.42**
Interviewed males[d]	35	.05	.13	−.24	.02	.21	.05	.35*
Females								
Mathematicians[a]	41	−.16	−.07	−.12	.07	−.42**	−.40**	.28
Psychology graduate students[b]	335	.14**	.17**	−.02	.07	.02	.07	.25**
College seniors[b]	51	.12	.11	−.02	−.02	.10	.14	.27*
Assessed females[c]	126	.29**	.32**	.04	.22*	.06	.21*	.40**
Interviewed females[d]	35	.27	.34*	.05	.21	.07	.35*	.40*

Note. D = Domino creativity; S = Schaefer creativity; A-1 = Welsh high origence, low intellectence; A-2 = Welsh high origence, high intellectence; A-3 = Welsh low origence, low intellectence; A-4 = Welsh low origence, high intellectence; CPS = Creative Personality Scale.
[a] Criteria = ratings by expert judges.
[b] Criteria = ratings by faculty members.
[c] Criteria = ratings by assessment staff.
[d] Criteria = ratings by interviewers.
* $p < .05$.
** $p < .01$.

the first six male samples, higher coefficients for CPS would be expected because these samples were used in the item analyses. The 35 interviewed males, however, were not used in the item studies and therefore constitute a true cross-validating sample. In this sample the only statistically significant coefficient was that of .35 for CPS.

For the five samples of women, the coefficient for CPS was the highest in four instances and statistically significant at the .05 level each time. The exception occurred in the sample of 41 women mathematicians. For these subjects, Welsh's A-3 scale had a statistically significant ($p < .01$) coefficient of −.42, as did his A-4 scale, for which the coefficient was −.40. The coefficient for CPS on this sample was .28, $p = .075$ in a two-tailed test. The last sample reported in Table 3, the 35 women seen only in life history interviews, constituted a true cross-validating sample. For these subjects, there were three statistically significant ($p < .05$) correlations between scales and the derived criterion rating of creativity: .34 for Schaefer's scale, .35 for Welsh's A-4 scale, and .40 for CPS.

In the six male samples from which item-analytic data were drawn, the median coefficient between CPS and the criterion ratings of creativity was .30, and in the small cross-validating sample of 35 males, the correlation between CPS and the inferred rating of creativity was .35. A similar summary for the female samples yielded a median coefficient of .28 for the samples used in the item analyses and a coefficient of .40 for the small cross-validating sample. These samples covered a wide range of ages, kinds of work, and circumstances of testing. They also involved the use of four vantage points in rating creativity; expert judges, faculty members, personality-assessment staff observers, and life-history interviewers. If one-tailed tests of significance are allowed, the new 30-item scale can claim a

statistically significant relationship with every criterion in every sample. If the two-tailed test is used, 10 of the 12 validity coefficients for CPS were significant at the .05 level of probability or beyond. On the basis of these findings, it seems reasonable to conclude that the new scale is a reliable and moderately valid measure of creative potential and that it may properly be included among the scales to be scored on the Adjective Check List.

Reference Note

1. Craik, K. H. Impression of a place: Effects of media, context, and personality. In S. Saegert (Chair), *Psychology of the Urban Environment*. Symposium presented at the annual meeting of the American Psychological Association, Toronto, August 1978. (Copies available from the Institute of Personality Assessment and Research, University of California, Berkeley, California 94720.)

References

Barron, F. Originality in relation to personality and intellect. *Journal of Personality*, 1957, *25*, 730–742.

Barron, F. The psychology of imagination. *Scientific American*, 1958, *199*, 150–156.

Barron, F. The psychology of creativity. In T. M. Newcomb (Ed.), *New directions in psychology: II*. New York: Holt, Rinehart & Winston, 1965.

Barron, F. *Artists in the making*. New York: Seminar Press, 1972.

Barron, F., & Welsh, G. S. Artistic perception as a factor in personality style: Its measurement by a figure-preference test. *Journal of Psychology*, 1952, *33*, 199–203.

Block, J. *The Q-sort method in personality assessment and psychiatric research*. Springfield, Ill.: Charles C Thomas, 1961.

Cashdan, S., & Welsh, G. S. Personality correlates of creative potential in talented high school students. *Journal of Personality*, 1966, *34*, 445–455.

Domino, G. Identification of potentially creative persons from the Adjective Check List. *Journal of Consulting and Clinical Psychology*, 1970, *35*, 48–51.

Domino, G. Assessment of cinematographic creativity. *Journal of Personality and Social Psychology*, 1974, *30*, 150–154.

Getzels, J. W., & Jackson, P. W. *Creativity and intelligence: Explorations with gifted students*. New York: Wiley, 1962.

Gough, H. G. Imagination—undeveloped resource. In *Proceedings of the first annual conference on research developments in personnel management*. Los Angeles: University of California, Institute of Industrial Relations, 1956. (Reprinted in S. J. Parnes & H. F. Harding [Eds.], *A source book for creative thinking*. New York: Scribner's, 1962).

Gough, H. G. Personality assessment in the study of population. In J. T. Fawcett (Ed.), *Psychological perspectives on population*. New York: Basic Books, 1973.

Gough, H. G. Studying creativity by means of word association tests. *Journal of Applied Psychology*, 1976, *61*, 348–353.

Gough, H. G., & Hall, W. B. A prospective study of personality changes in students in medicine, dentistry, and nursing. *Research in Higher Education*, 1973, *1*, 127–140.

Gough, H. G., & Heilbrun, A. B., Jr. *The Adjective Check List manual*. Palo Alto, Calif.: Consulting Psychologists Press, 1965.

Gough, H. G., & Woodworth, D. G. Stylistic variations among professional research scientists. *Journal of Psychology*, 1960, *49*, 87–98.

Guilford, J. P. Creativity. *American Psychologist*, 1950, *5*, 444–454.

Guilford, J. P., Wilson, R. C., Christensen, P. R., & Lewis, D. J. A factor analytic study of creative thinking: I. Hypotheses and descriptions of tests. *Reports from the Psychological Laboratory* (No. 4). Los Angeles: University of Southern California, 1951.

Hall, W. B. A technique for assessing aesthetic predispositions: Mosaic Construction Test. *Journal of Creative Behavior*, 1972, *6*, 225–235.

Hall, W. B., & MacKinnon, D. W. Personality inventory correlates of creativity among architects. *Journal of Applied Psychology*, 1969, *53*, 322–326.

Helson, R. Personality characteristics and developmental history of creative college women. *Genetic Psychology Monographs*, 1967, *76*, 205–256.

Helson, R. Women mathematicians and the creative personality. *Journal of Consulting and Clinical Psychology*, 1971, *36*, 210–220.

Helson, R. The creative spectrum and authors of fantasy. *Journal of Personality*, 1977, *45*, 310–326.

Helson, R., & Crutchfield, R. S. Mathematicians: The creative researcher and the average PhD. *Journal of Consulting and Clinical Psychology*, 1970, *34*, 250–257.

Kanner, A. D. Femininity and masculinity: Their relationship to creativity in male architects and their independence from each other. *Journal of Consulting and Clinical Psychology*, 1976, *44*, 802–805.

LaRussa, G. W. Portia's decision: Women's motives for studying law and their later career satisfaction as attorneys. *Psychology of Women Quarterly*, 1977, *1*, 350–364.

MacKinnon, D. W. The nature and nurture of creative talent. *American Psychologist*, 1962, *17*, 484–495.

MacKinnon, D. W. Creativity and images of the self. In R. W. White (Ed.), *The study of lives: Essays on personality in honor of Henry A. Murray*. New York: Atherton Press, 1963.

MacKinnon, D. W. Personality and the realization of creative potential. *American Psychologist*, 1965, *20*, 273–281.

MacKinnon, D. W., Crutchfield, R. S., Barron, F., Block, J., Gough, H. G., & Harris, R E. *An assessment study of Air Force officers. Part I: Design of the study and description of the variable* (USAF WADC Tech. Rep. 58-91, ASTIA Doc. No. AD 151 050). Lackland Air Force Base, Texas: Personnel Laboratory, 1958.

MacKinnon, D. W., & Hall, W. B. Intelligence and creativity. In H. W. Peter (Chair), The measurement of creativity. Colloquium 17 in *Proceedings, XVII International Congress of Applied Psychology, Liege, Belgium, 25th–30th July 1971* (Vol. 2). Brussels: Editest, 1972.

Mednick, S. A. The associative basis of the creative process. *Psychological Review,* 1962, *69,* 220–232.

Schaefer, C. E. The self concept of creative adolescents. *Journal of Psychology,* 1969, *72,* 233–242.

Schaefer, C. E. Follow-up study of a creativity scale for the Adjective Check List. *Psychological Reports,* 1972, *30,* 662.

Schaefer, C. E. A five-year follow-up study of the self-concept of creative adolescents. *Journal of Genetic Psychology,* 1973, *123,* 163–170.

Smith, J. M., & Schaefer, C. E. Development of a creativity scale for the Adjective Check List. *Psychological Reports,* 1969, *25,* 87–92.

Stein, M. I., & Heinze, S. J. *Creativity and the individual.* Glencoe, Ill.: Free Press, 1960.

Taylor, C. W. Thinking and creativity. *Annals of the New York Academy of Science,* 1960, *91,* 108–127.

Wechsler, D. *The measurement and appraisal of adult intelligence* (4th ed.) Baltimore, Md.: Williams & Wilkins, 1958.

Welsh, G. S. *Gifted adolescents: A handbook of test results.* Greensboro, N.C.: Prediction Press, 1969.

Welsh, G. S. *Creativity and intelligence: A personality approach.* Chapel Hill, N.C.: Institute for Research in Social Science, 1975.

Welsh, G. S. Personality correlates of intelligence and creativity in gifted adolescents. In J. C. Stanley, W. C. George, & C. H. Solano (Eds.), *The gifted and the creative: A fifty-year perspective.* Baltimore, Md.: Johns Hopkins University Press, 1977.

Received October 3, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 8, 1406-1411

Conflict and Avoidance in the Helping Situation

S. Mark Pancer, Linda M. McMullen, Randal A. Kabatoff,
Kent G. Johnson, and Carole A. Pond
University of Saskatchewan, Saskatoon, Canada

Two field studies examined avoidance behavior in individuals asked for a charitable donation. Subjects were observed as they walked past a table placed in a tunnel leading to the university library. The first study showed that subjects tended to maintain a greater distance from the table when it was set up as a donation center than when it was not. In addition, subjects maintained greater distances when a female confederate sat at the table than when no one was at the table. Study 2 replicated the findings of Study 1. It found, in addition, that even greater distances were maintained from the table when a handicapped person sat at the table than when the person at the table was not handicapped. This avoidance behavior is discussed with reference to the kinds of conflict inherent in helping situations.

The helping situation is fraught with conflict. Usually there are strong normative pressures to help those in need and to give aid to those who have been innocently victimized (see Harris & Meyer, 1973; Konečni, 1972). Coupled with these pressures, however, are equally strong forces that tend to inhibit helping. Research has shown, for example, that appeals for help often evoke reactance (Jones, 1970). In certain situations, helping may also expose the helper to danger or to the ridicule of others (see Latané & Darley, 1976). Also, costs in time, effort, and money are often involved in giving help.

One consequence of the conflict inherent in many helping situations may be what Schwartz (1970) refers to as a "redefinition" of the helping situation. The individual redefines the situation to perceive himself/herself as less responsible for helping or the situation as less needful of help. This cognitive redefinition of the helping environment serves to reduce the press of social helping norms and thereby lessens the conflict intrinsic to such situations. Such redefinitions may often be difficult to accomplish, however, as demonstrated in the research of Darley and Latané (1968). They report anecdotal evidence of subjects under much emotional stress resulting from not helping someone who was quite obviously in need of help.

Another means of reducing helping-related conflict is simply to avoid such situations. Canvassers for various charities (e.g., Wilsdon, Note 1) have long noted a tendency in potential donors to ignore the existence of the charity worker or even to go some distance out of their way to avoid the donation site. Avoidance is psychologically perhaps the least costly of conflict-reduction strategies, since it spares the person from any cognitive distortions that might have been necessary to redefine the situation.

The purpose of the present study was to demonstrate the avoidance that can occur in helping situations and to determine some of the factors that may influence the extent of this avoidance. In the studies reported, sub-

This research was supported by Grant 3-771-531 from the University of Saskatchewan's Humanities and Social Sciences Research Fund. The authors thank Marvin Brown and Ken McDowell for their comments on earlier drafts of this article. We express special appreciation to the Saskatoon branches of United Nations Children's Fund (UNICEF) and the United Way for providing canvassing materials and for allowing us to canvass on their behalf.

Requests for reprints should be sent to S. Mark Pancer, Department of Psychology, University of Saskatchewan, Saskatoon, Saskatchewan, Canada S7N 0W0.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3708-1406$00.75

jects were observed as they walked past a table set up in a tunnel connecting the arts and library buildings on a university campus. It was hypothesized that subjects would show greater avoidance of the area around the table when it was set up as a donation center for a well-known charity than when it was not set up in this way.

Another variable examined was the presence or absence of an experimental confederate at the donation table. Research has shown that personal (or even personalized) requests are more effective in eliciting helping. This may be true for a number of reasons. Anonymous or impersonal requests (such as those represented by placing an unattended donation canister on a street corner) may not be perceived to be as strong as personal requests. Also, the person canvassing for donations may serve as a helping model. It has often been demonstrated that models serve to increase helping, presumably by making helping norms more salient (e.g., Bryan & Test, 1967). Although personal requests often elicit more helping, however, they may also produce greater avoidance. They are even more likely than impersonal requests to produce reactance in the potential donor because of the strong normative pressures implied. In addition, being approached or asked for help by a strange person may represent an invasion of one's personal space. Again, while this may produce more helping (Baron, 1978; Baron & Bell, 1976; Konečni, Libuser, Morton, & Ebbesen, 1975), it is also likely to produce more avoidance. The presence of a confederate at the donation table, then, may serve to intensify the conflicting demands present in the helping situation. This may result in an increased level of helping (and therefore a greater tendency to approach the table) or greater avoidance of the donation site.

Study 1

Method

Summary of design. The experiment was a 2 × 2 factorial. Subjects were monitored as they walked past a table that did or did not contain a United Way donation box (donation factor) and at which there was or was not a female confederate seated (person factor). The measure of avoidance was the distance subjects maintained from the table as they walked past.

Subjects. The subjects were 180 people (45 in each of the four experimental conditions) who were randomly selected as they walked through a tunnel that connected the main library to the arts building on the University of Saskatchewan campus. Only subjects walking alone from the library towards the arts building were monitored.

Experimental site. Approximately halfway through the tunnel, a chair and a table had been positioned along one wall (see Figure 1). The tunnel floor from the outside edge of the table to the opposite wall was sectioned into six lanes, each lane measuring 76 cm in width (approximately one body width). These lanes were marked by placing small pieces of black tape on a 2-cm expansion strip that ran the width of the tunnel, through the center of the table. Experimental observers sat in a lounge area approximately 6.5 m from the table, from which subjects could be clearly observed as they were both approaching and passing the table.

Procedure. The experiment was conducted during the afternoon of 1 day and the morning of the next. No subjects were observed within 10 minutes of breaks between classes, due to the large numbers of people using the tunnel at these times. This restriction also prevented individuals who were in a particular hurry to go to (or away from) class from being used as subjects. Six volunteer undergraduate observers (three on each of the 2 days) sat unobtrusively in the lounge area across from the table. As people approached the table from the library, each observer selected a subject who met the criterion of being alone, made certain that none of the other observers was monitoring that same subject, and recorded the lane through which the subject walked as he or she passed the table. As soon as a recording had been made, the observer selected the next subject walking alone and unmonitored by the other observers. Each observer monitored the same number of subjects under each of the four experimental conditions. Prior testing had indicated greater than 95% interrater agreement as to which lanes subjects had used in passing the table.

On each of the 2 days, the conditions were arranged in a different random order, and approximately half the subjects in each condition were observed on each day of the study. In the two person-present conditions, a female undergraduate confederate sat on the chair behind the table. She made no attempt to solicit donations but sat silently and watched the activity in the tunnel, avoiding direct eye contact with subjects who were being monitored. In the no-person-present conditions, the chair was empty.

In the two donation conditions, a cardboard donation box (34 × 23 × 47 cm) was positioned on the table near the end closest to approaching subjects. It was covered on all four sides and on the top with posters from the donation agency (United Way). Each of these posters consisted of a white back-

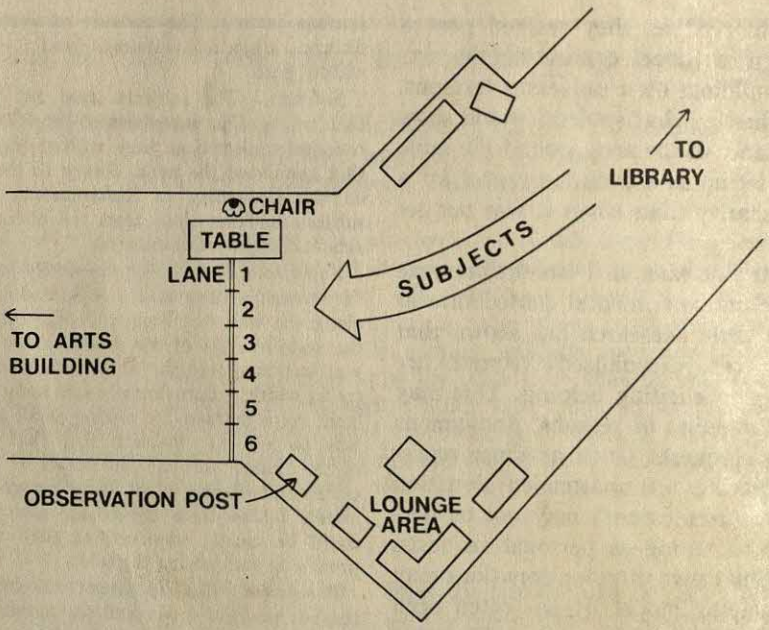


Figure 1. Experimental site.

ground with 2 cm of black lettering on the top and bottom and the agency's logo in orange in the center. The lettering on the top of each poster read "Thanks to you it's working," and on the bottom it read "United Way." In addition, a scaled-down version of the poster (17 × 25 cm) was placed behind the box so that it was clearly visible to people walking from the library. In the no-donation conditions, the table was bare.

Results and Discussion

Mean distances maintained from the table as subjects passed by are presented in Table 1.[1] Cochran's test for homogeneity of variance revealed significant differences among cell variances ($C = .366$, $p < .05$). Distance scores were therefore converted to reciprocals before being submitted to analysis of variance. Transformed scores appeared to be homogeneous ($C = .310$, *ns*).

Analysis of variance of the transformed data revealed a main effect for the donation factor, $F(1, 176) = 6.03$, $p < .025$. Subjects maintained a greater distance from the table when it displayed a poster and donation box than when no donation request was implied. A main effect was also obtained for the person factor, $F(1, 176) = 13.46$, $p < .001$, subjects maintaining a greater distance from the table when the female confederate was present than when no one was at the table. The Donation × Person interaction was not significant, $F(1, 176) = .40$, *ns*.

Only three subjects made donations, two in the person and one in the no-person condition. Analysis of variance of the distance measure excluding these subjects revealed the same patterns of significant differences.

These results suggest that people do actively avoid situations involving charitable donations. The results further indicate that personal appeals will produce more avoidance than anonymous appeals. It is unlikely that the presence of the confederate produced more avoidance because of the normative pressures implied or because the appeal was somehow stronger; her presence produced more avoidance in the nondonation conditions as well as in the donation conditions. This suggests that much of the avoidance may stem from the invasion of personal space that can occur when a personal appeal is made.

[1] Subjects never used Lanes 5 or 6 in walking to the arts building and only infrequently used Lane 4. This was most likely because the lanes with lower numbers represented the shortest path to the arts building.

Table 1
Distance Maintained from Table: Study 1

Donation factor	Person factor: Person present	Person factor: No person present
Donation sought	1.71	1.33
No donation sought	1.42	1.20

Note. Distances are in terms of 76-cm units.

Study 2

The second experiment was conducted to replicate the findings of Study 1 and further to examine differences between personal and impersonal appeals in evoking avoidance behavior.

Study 1 did not control for the effects of sex of subject or the nature of the charity involved. It is possible that differences in avoidance behavior were the result of having different proportions of males and females in each of the cells. It may also be possible that subjects' feelings about the particular charity involved may have produced a different pattern of avoidance than if another charity had been used. For these reasons, sex of subject was included as a factor in Study 2, and the appeal was made on behalf of the United Nations Children's Fund (UNICEF), a different charity from that used in Study 1.

It is often the case in appeals for donations that members of the disadvantaged group act as canvassers. While potential donors may be more sympathetic to handicapped canvassers, the research literature suggests that disabled individuals can also elicit greater avoidance. For example, Acton, Bentley, Matheson, & Stewart (Note 2) found that subjects avoided looking at a disabled person more than at a nondisabled person during the course of an interview. In a study by Kleck, Ono, and Hastorf (1966), subjects interacting with a handicapped individual terminated the interactions sooner than did subjects interacting with a nonhandicapped individual. In Study 2, subjects again walked by a table that did or did not display a donation box and at which a handicapped confederate, a nonhandicapped confederate, or no one was seated. It was hypothesized that greater avoidance would occur when the handicapped individual sat at the table, less avoidance would occur when the nonhandicapped person was at the table, and least avoidance would result when no one was seated at the table.

Method

Summary of design. The experimental design was a $2 \times 2 \times 3$ factorial. Male and female subjects (sex factor) were monitored as they walked past a table that was or was not set up as a center for donations (donation factor). Seated at the table was a handicapped person, a nonhandicapped person, or no one (person factor). Again, the dependent measure was the distance maintained by subjects as they walked past the table.

Subjects. The subjects were 462 people (51 males and 26 females in each of six experimental conditions) who were randomly selected as they walked from the library to the arts building as described in Study 1.

Experimental site. The site used was the same as in Study 1. The only change made was in the width of the lanes marked along the tunnel floor. The lanes had been reduced from 76 cm to 62 cm in width. This provided for finer distance discriminations with no apparent loss in interrater reliability.

Procedure. Subjects were observed during three sessions, in the morning and afternoon of one day and in the morning of the following day. Approximately the same number of subjects were observed in each of the three sessions. Two observers monitored subjects in the same manner as in Study 1. The experimental conditions included all six combinations of the donation factor and the person factor. These six conditions were arranged in a different random order for each of the three experimental sessions.

The three donation conditions were somewhat different from those in Study 1, since a different charity was being used. In these conditions a sign reading "Please donate to UNICEF" was hung from the end of the table facing oncoming traffic. A donation box ($18 \times 18 \times 23$ cm) was positioned on the table as in the first experiment. In the no-donation conditions, the table was bare.

In the handicapped-person conditions, a male confederate sat in a wheelchair near the end of the table that faced oncoming subjects. The wheelchair was backed against the wall and slanted toward the traffic at a 45° angle. Both the wheelchair and the donation sign were clearly visible to approaching subjects. In the nonhandicapped-person conditions, the confederate sat in an ordinary chair, positioned exactly as the wheelchair had been. In the no-person conditions, this chair was empty.

Table 2
Distance Maintained from Table: Study 2

Donation factor	Person factor		
	Handicapped	Nonhandicapped	No person
Donation sought	2.08_a	1.99_a	1.30_{bd}
No donation sought	1.66_b	1.44_{bc}	1.13_d

Note. Distances are in terms of 62-cm units. Means carrying different subscripts differ at the $p < .05$ level of significance.

Results

Analysis of variance of the distance measure revealed no main effects for sex of subjects and no interactions involving this variable. Consequently, the data were collapsed across sex and reanalyzed using a 2 × 3 (Donation × Person) analysis of variance. Mean distances are presented in Table 2. Cochran's test for homogeneity of variance again revealed significant differences among cell variances ($C = .232$, $p < .05$). Distance scores were therefore transformed as in Study 1 before being submitted to analysis of variance. Transforming the scores appeared to result in homogeneous cell variances ($C = .212$, *ns*).

Analysis of variance of the transformed data revealed a main effect for the donation factor, $F(1, 456) = 45.03$, $p < .0001$. As in the first experiment, subjects maintained a greater distance from the table when it was set up as a donation station than when it was not. A main effect was also obtained for the person factor, $F(2, 456) = 60.96$, $p < .0001$. Planned comparisons showed that the greatest distances were maintained from the table at which the handicapped person was seated, intermediate distances were maintained from the table at which the nonhandicapped person was seated, and least distances were maintained when no one was at the table, $t(456) = 2.17$, $p < .05$ for handicapped versus nonhandicapped means; $t(456) = 8.29$, $p < .001$ for nonhandicapped versus no-person means; $t(456) = 10.46$, $p < .001$ for handicapped versus no-person means.

Analysis of variance also yielded a marginal Donation × Person interaction, $F(2, 456) = 2.84$, $p < .06$. Multiple comparisons of the cell means (indicated in Table 2) show that while the donation and no-donation means did not differ when there was no person at the table, they did differ significantly when either a handicapped or a nonhandicapped confederate sat at the table.

Only a relatively small number of people donated money during the course of the experiment. Twelve people made donations, in all; 5 when the handicapped confederate was present, 6 when the nonhandicapped confederate was present, and only 1 when no one was at the table. Naturally, these people were recorded as passing the table in Lane 1. Prior analysis that excluded these subjects revealed the same patterns of significant differences.

General Discussion

These results provide support for the hypothesis that people will often avoid situations in which they are asked to give help. The main effects found in both studies for the donation factor suggest that the appeal itself, regardless of its impersonal or personal nature, elicits avoidance behavior. The main effects found for the person factor in both studies indicate that another element involved in producing avoidance on the part of potential donors may be the invasion of personal space that can occur when one is approached by an unfamiliar person seeking a donation. However, the Person × Donation interaction found in Study 2 suggests that there may be another important component operating in personal appeals for help. The fact that personal appeals (by a handicapped or nonhandicapped individual) yielded greater avoidance relative to the corresponding nondonation conditions than did nonpersonal appeals suggests that personal appeals represent more than an invasion of personal space. They may increase the strength of the appeal and thereby influence the donor's perceived choice as to whether or not to give. This would result in greater reactance and hence greater avoidance. Further support for this interpretation comes from the fact that the personal appeals, while

eliciting more avoidance, also produced more donations than the anonymous appeal.

Presumably, avoidance will be a function of the relative strength of the forces that motivate and inhibit helping. If those factors that motivate helping are much more powerful than those that inhibit helping, the individual will give help and suffer no conflict in doing so. If, on the other hand, the inhibiting factors are much stronger than the motivating factors, the individual will *not* help but, again, will suffer no conflict in doing so. Conflict (and hence avoidance) will be maximal when the forces that motivate helping and those that inhibit helping are equally strong.

These studies address an aspect of helping not often considered in helping research. Helping studies have typically been quite spare in the kinds of dependent variables they examine. Often the amount of help given is the only dependent variable considered. The results of the current research indicate that in looking only at amount of helping, investigators may be ignoring many aspects of the helping *situation*. A knowledge of the amount and kind of avoidance produced, for example, could aid greatly in understanding why people are often so reluctant to give help. Such considerations might also lead to more effective appeals, in which helping entails less conflict and aversion for the potential helper.

Reference Notes

1. Wilsdon, L. Canadian Save the Children Fund. Personal communication, February 1978.
2. Acton, R. G., Bentley, K. L., Matheson, D. W., & Stewart, N. J. *Effects of interviewer's status and presence of physical disability on interpersonal distance and sympathy toward interviewer.* Unpublished manuscript, University of Saskatchewan, 1976.

References

Baron, R. A. Invasions of personal space and helping: Mediating effects of invader's apparent need. *Journal of Experimental Social Psychology*, 1978, *14*, 304–312.

Baron, R. A., & Bell, P. A. Physical distance and helping: Some unexpected benefits of "crowding in" on others. *Journal of Applied Social Psychology*, 1976, *6*, 95–104.

Bryan, J. H., & Test, M. A. Models and helping: Naturalistic studies in aiding behavior. *Journal of Personality and Social Psychology*, 1967, *6*, 400–407.

Darley, J. M., & Latané, B. Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, 1968, *8*, 377–383.

Harris, M. B., & Meyer, F. W. Dependency, threat, and helping. *Journal of Social Psychology*, 1973, *90*, 239–242.

Jones, R. A. Volunteering to help: The effects of choice, dependence, and anticipated dependence. *Journal of Personality and Social Psychology*, 1970, *14*, 121–129.

Kleck, R., Ono, H., & Hastorf, A. H. The effects of physical deviance on face to face interaction. *Human Relations*, 1966, *19*, 425–436.

Konečni, V. J. Some effects of guilt on compliance: A field replication. *Journal of Personality and Social Psychology*, 1972, *23*, 30–32.

Konečni, V. J., Libuser, L., Morton, H., & Ebbesen, E. B. Effects of a violation of personal space on escape and helping responses. *Journal of Experimental Social Psychology*, 1975, *11*, 288–299.

Latané, B., & Darley, J. M. Help in a crisis: Bystander response to an emergency. In J. W. Thibaut, J. T. Spence, & R. C. Carson (Eds.), *Contemporary topics in social psychology*. Morristown, N.J.: General Learning Press, 1976.

Schwartz, S. H. Moral decision making and behavior. In J. Macaulay & L. Berkowitz (Eds.), *Altruism and helping behavior: Social psychological studies of some antecedents and consequences*. New York: Academic Press, 1970.

Received October 6, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 8, 1412–1417

The Importance of Consistency of Modeling Behavior Upon Imitation: A Comparison of Single and Multiple Models

Peter A. Fehrenbach, David J. Miller, and Mark H. Thelen
University of Missouri—Columbia

The present study was designed to assess the differential effects of single versus multiple adult models on children's expression of preferences when the modeled preferences are either consistent or divergent. Second- and third-grade children were exposed to either a single model who made two choices per trial or to two models (multiple models) who made one choice each per trial. The modeled choices were either identical (consistent) or different (inconsistent) from each other on every trial of two four-option preference tasks. A single-response control condition was included, in which a single model made one choice per trial. As expected, children who observed multiple models whose choices were consistent (multiple–consistent) imitated more than any other group. Children who were exposed to two models whose choices differed from each other on every trial (multiple–inconsistent) showed little imitation of either model. An intermediate amount of imitation was observed for the single model conditions, with the single model who made two identical choices per trial (single–consistent) resulting in more imitation than the single model whose choices were inconsistent on each trial (single–inconsistent). The single–consistent group and the single-response control group did not differ in imitation, but the two inconsistent modeling groups showed significantly less imitation than the control group.

Research on the effects of modeling typically has been conducted within a methodological framework in which a subject observes the behavior of a single model. Yet in naturalistic settings, individuals often observe the responses of more than one person to a given situation. This phenomenon has been called multiple modeling and may involve similar or divergent responses on the part of the models (Bandura & Walters, 1963). The present study was designed to assess the differential effects of single versus multiple adult models on children's imitation of preferences when the modeled preferences are either consistent or divergent.

Previous research suggests that multiple models are more likely than a single model to facilitate maintenance of reductions in the fear of dogs in children (Bandura & Menlove, 1968). However, in that study the multiple modeling subjects viewed the models interacting with different dogs, whereas the single model interacted with only one dog. Thus, the number of observed models and the number of feared stimuli were confounded. With such a confound removed, greater reductions in snake avoidance (Kazdin, 1974) and larger increases in self-reported assertive social skills (Kazdin, 1975, 1976) have been reported when multiple covert modeling procedures rather than single covert modeling procedures were used.

Other research has examined the effects of live multiple modeling cues on children's standards of self-reward, sharing, and resistance to temptation. In support of the notion that consistently deviant modeling cues operate additively, Allen and Liebert (1969) found that exposure to two deviant adult models weakened children's adoption of previously learned standards of self-reward more than exposure to a single deviant model. Also, Liebert and Fernandez (1970) found that

Requests for reprints should be directed to Mark H.Thelen, Department of Psychology, University of Missouri—Columbia, Columbia, Missouri 65211.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3708-1412$00.75

nursery school boys were more likely to imitate specific norms for sharing when multiple models demonstrated the norm as compared with a single model.

In making a direct comparison of a single model with two models, Allen and Liebert (1969) and Liebert and Fernandez (1970) confounded the number of models with the number of exposures to the modeled behavior. In both studies the single model condition consisted of essentially one exposure to the deviant or generous model's behavior, whereas in the multiple modeling condition subjects observed twice the number of deviant or generous responses. The greater behavioral change observed following multiple modeling may have been simply a function of increased exposure to the deviant or generous behavior, regardless of the number of models.

Other support for the hypothesis that modeling cues operate additively comes from McMains and Liebert (1968), who have shown that children exposed successively to two models, one of whom behaved in accordance with rules (strict model) and the other in a deviant fashion (lenient model), showed more rule violation than children exposed to two strict models and less rule violation than children who observed two lenient models. Wolf (Note 1) obtained similar effects of multiple models upon children's resistance to temptation.

Two methodological considerations limit the generalizability of the findings of McMains and Liebert (1968) and Wolf (Note 1) with respect to the effects of discrepant multiple models upon imitation. First, in many situations in which multiple models demonstrate discrepant behaviors (e.g., in the expression of preferences or judgments), the observer has the option of not imitating either model in addition to the option of imitating one or more of the models. In the McMains and Liebert (1968) and Wolf (Note 1) studies, nonimitation was impossible to the extent that the subject's violation of the rule could be taken as indicating imitation of the lenient model and adherence to the rule as imitation of the strict model. Second, in both studies, the child's behavior was measured after the modeling manipulations and when the child was alone. Under conditions where a child must respond in the presence of models, it is more likely that the observer will anticipate the social consequences of imitating or not imitating. Thus, when two models are divergent in their behaviors, the observers may perceive imitation of one model as offensive to the other, nonimitated model. Given the option of nonimitation, such perceptions could result in lower levels of imitation of either model than would be predicted by the hypothesis that modeling cues operate additively.

Different kinds of social factors seem to be operating when an individual responds in the presence of a single model. In this case the specific characteristics and behavior of the single model are more important in determining imitation. For example, an adult model who demonstrates consistency in preferences is likely to be viewed as more certain and deserving of emulation than an adult who is inconsistent or vacillating in his or her preferences.

The present investigation differed methodologically from previous studies on multiple modeling in several ways. The experimental tasks involved the expression of preferences, rather than self-initiated standards of behavior, and were performed in the presence of the model(s). The number of exposures to modeled behavior was controlled by having the single model make two responses, the same number of responses that occurred in the multiple modeling conditions where two models responded once each. Finally, a 2×2 factorial design was employed with the number of models (single vs. multiple) and type of modeling (consistent vs. inconsistent responses) as the main factors. A single model – single response control group was also included.

It was expected that the effects of multiple modeling on imitation of preferences would depend upon whether the models were consistent or divergent with each other in their choices. When both models expressed the same preference, high levels of imitation were expected. When the models expressed preferences different from each other, very little imitation of either model was expected. An intermediate amount of imitation was pre-

dicted for single model subjects, the inconsistent – single modeling condition resulting in less imitation than a consistent model.

Method

Subjects

The subjects were 80 second- and third-grade children (40 female and 40 male) from a primarily middle-class public elementary school. The experimenter was a senior undergraduate male, and the models were two female undergraduates. The subject pool was counterbalanced across the four experimental conditions and one control condition with respect to sex, grade, seating position of the model(s), the order in which the models responded in the multiple modeling conditions, and the model confederate utilized in the single modeling conditions. Each of the five conditions contained 16 subjects.

Tasks and Materials

Two limited-option preference tasks were used in this study. The colors task involved 10 5 × 8 inch index cards with four 1-inch color squares attached horizontally across the cards. The color squares varied only slightly in shades of the same basic color on any given card.

The faces task consisted of 10 emotionally expressive faces, each mounted on a 5 × 8 inch index card. Below each face were typed four nonsense syllables, which were three-letter consonant-vowel-consonant trigrams with an associative value of less than 15% (Archer, 1960).

Procedure

The experimenter escorted the subject individually to the experimental room, and the model arrived shortly after the subject. In the multiple modeling conditions, the second model arrived shortly after the first model. The models were introduced to each other and to the experimenter in order to decrease suspicion on the part of the subject that the models knew each other or the experimenter. The subject and the models sat across a table from the experimenter, the subject being seated between the two models in the multiple modeling conditions. The experimenter then explained that he was seeking the assistance of students and teacher aides in deciding which colors to paint some classrooms.

In both of the single model conditions, the model and subject were told that each would have a chance to pick a color. In order to determine who went first, the experimenter drew their names from a box. The model's name was always selected first, and she was told that she could therefore go twice before the subject made a selection from the colors. Both model and subject were told that each choice they made could be the same or different from the previous choices.

The experimenter then administered 10 trials on the colors task. On each trial in the single–consistent condition, the model always pointed on her second chance to the *same* color square that she had selected on her first turn. In the single–inconsistent condition, the model always selected a *different* color square on her second turn. In all conditions the model's choices were randomly determined.

After 10 trials on the colors task, the experimenter told the subject and model that he was writing a children's book and needed their help in picking "crazy names" to go with faces. They were asked to pick the name they thought went with the face. Ten trials on the faces task were then administered. As on the colors task, the model in the single–consistent condition pointed to the same name twice. In the single–inconsistent condition the model picked two different names on each trial.

In both multiple modeling conditions, the models and subject were told that each would have a chance to pick colors. Again, the experimenter drew names from the box to determine the order to be followed in making their choices. The models' names were always drawn before the subject in a prearranged, counterbalanced order. The rest of the instructions for both tasks were essentially the same as in the single model conditions, except that both the models and subject were told that they could make one choice of colors (or names) per trial. On each trial in the multiple–consistent condition, the second model always expressed the same preference as the first model. In the multiple–inconsistent condition, the second model always selected a different color or name from the first model.

The order of responding in the single-response control condition was established in the same way as in the two single modeling groups. However, the model was instructed to go first and to make only one choice per trial.

Results

The number of trials on which the subject matched a model's choice on the color task and the faces task and the sum of the matched responses on colors plus faces were used as the main measures of imitation. In general, the results of the analyses of the separate tasks are consistent with those obtained when matched responses on the two tasks were summated. Therefore, unless otherwise indicated, only the results of colors plus faces will be reported. Preliminary analyses indicated no significant effects on any of the dependent measures with respect to model confederates, seating position of the model(s), or the order

in which the multiple models made their choices. Therefore, the data were pooled across these variables in subsequent analyses.

Table 1 presents the means and standard deviations of the number of matched responses on colors plus faces tasks for the five groups. Inspection of this table reveals that the predicted interaction of number of models and consistency of modeled responses was obtained. To test the reliability of this finding, a $2 \times 2 \times 2 \times 2$ analysis of variance was performed, the factors being number of models (single vs. multiple), type of modeling (consistent vs. inconsistent), grade of subject (second vs. third), and sex of subject. A significant main effect for number of models and type of modeling was obtained, but the predicted interaction of number of models and type of modeling was also highly significant, $F(1, 48) = 20.71$, $p < .001$.

Two-tailed t tests indicated that as predicted, more imitation occurred in the multiple–consistent group than in the single–consistent condition, $t(48) = 5.92$, $p < .001$. Also as predicted, the single–consistent condition resulted in more imitation than the single–inconsistent condition, $t(48) = 2.01$, $p < .05$. Finally, although the means are in the predicted direction, the mean for the single–inconsistent group was not significantly greater than the mean for multiple–inconsistent group.

In summary, multiple models who were consistent with each other in the expression of preferences more strongly influenced the children to imitate than a single model who made

Table 1
Means and Standard Deviations of the Total Number of Matched Responses on the Colors Plus Faces Tasks

Condition	No. of matched responses	
	M	*SD*
ultiple–consistent	10.56	4.93
ultiple–inconsistent	2.06	2.32
ngle–consistent	4.63	2.42
ngle–inconsistent	2.56	2.00
ngle-response control	4.00	2.03

two consistent choices. However, multiple models who showed divergent preferences were imitated very little. These predicted results were obtained despite the fact that the probability of obtaining matching responses by chance alone was higher in the inconsistent conditions, where two of the four options were modeled ($p = .50$), than in the consistent conditions, where chance imitation was only one out of four ($p = .25$). Imitation was significantly greater than chance in the multiple–consistent condition, $t(15) = 3.92$, $p < .01$, but significantly less imitation than that expected by chance alone was obtained in the multiple–inconsistent and single–inconsistent conditions, $t(15) = -13.67$, $p < .001$, and $t(15) = -14.89$, $p < .001$, respectively.

An unexpected three-way interaction of number of models, type of modeling, and sex of subject was obtained, $F(1, 48) = 4.02$, $p < .05$. Separate comparisons of males and females in all four experimental groups revealed that the only significant sex differences occurred in the multiple–consistent condition, where the female subjects imitated more than males, $t(48) = 4.50$, $p < .001$.

Finally, nonorthogonal planned comparisons of the single-response control condition with the four experimental groups were performed. The multiple–consistent condition evidenced more imitation than the control group, $t(75) = 6.28$, $p < .001$ (all t tests are two-tailed). The single–consistent group did not differ in imitation from the single-response control group. However, the multiple–inconsistent group ($M = 1.06$) imitated less than the control group ($M = 2.31$) on faces, $t(75) = -2.27$, $p < .05$; there was only a trend in the same direction on colors plus faces, $t(75) = -1.86$, $p < .10$. Similarly, subjects in the single–inconsistent condition ($M = 1.06$) imitated less than control subjects ($M = 2.31$) on faces, $t(75) = -2.02$, $p < .05$; a trend in the same direction was obtained for colors plus faces, $t(75) = -1.38$, $p < .20$.

In summary, under the present conditions it appears that exposure to a single model making two consistent responses does not increase imitation over that obtained after observing a single model responding once. The likelihood of imitation does increase, however,

when multiple models respond consistently with each other. While it thus appears that multiple models are required to increase imitation, imitation may be decreased by inconsistent modeling, regardless of the number of models. This is evident by the finding of equivalent reductions in imitation in both inconsistent groups relative to the single-response control group and the single–consistent group.

Discussion

The results provide support for the notion that observing multiple models can increase imitation of preferences in children significantly over that obtained after observing a single model. These results were obtained when the number of exposures to modeled behavior was equated in both the single and multiple model conditions. Previous studies comparing single and multiple modeling have failed to control for the number of exposures to modeled behavior. However, the results also strongly suggest that a greater degree of imitation with multiple models can only be expected when the multiple models present consistent behaviors to imitate. When multiple models appear to be in conflict or are otherwise divergent in their behavior, very little imitation of either model can be expected.

The finding of a dramatic attenuation of imitation in the multiple–inconsistent condition warrants some explanation. While much of the research on imitation has emphasized the unidirectional impact of modeling on the subsequent behavior of a naive subject, more recent studies have considered the reciprocal-influence process involved in the ongoing model–imitator interaction. It has been found, for example, that being imitated is reinforcing to the model and leads to reciprocal imitation and to increased attraction on the part of the model toward the imitator (Thelen, Dollinger, & Roberts, 1975). Given that the effects of being imitated are often positive, it is likely that children who observed two high-status adults (supposedly teacher's aides) expressing different preferences were conflicted regarding which one to imitate. Children in the second and third grades may realize the possible negative social consequences (e.g., lowered attractiveness) of choosing to imitate or conform with one adult rather than another. Given the option not to imitate, as was the case in this study, the children may have chosen to avoid the conflict by imitating neither model.

A different explanation is suggested when one notes that, although they were in the predicted direction, the differences between the single–inconsistent and multiple–inconsistent conditions were not significant. These results, considered in conjunction with the finding that both inconsistent conditions resulted in less imitation than the two consistent conditions, indicate that the crucial variable accounting for the low levels of imitation in the multiple–inconsistent condition may be the inconsistency of modeling cues per se, rather than social interaction variables associated with the number of models.

A third explanation for the low levels of imitation in the multiple–inconsistent condition involves the possibility that the subjects were actually imitating the nonimitative (divergent) behavior of the second model. The second model was thus demonstrating oppositional behavior toward the first model and giving tacit acknowledgment that the child could similarly make a choice different from that of the first model. Brehm (1977) found that when one adult's attempt to influence children's choice preferences was followed by a second adult's suggestion that they should choose whatever they wanted, first graders (but not fifth graders) acted in opposition to the first adult. Perhaps simply observing conflict in preferences among high-status adults disposes a child not to imitate either adult's choices. The potential effects upon socialization resulting from a child's receiving conflicting modeling cues from such significant adults as parents remains to be determined in research conducted in more naturalistic settings than those employed in the present study.

With respect to the single model conditions, it appears that doubling the number of exposures to modeled choices in the single–consistent conditions did not result in more imitation than the single model - single-response control condition. However, when the single

nodel made two divergent choices, the chil-lren were less likely to imitate either choice han when the model made the same choice wice. The difference in imitation between hese two single model groups may have been n part a function of the lowered esteem in vhich the inconsistent model was held relative o the consistent model. Or, as mentioned arlier, the inconsistency of modeled informa-ion (cues) may result in reduced imitation ndependent of social factors.

This study presents strong support for the ypothesis that multiple models are more ffective in producing imitative behavior than single model. This finding is only supported, owever, when the models behave consistently vith each other. Oppositional behavior or low evels of imitation or compliance are likely to esult when two models conflict in their be-avior or when a single model behaves incon-istently. Further research conducted in more aturalistic settings is necessary to determine vhether inconsistency in modeling cues pre-ented by either single or multiple models ignificantly affects socialization. For example, f parents in marital conflict provide divergent nodeling cues, what effect might this have on heir children's behavior?

Reference Note

Wolf, T. M. *The effects of multiple modeling on resistance to deviation.* Paper presented at the biennial meeting of the Society for Research in Child Development, Santa Monica, California, April 1969.

References

Allen, M. K., & Liebert, R. M. Effects of live and symbolic deviant-modeling cues on adoption of a previously learned standard. *Journal of Personality and Social Psychology,* 1969, *11,* 253–260.

Archer, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychological Monographs,* 1960, *74*(10, Whole No. 497).

Bandura, A., & Menlove, F. L. Factors determining vicarious extinction of avoidance behavior through symbolic modeling. *Journal of Personality and Social Psychology,* 1968, *8,* 99–108.

Bandura, A., & Walters, R. H. *Social learning and personality development.* New York: Holt, Rinehart & Winston, 1963.

Brehm, S. S. The effect of adult influence on children's preferences: Compliance versus opposition. *Journal of Abnormal Child Development,* 1977, *5,* 31–41.

Kazdin, A. E. Comparative effects of some variations of covert modeling. *Journal of Behavior Therapy and Experimental Psychiatry,* 1974, *5,* 225–231.

Kazdin, A. E. Covert modeling, imagery assessment, and assertive behavior. *Journal of Consulting and Clinical Psychology,* 1975, *43,* 716–724.

Kazdin, A. E. Effects of covert modeling, multiple models, and model reinforcement on assertive behavior. *Behavior Therapy,* 1976, *7,* 211–222.

Liebert, R. M., & Fernandez, L. E. Effects of single and multiple modeling cues on establishing norms for sharing. *Proceedings of the 78th Annual Convention of the American Psychological Association,* 1970, *5,* 437–438. (Summary)

McMains, M. J., & Liebert, R. M. Influence of discrepancies between successively modeled self-reward criteria on the adoption of a self-imposed standard. *Journal of Personality and Social Psychology,* 1968, *8,* 166–171.

Thelen, M. H., Dollinger, S. J., & Roberts, M. C. On being imitated: Its effects on attraction and reciprocal imitation. *Journal of Personality and Social Psychology,* 1975, *31,* 467–472.

Received October 11, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 8, 1418–1420

Self-Serving Attributional Biases: Perceptual or Response Distortions?

Gifford Weary
Ohio State University

The present article examines Miller's analysis of what constitutes a self-serving attributional bias. It is argued that his delineation of different types of self-serving attributions is not supported by the empirical evidence collected to date and that what previous authors (e.g., Miller & Ross) have viewed as a perceptual bias in the causal inference process may be better seen as a response bias or as a strategic self-presentation designed to maximize public esteem.

In Miller's (1978) recent reply to my re-examination of the notion of self-serving biases in the causal inference process (Bradley, 1978),* he discusses the merit of my broadened self-serving bias formulation and addresses the important issue of what constitutes a self-serving attribution. While an examination of this latter issue is long overdue, I will argue in the present article that the analysis presented by Miller serves only to obfuscate issues in an already confusing area of inquiry. Moreover, I will contend that in his reply, Miller (a) incompletely summarized my formulation of self-serving attributional biases and (b) failed to address important criticisms raised in my examination of Miller and Ross's (1975) information-processing analysis of data that have been interpreted as reflecting motivational biases.

In my review (Bradley, 1978) of the empirical evidence relevant to the proposition that self-serving biases modify attributions of causality, I presented a broadened formulation to account for seemingly counterdefensive attributions (i.e., attributions indicating greater acceptance of responsibility by the attributor for negative than for positive outcomes attendant upon his/her behavior). Specifically, I suggested that self-serving attributions may be viewed as public self-presentations designed to maximize public esteem and that *under certain conditions* such esteem needs may be best served by accepting responsibility for negative outcomes.

While Miller (1978) feels this broadened formulation "has the virtue of parsimony" (p. 1221), he also states that it "seems to vitiate the virtue of parsimony by being over-inclusive and allowing *any* asymmetrical attributions for success and failure to be interpreted as being self-serving in nature" (p. 1221). It is important to note, however, that certain major *conditions* under which one would expect esteem motives to lead to defensive versus counterdefensive attributions were specified in my earlier article. It was pointed out that in all of the studies that have offered support for counterdefensive attributional processes, there has existed the opportunity to compare simultaneously the subject's self-attribution for his or her own task performance with his or her subsequent behavior on a similar task or with another's causal attribution for the subject's outcome. I suggested (1978, pp. 64–66) that the probable *simultaneous comparison* of and *potential inconsistency* between one's description of causality and (a) one's subsequent behavior and/or (b) another's description of causality for one's performance outcomes may prevent an individual from taking too much credit fo

The author thanks John H. Harvey for his helpful comments on an earlier draft of this paper.

* Some of the author's previous publications have appeared under the name "Gifford Weary Bradley."

Requests for reprints should be sent to Gifford Weary, Department of Psychology, Ohio State University, 164 West 19th Avenue, Columbus, Ohio 43210.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3708-1418$00.75

success or denying blame for failure. In view of this reasoning, I contend that the broadened formulation is not "overinclusive" and is not invulnerable to disconfirmation.

In an attempt to define a self-serving attribution, Miller (1978) suggests that attributions for positive and negative outcomes can represent either distortions in perception of causality or distortions in descriptions of causality. That is, "they can serve to protect or enhance how the person perceives himself or herself" or "they can serve to protect or enhance how *others* perceive the person" (p. 1221). Moreover, Miller argues that my analysis of self-serving attributional phenomena has obscured two distinct psychological processes, because I use the term *self-serving* to refer to attributions that serve one's public image, whereas previous authors have used this term to refer to attributions that serve one's private image.

There are several problems associated with this statement. First, as Miller (1978) himself admits, there is little evidence to support the proposition that "we are prone to alter our *perception* of causality so as to protect or enhance our self-esteem" (p. 1222). In most of the studies reviewed by Miller and Ross (1975) and myself (Bradley, 1978), subjects' performances and causal interpretations were public in nature; consequently, subjects' causal ascriptions for their positive and negative outcomes likely reflected "distortion in descriptions or self-presentations, not perceptions" (Miller, 1978, p. 1222).[1] Second, the preceding statement should not be taken to imply that private self-esteem needs are not influenced by distortions in descriptions of causality. For example, if the self-serving bias effect is mediated by a desire to maintain or gain a positive public image, then positive evaluations from others may serve to enhance the esteem one feels for oneself; conversely, negative evaluations from others may threaten the individual's positive, private self-image. Based upon the empirical evidence collected to date, then, what previous authors (e.g., Miller & Ross, 1975) have viewed as a perceptual bias in the causal inference process may be better viewed as a response bias or strategic self-presentation designed to maximize public esteem (Bradley, 1978) and, secondarily, private self-esteem.[2]

In their 1975 review, Miller and Ross proposed that the results of many of the studies often cited as support for self-serving attributional biases could readily be "interpreted in information-processing terms" (p. 224). Specifically, they contended that the observed tendency for individuals to accept greater responsibility for positive than for negative outcomes may occur for any or all of several reasons: (a) Individuals intend and expect success more than failure and are more likely to make self-ascriptions for expected than for unexpected outcomes, (b) perceived covariation between response and outcome may be more apparent for individuals experiencing a pattern of increasing success than for individuals experiencing constant failure, and (c) people erroneously base their judgments of the contingency between response and outcome in terms of the occurrence of the desired outcome (i.e., success) rather than any actual degree of contingency.

In my examination of the explanatory strength of these alternative interpretations of findings offered as evidence for self-serving attributional biases, I concluded that they

---

[1] The major issue underlying Miller's differentiation of two types of self-serving attributions involves the *locus* of distortion in the causal inference process (i.e., a perceptual vs. a response bias). However, I would argue that (a) there are probably multiple points of bias in the attribution process, from input to output, and (b) present evidence does not allow us to determine the precise point at which bias or defensiveness may be occurring. At this point in time, a more pertinent empirical question would seem to be how and under which conditions self-serving motivations influence attributions of causality.

[2] Miller suggests that securing behavioral measures might help future researchers determine whether individuals distort causal ascriptions for their positive and negative outcomes in order to serve their public or their private images. Since private self-esteem needs might be influenced by public self-presentations, it is not clear that behavioral measures would help determine whether the self-serving bias effect represents distortions in descriptions or perceptions. Moreover, it seems reasonable to argue that written or verbal measures of causal responsibility are in fact behavioral measures and that individuals may infer their private self-attributions from their public, behavioral statements of causality (cf. Bem, 1972).

suffer from conceptual incompleteness and a lack of parsimony. It is indeed unfortunate that Miller chose not to use his recent reply to my 1978 review as a forum for addressing these criticisms; while many of Miller and Ross's (1975) points are meritorious and deserve *direct* investigation, I suspect such experimentation will not be forthcoming until the alternative interpretations proposed by these authors receive considerable theoretical clarification.

In conclusion, it is contended that the validity and explanatory power of Miller and Ross's (1975) information-processing analysis of data presumably reflecting motivational biases in the causal inference process remains undetermined. In addition, it is concluded that the empirical evidence related to the notion of self-serving biases in causal attributions is most consistent with a self-presentation formulation holding that individuals alter their descriptions of causality so as to enhance or protect others' perceptions of them.

References

Bem, D. J. Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6). New York: Academic Press, 1972.

Bradley, G. W. Self-serving biases in the attribution process: A reexamination of the fact or fiction question. *Journal of Personality and Social Psychology*, 1978, *36*, 56–71.

Miller, D. T. What constitutes a self-serving attributional bias?: A reply to Bradley. *Journal of Personality and Social Psychology*, 1978, *36*, 1221–1223.

Miller, D. T., & Ross, M. Self-serving biases in the attribution of causality. Fact or fiction? *Psychological Bulletin*, 1975, *82*, 213–225.

Received October 16, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 8, 1421-1431

The Bias Phenomenon in Attitude Attribution: Actor and Observer Perspectives

Arthur G. Miller, Robert Baer, and Peter Schonberg
Miami University (Ohio)

Extensive research in attitude attribution has shown that observers infer correspondent attitudes from behavior that is externally constrained. The present study concerned (a) the degree to which actors anticipate observers' manifestation of this bias, (b) the effect of presenting constraint information directly from the actor as compared to the experimenter's instructions, and (c) the relationship between constraint experienced by the actor and the persuasiveness of essays produced under different constraint levels. Essay writers clearly predicted that observers would infer correspondent attitudes even when the position had been randomly assigned to the writer. This was true, although to a diminished extent, when the writer was under genuinely high constraint. When actors expected observers to have precise information regarding their actual constraint, however, they anticipated that observers would recognize the attributional implications of such information. Data from observers corroborated the actors' predictions. When observers were not given this information, their attributions were based solely on essay content and indicated no recognition of the different freedom levels experienced by the essay writers. Essays written under different constraint levels were judged by observers to be generally of similar and fairly respectable persuasiveness. It is suggested that the bias phenomenon may be a consequence of presenting observers with essays more persuasive than they expect from a writer who in fact disagrees with the assignment.

An individual's prediction of the attribution that others will make to him or her is an issue of practical as well as conceptual interest. Accurate predictions should enhance effectiveness in managing the impression one makes upon others. Conflicts could be eased by anticipating how others will view the causes of one's actions. Research attention has focused recently on attributional bias, in particular the tendency to overestimate the importance of personal or dispositional determinants relative to situational influence—what Ross (1977) has termed the "fundamental attribution error," which was a central concept in the article by Jones and Nisbett (1971). The present study was primarily concerned with three issues. The first was the degree to which actors anticipate or predict that observers of their behavior will manifest this particular bias. The second related to the source of constraint information—the effect of presenting constraint information *directly* from the actor and determining the impact of this kind of stimulus upon observers' attributions as well as upon actors' predictions of observer judgments. A final concern was the relationship between the amount of constraint experienced by the actor and the quality of the behavior produced under different constraint levels.

The procedure was a variation of a paradigm used by Jones and Harris (1967), in which subjects were told that a target person had written an essay under choice or no-choice instructions. Perhaps the most celebrated

Portions of this research were reported at the annual meeting of the American Psychological Association, Toronto, Canada, 1978.

Requests for reprints should be addressed to Arthur G. Miller, Department of Psychology, Miami University, Oxford, Ohio 45056.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3708-1421$00.75

finding of that study was the observation that subjects attributed attitudes in line with the essay content or direction (pro or anti) even when the position had been assigned to the writer. This finding has been reported in many experiments and is widely regarded as observer bias in the direction of personal causality (Jones, 1976, 1979; Jones & McGillis, 1976; Miller, 1974, 1976; Monson & Snyder, 1977; Ross, 1977; Snyder & Jones, 1974).

In the first part of the present experiment, actors were asked to predict the judgments of observers in the attitude attribution paradigm. Subjects were asked to write essays on affirmative action; they were either given a choice of position or were assigned to write in the *anti* or *pro* direction. After writing their essays, they were asked to predict the average attitude attribution that observers would make after reading the essays and after being informed of the instruction condition. Would actors predict attributions according to the model of correspondent inference theory? This would be shown by predicting correspondent inferences under the choice condition and predicting a normative or base-rate attitude, unrelated to essay directionality, in the assign conditions (Jones & McGillis, 1976). Actors should thus expect observers to feel confident that the essay represented the actor's true opinion in the choice condition but should expect observers to be in a very real dilemma in the no-choice or assign conditions and to make an attribution that would be the "best guess" under these circumstances.

Actors have an advantage over observers in being *explicitly* aware of the randomness of their essay assignment, their precise constraint in the actual writing of the essay, and their personal opinion on the essay topic. With these insights, actors could be expected to predict the discounting of their own behavior in the assign conditions. Conceptually, the necessary information (for example, randomness of assignment) is more salient to actors than to observers. Actors would therefore have at least the option of predicting the discounting strategy—a pattern dictated by attribution theory but one that observers do not appear inclined to endorse. Of particular interest were the predictions of actors who were asked to write essays contrary to their true opinion. Would these genuinely constrained actors expect observers to attach little informational value to their essay, or would they expect observers to base attributions upon the content of their essay? Perhaps the more likely outcome, given the available research data from observer subjects, is that actors, constrained or not, would predict correspondent inferences under the no-choice instruction. This latter outcome would suggest that the bias effect in attitude attribution not only results from the observer's lack of precise information regarding constraint but reflects a more constructive, intuitive strategy that subjects bring to their task of estimating attitudes from essays.

A second issue in the present study was the source of information concerning the actor's constraint. Jones and Harris (1967) and others have demonstrated bias by showing that perceivers appear to acknowledge neither the randomness of the writer's essay assignment in the no-choice conditions nor—a result of that randomness—the low informational or diagnostic value of the essay itself. There is a certain vagueness or abstractness to the no-choice instruction, however, in that it constitutes the *experimenter's* designation of the actor's situation but provides no explicit information concerning the *actor's* reaction to the essay assignment. In actual social interaction, the actor's personalized definition of the situation might be available to the observer. One might imagine a disgruntled writer muttering, "Oh, what luck, I would have to get *that* position—can I switch?" Would observers discount this type of information about constraint, coming directly from the actor? Would actors expect them to discount it? This issue was examined by comparing the impact of constraint information presented in the traditional, externally designated manner with information based upon the actor's own definition of the situation. It was expected that information concerning the actor's actual constraint would have a marked influence upon attributions of attitude, as well as predicted attributions, based on the essay. Specifically, indications of high constraint from the actor were expected to result in a discounting of the behavior.

A final concern was the persuasiveness of the essays produced in the assign conditions. Two relationships were examined. The first concerned the correlation between observers' attributions of attitude and their rating of the persuasiveness of the essay. The second concerned the degree to which essays composed under different constraint levels were judged as more or less persuasive. Data on these matters were considered as they might help explain the tendency for attributions to be associated with essay content in the no-choice conditions.

Method

Overview

In the first phase of the experiment, subjects filled out an attitude inventory, including an item that dealt with affirmative action. They were next asked to write a short essay and were either allowed to choose a position to defend or were assigned the pro or the anti position on the issue of affirmative action. In one condition, subjects were asked to rate their degree of freedom after having written their essay and to note this judgment at the bottom of the essay. In another condition, subjects were not immediately asked for this judgment. All subjects then estimated the average attribution that would be made by a group of observers who would read their essay and would judge the true attitude of the writer. Subjects were told that the readers would be aware of their condition (choice or assign). The subjects who rated their freedom after writing their essays were informed that observers would also have access to this rating. In a second phase of the research, subjects (observers) were given sets of essays written under varying constraint levels in the earlier part of the experiment. These subjects, who either were or were not shown the writer's freedom rating, were asked to estimate the writer's personal attitude on the issue of affirmative action and to indicate their opinion of the essay's persuasiveness. Subjects were then debriefed.

Subjects

In the essay-writing part of the experiment, the subjects were 276 undergraduates participating in research as a course requirement. Ninety subjects participated in the observer phase of the experiment. Subjects were seated at individual desks. Instructions were in written form, and subjects were seen in small groups (10 to 25). Subjects with procedural questions were aided individually by the experimenter.

Procedure

Actor predictions of observer attributions. The introduction to the experiment explained that it dealt with attitudes and the perception of beliefs. Subjects were informed that after filling out a short attitude inventory, they would be asked to write a short essay on one of the items and to fill out a questionnaire dealing with the essay that they had written. The key item on the 10-item attitude survey read as follows: "Women and other minorities should be given preference in job markets and other places where they are under-represented, assuming that they are qualified." Responses were made on a 10-point scale (1 = extremely disagree, 10 = extremely agree). In the second part of the experiment, described as a debate simulation, subjects were asked to consider the statement on affirmative action, as worded above, and to write a short, convincing essay in approximately 20 minutes. Subjects were randomly assigned to one of three categories listed at the top of the sheet upon which the essay was to be written. For each subject one category was checked in red: (a) choice of position, (b) assign pro, or (c) assign anti.

After writing their essays, subjects were randomly placed in one of two conditions. In one condition, to be termed *observer informed,* subjects were asked to rate the degree of choice or freedom they had while writing their essay. The definition of the low end of the scale (1 = no choice or freedom) was "as if you had copied the essay verbatim from another source." The high end (10 = complete choice and freedom) was defined as "indicating the completely free expression of your beliefs." This rating was located near the bottom of the $8\frac{1}{2} \times 14$ inch sheet upon which the essay had been written. In the other condition, to be termed *observer not informed,* subjects were not asked to make this freedom rating. All essays were collected. At this point, subjects were not aware of the specific questions that would follow.

Subjects then received questionnaires. The first item asked them to estimate the average attribution of a group of observers asked to read their essay and to guess their true opinion on affirmative action (1 = strongly against, 10 = strongly in favor). Subjects were informed that these observers would be shown the essay sheet and would be aware of the essay instructions. Subjects in the observer-informed condition, who had rated their freedom upon completion of their essay, were also told that observers would have access to this rating. After estimating the observers' attributions of attitude, subjects filled out several filler items on the two-page questionnaire. For subjects in the observer-not-informed condition, the freedom rating was included as the final item on the second page. The timing of the freedom rating and its mention (or omission) when predictions of observers' attributions were requested thus differentiated the observer-informed and observer-not-informed conditions. The design of this part of the experiment was a 3×2 factorial (Choice,

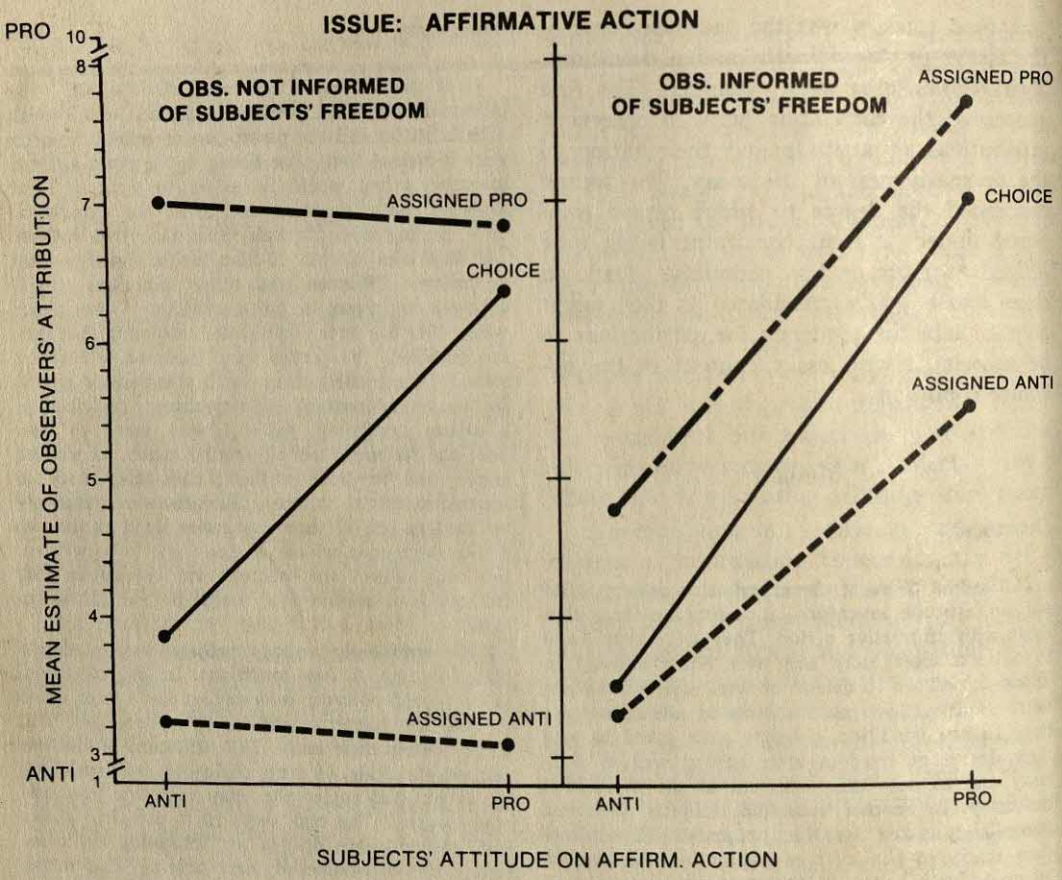


Figure 1. Mean estimate of observers' attributions as a function of writer's attitude, essay condition, and instruction regarding informing observers of writer's freedom. (Obs. = observer.)

Assign Pro, Assign Anti × Observer Informed/Not Informed).

Observer data. Ninety essays produced in the assign-pro ($n = 45$) and assign-anti ($n = 45$) conditions of *observer informed* were selected as stimuli for this phase of the experiment. The fact that the writer's freedom rating was located below the essay in this condition facilitated the presentation or omission of this information as the independent variable manipulation.[1] Ten copies of each of the 90 essays were made, with the writer's freedom rating masked on half. The essays were divided into nine sets, each set containing five assign-pro and five assign-anti essays written by actors who varied in their personal belief on affirmative action. Ten subjects were assigned to each set of essays, five subjects being given copies that included the freedom rating and five being given copies (of the same essays) that masked out the freedom rating. Subjects were asked to read each essay and make two judgments: an estimation of the writer's true attitude on affirmative action (1–10 scale) and an evaluation of the essay's persuasiveness (1 = very low, 10 = very high). Mean attitude attributions and persuasiveness ratings were then computed for each essay based on the judgments of subjects who were ($n = 5$) and were not ($n = 5$) shown the writer's freedom rating.

[1] The use of essays from the observer-informed condition was based strictly on the convenience of having the freedom rating located below the essays. There are no indications that the essays produced in this condition differed systematically from those in the observer-not-informed condition. The mean freedom ratings in the observer-informed conditions (assign pro = 5.17, assign anti = 5.84) did not differ significantly from those in the observer-not-informed conditions (assign pro = 4.89, assign anti = 5.50). In all conditions, subjects who were assigned a position counter to their personal belief reported less freedom than those assigned a position congruent with their belief. Correlations between own attitude and rated freedom for subjects in the observer-informed conditions were: assign pro—$r = .66$, assign anti—$r = -.52$; for subjects in the observer-not-informed conditions; assign pro—$r = .49$; assign anti—$r = -.41$.

Results

Actors' Estimates of Observers' Attributions

The subjects (actors) were first divided into two groups based on their attitude toward affirmative action, obtained in the original attitude inventory. Subjects who checked Responses 1–5 were categorized as *anti*; those checking 6–10 were categorized as *pro*. Subjects' predictions of observers' attributions were submitted to a $3 \times 2 \times 2$ analysis of variance, the independent variables consisting of the writer's essay condition (choice, assign pro, or assign anti), the writer's true opinion (anti or pro), and whether the writer did or did not expect the observer to have access to the writer's freedom rating (observer not informed or observer informed). Mean predictions of observer attributions are shown in Figure 1.

Two main effects achieved significance. Subjects assigned to write a pro essay predicted more pro affirmative action attributions than subjects in the choice condition, with subjects assigned to write an anti essay predicting the most anti attitude attributions ($M_{pro} = 6.39$, $M_{ch} = 5.17$, $M_{anti} = 3.68$), $F(2, 264) = 35.61$, $p < .01$. Subjects who were in fact pro affirmative action predicted more pro attributions than did subjects who were anti affirmative action ($M_{pro} = 6.04$, $M_{anti} = 4.18$), $F(1, 264) = 43.14$, $p < .01$. In terms of the central questions of this research, however, the data are best understood in terms of the Observer Not Informed/Observer Informed × Writer's Attitude interaction, $F(1, 264) = 15.05$, $p < .01$.

Consider the left panel for the observer-not-informed results. Subjects writing essays under the choice conditions predicted that observers would make correspondent attributions. Indeed, subjects under choice conditions invariably do choose to write in defense of their own attitude. However, subjects in the assign-pro and assign-anti conditions anticipated that observers would base their attributions solely on the essay itself. The subject's own attitude bears no relationship to the attribution the subject expected observers to make. This is a clear instance of the prediction of the "fundamental attribution error."

The right-hand panel shows the results for the observer-informed conditions. The result for the choice conditions is very similar to the observer-not-informed case, with subjects predicting correspondent inference. However, subjects in the two assign conditions clearly anticipate that observers will use the freedom rating information. The subjects' own attitudes thus bear a very strong relationship to the attributions they expect informed observers to make. There is still an effect of the particular assignment, however, in the observer-informed condition. Within the writer-anti and writer-pro conditions, the difference between the assign-pro and assign-anti means is significant at the .05 level. However, neither mean is significantly different from the choice-condition mean. There is, in summary, a strong demonstration of actors' predicting that observers will attribute attitudes on the basis of essay direction. If they expect observers to be informed of their actual constraint level, however, there is a marked modification in actors' predicted attributions, with writers expecting their true opinions to "come through" no matter what their particular assignment.

Also of interest is the relationship between the actor's self-rated freedom in the essay task and his or her prediction of observer attributions. For this analysis, only the responses from subjects in the two assign conditions were considered. Subjects were divided into three freedom categories based on their responses to the 10-point scale (1–3 = low; 4–7 = medium; 8–10 = high). Predictions of observer attributions were submitted to a $3 \times 2 \times 2$ analysis of variance (Freedom Level × Assign Pro/Anti × Observer Informed/Not Informed). Mean estimates of observer attributions are shown in Figure 2.

Subjects assigned to write pro essays estimated that observers' attributions would, in general, be more pro affirmative action than did subjects assigned to write anti essays, $F(1, 165) = 110.70$, $p < .01$. Within the pro and anti conditions, as self-rated freedom increased, subjects indicated that they expected observers' attributions to be more extreme in

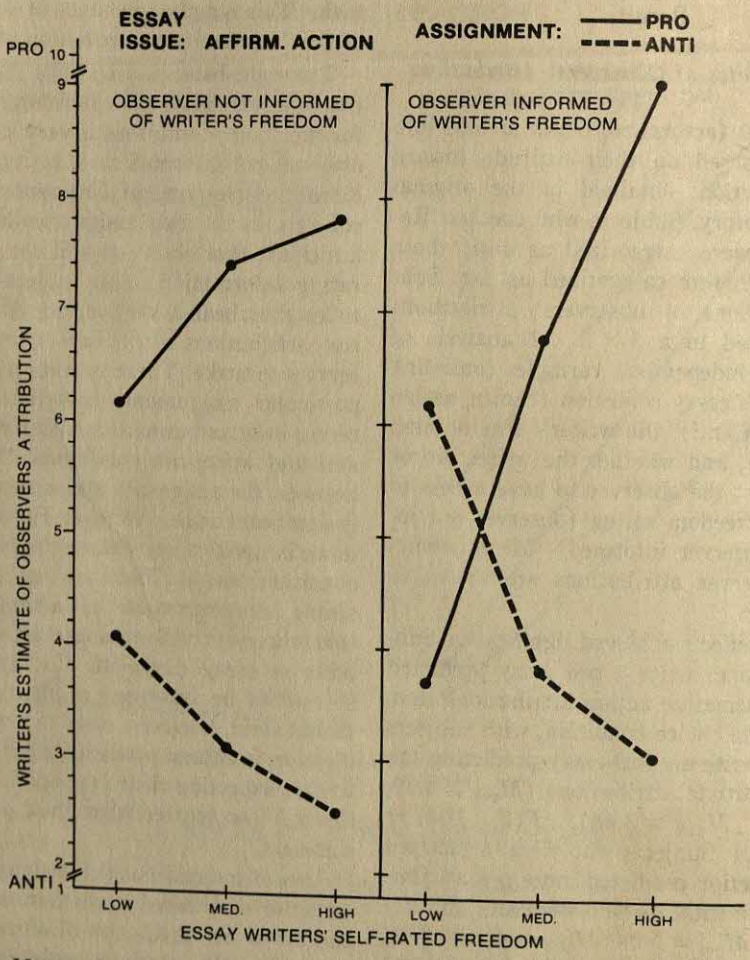


Figure 2. Mean estimate of observers' attitude attributions as a function of writer's freedom, essay condition, and instruction regarding informing observers of writer's freedom (choice condition omitted). (Affirm. = affirmative.)

the direction of their essays, for the Freedom × Condition interaction, $F(2, 165) = 34.42$, $p < .01$. Central to the issue addressed in this research, however, is the three-way Observer × Freedom × Condition interaction, $F(2, 165) = 7.44$, $p < .01$. When essay writers presumed freedom knowledge on the part of the observers (see the right panel of Figure 2), they clearly expected observers to use this information in an appropriate way. Thus, for writers who were under genuine constraint (low freedom), predicted attributions did not reflect their essay assignment condition but were in fact in the direction opposite to that condition. When essay writers did not expect observers to know their freedom ratings, they expected observers to make attributions in line with essay content (see the left panel of Figure 2). There is a clear suggestion, even in the observer-not-informed condition, that subjects experiencing less freedom expected observers to make less extreme attributions. However, even in the low-freedom condition, the assign-pro (6.16) and assign-anti (4.07) means are significantly different ($p < .05$).[2]

[2] To assess the reliability of this result, a replication of the observer-not-informed condition was conducted using a different attitude issue (allowing

From both views of the data (Figures 1 and 2), actors expect observers to base their attributions substantially on the content of the essay, provided that observers are not informed of the actor's actual freedom in the essay task. If actors do expect observers to be given this information, their predicted attributions assume a much different pattern, one that is closely and logically related to the writer's actual attitude and constraint level.

Observer Data

The mean attribution and persuasiveness judgments, which had been computed for each of 90 essays, were first divided into three categories based on the actual freedom ratings given by the writers of the essays (1–3 = low; 4-7 = medium; 8–10 = high). These means were then submitted to a 2 × 3 × 2 (Assign Pro/Anti × Freedom Category × Observer Informed/Not Informed) analysis of variance; means for these analyses are shown in Figure 3.[3]

Ratings of essay persuasiveness were generally similar, with an indication of a somewhat lower quality essay in the low-freedom, assign-anti condition. A Direction × Freedom Category interaction, $F(2, 168) = 4.33$, $p < .05$, indicated that persuasiveness ratings were similar in the three categories of assign pro (low = 5.85, medium = 5.81, high = 5.52), but were somewhat lower for assign-anti essays written with less freedom (low = 4.71, medium = 5.93, high = 6.03).

The attribution data, shown in the lower panels of Figure 3, indicate significant effects for essay direction, $F(1, 168) = 137.98$, $p < .01$, and the Direction × Freedom Category Interaction, $F(2, 168) = 16.58$, $p < .01$. Attitude attributions thus were more positive in the assign-pro than in the assign-anti conditions and were increasingly correspondent with essay directionality under conditions of higher freedom. The latter result, however, was unique to the observer-informed condition. The significant Direction × Freedom Category × Observer Informed/Not Informed interaction, $F(2, 168) = 9.73$, $p < .01$, indicates that freedom category had no influence on attitude attributions in the observer-not-informed conditions, whereas the impact of this information was very clear in the observer-informed condition. Data in the left panel indicate that attitude attributions were based on the content of the essays and reveal no recognition on the part of observers of the wide array of freedom or constraint levels actually experienced by the writers of these essays. The fact that the attributions are similar at each freedom category also suggests that the essays were perceived as similar in persuasiveness across freedom categories.[4] Data in the right panel, however, suggest that observers utilized the freedom ratings as a distinctive clue to the writer's personal attitude, with essay content being totally discounted at the low-freedom level.

Correlations among various judgment items are shown in Table 1. Without the freedom information, observers' attributions of attitude did not relate meaningfully to judgments from the writers (i.e., their real attitude, rated freedom, and predicted attributions). Subjects in the observer-informed conditions, however, attributed attitudes that related substantially and logically to writer judgments. Correlations

undergraduates to own and drive cars on campus). Essay directionality was, as expected, a strong determinant of predicted attributions, $F(1, 77) = 25.85$, $p < .0001$. A Freedom Level × Essay Direction interaction again characterized the data, $F(1, 77) = 4.18$, $p < .05$, with writers under low freedom predicting less correspondent attitudes (assign pro = 6.09, assign anti = 4.68) than writers under high freedom (assign pro = 7.06, assign anti = 3.84).

[3] Note that the same essays constituted the stimuli for the judgments in both the left and right panels of Figure 3. Judgments in the right-hand panels were obtained from subjects having access to the writer's freedom rating; data in the left-hand panels were obtained from subjects not given this information.

[4] If essays written under less freedom were lower in persuasiveness, attributions could reflect this diminution in strength. It would then not be possible to interpret less correspondent attributions as a pure reflection of the observer's knowledge of low freedom ratings per se, in that similarly less extreme attributions would be expected to accompany weaker essays. However, the persuasiveness ratings, as well as the attributions of attitude in the observer-not-informed conditions, strongly suggest that observers were not responding to systematic differences in essay persuasiveness at the various freedom levels.

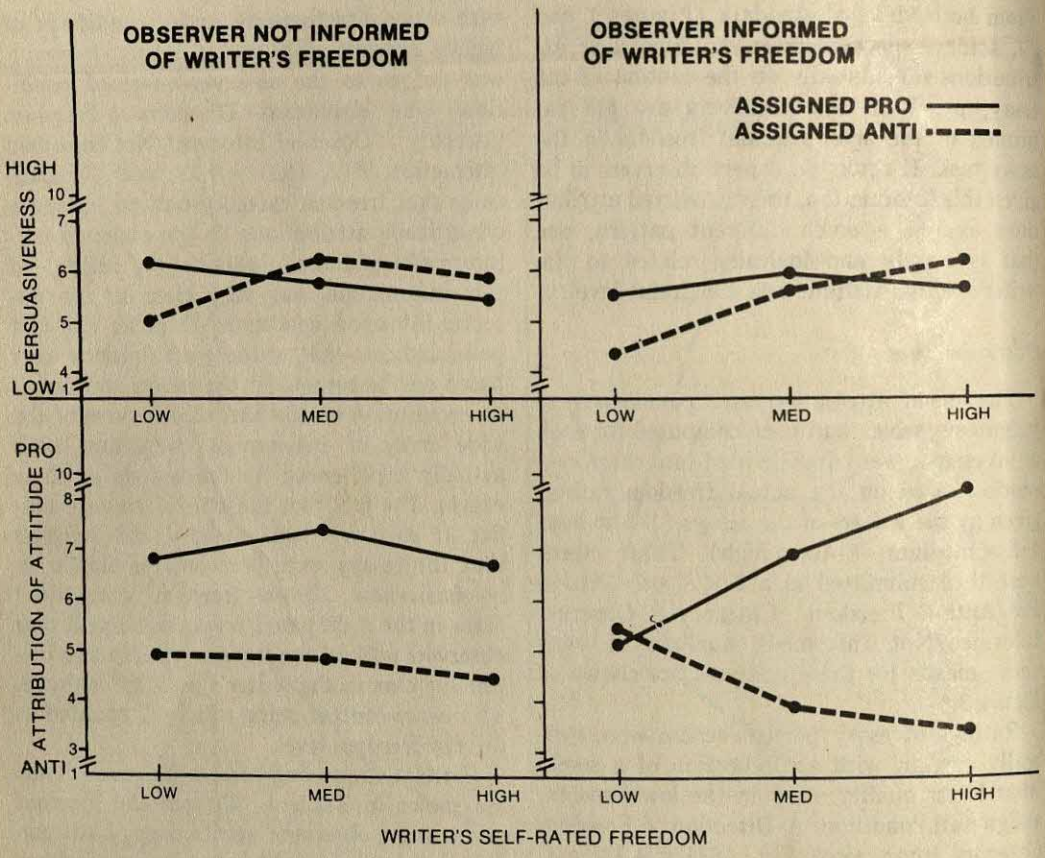


Figure 3. Mean persuasiveness ratings and attitude attributions for essays written at three freedom levels by observers who are or are not informed of the writer's freedom rating.

involving persuasiveness judgments corroborate the analysis of variance results for the data in Figure 3. In the assign-pro conditions, persuasiveness ratings did not correlate significantly with the writer's attitude or freedom ratings, whereas these correlations were somewhat higher in the assign-anti conditions. This suggests, as noted above, that writers in the assign-anti condition who were in favor of affirmative action wrote somewhat less persuasive essays. Finally, the correlations involving the two judgments made by observers —attitude attributions and persuasiveness ratings—were highly significant. The extremity or polarity of attitude attributions, that is, their degree of correspondent inference, related strongly to the perceived persuasiveness of the essays.

Discussion

Actors, in the role of essay writers, predicted that observers would attribute attitudes using the content of the essays—a pattern resembling that of observer data found in many previous experiments and in the present one as well. Even genuinely constrained actors estimated that observers would attach informational value to their essays, although these predictions were less extreme in the direction of essay assignment. Perhaps subjects under high constraint wrote, or felt that they wrote, less persuasive essays. The persuasiveness ratings do not suggest that this was a major factor, although the possibility exists, to a degree, for the anti affirmative action essays. Writers assigned to write against their own beliefs may have expected observers to be

Table 1
Correlations in Experimental Conditions

Item	Observer not informed		Observer informed	
	Assign anti	Assign pro	Assign anti	Assign pro
Observers' attribution and writer's attitude on essay topic	.12	−.19	.44**	$.49_a$**
Observers' attribution and writer's freedom rating	−.20	−.03	$-.59_a$**	$.73_a$**
Observers' attribution and writer's prediction of observers' attribution	.29	−.10	$.63_a$**	$.60_a$**
Observers' persuasiveness rating and writer's attitude	−.30*	−.23	−.36*	−.02
Observers' persuasiveness rating and writer's freedom rating	.23	−.27	.39**	.11
Observers' persuasiveness rating and observers' attribution of attitude	−.60**	.69**	−.61**	.53**

Note. Correlations in observer-informed conditions with subscript *a* differ significantly from corresponding correlations in observer-not-informed conditions at or beyond $p < .05$. $n = 45$.
*$p < .05$.
**$p < .01$.

insightful regarding their constraint. However, this expectation of empathy or accuracy on the part of observers was unwarranted, because observers were completely unresponsive to the constraint variations that actually existed in the assign conditions unless they were specifically informed of the writer's self-rating of constraint.

The present results provide further evidence that perceivers' attributions—or, in this instance, actors' predictions about them—do not accord with the hypotheses derived from correspondent inference theory—a theory that calls for an estimation of the average or modal attitude in the no-choice conditions (Jones & McGillis, 1976, p. 405). The phrase *observer error* or *bias* has been used to describe these results because the perceiver does not appear to appreciate the full implications of randomly assigning the essay position to the writer—that is, the inherent lack of a systematic match between the writer's personal view and the position assigned to him. The constrained actor is conceptually in a particularly ideal position to predict the "correct" (i.e., noncorrespondent) attribution because, relative to the observer, the actor is informationally enriched. The results supported this line of reasoning to a degree, in that actors under low freedom did predict less correspondent attributions. Even these subjects, however, predicted attributions that related to the essay content. It would seem, moreover, equally "unjustified" for subjects in the high-freedom category to predict such extremely correspondent attributions of attitude. There was no logical reason for these actors under high freedom (in the observer-not-informed conditions) to expect that observers would know that they were, in fact, under such minimal constraint. What might account for this persistent tendency to base attributions on essay content and to expect others also to produce such judgments?

Research in a different but related area of attribution theory may be helpful in understanding the basis of the present findings. A number of experiments have addressed the issue of the use of consensus information in the context of Kelley's theory (1967). At issue, specifically, is the tendency for subjects to fail to use such information in a theoretically proper way. The reasons for this particular kind of attribution error are themselves complex. Some have suggested that consensus (i.e., base-rate or normative) information is pallid and abstract vis-à-vis the concrete realism of the behavior per se (Borgida & Nisbett,

1977). Others have reasoned that many research paradigms unwittingly capitalize upon subjects' lack of an intuitive grasp of the meaning of random assignment (Wells & Harvey, 1977). The attitude attribution paradigm would seem vulnerable to the same lines of reasoning. The essay itself, in terms of occupying the perceivers' time and attention, is vivid in comparison to the brief, relatively subtle impact of the no-choice instructions. Furthermore, discounting of the behavior under no-choice instructions requires that subjects understand explicitly the statistical or probabilistic implications of the essay position being randomly assigned to the writer. In this context, Wells and Harvey (1977) obtained judgments that were more in line with theoretical predictions from a group of subjects who were indoctrinated as to the meaning of random assignment. It would seem that, at the least, a manipulation of this magnitude would be required to produce similar means in the assign-pro and assign-anti conditions of the attitude attribution paradigm. To designate what subjects produce in these experiments as "error" or "bias" may, in this view, be questionable. The paradigm requires that the subject attend to an essay and make a difficult attributional response. A model ultimately stipulating that the essay should be totally discounted may itself be committing an "error of implausibility." The fact that attitude attributions are often significantly less correspondent in the assign conditions than in the choice conditions may be more of a judgmental triumph or accomplishment than has generally been acknowledged.

Another interpretation of these findings focuses not on what subjects err in doing or fail to accomplish but instead on the active, cognitive strategies that they bring to their task in this paradigm. A clue to one such strategy may be found in the matter of essay persuasiveness. Mean ratings were generally in the 5–6 range, that is, better than average, although they were based on essays assigned to writers. Actors may have been able to write better essays than they expected to be able to write, particularly when the assignment was not in line with their own attitude. Having written such essays, they may have suspected that observers would also consider the essays to be better than one might expect from someone assigned to defend a position uncongenial to his or her personal belief.

In the essay task, therefore, people may give themselves less than due credit for the ability to produce a noncorrespondent essay. They are better able to comply with the experimenter's assignment than they expect. The essays are then written in a modal range of "average to good." The attributor, however, is still operating on the premise that people assigned a position will do a poor job unless they happen to agree with the assignment. Attributions of attitude in the form of essay-correspondent judgments—what Ross (1977) and Jones (1979) term the "fundamental attribution error"—are, in this view, the logical judgment to make, but for the wrong (a priori) reasons. That is, given the strategy of basing the extremity of one's attribution on perceived persuasiveness, it is logical to attribute attitudes that are, to a degree, in line with the essays—many of the essays *are,* in fact, reasonably persuasive. The error lies in not recognizing that many writers are capable of producing respectable essays even if they personally disagree with what they are writing. There is limited but suggestive support for this line of reasoning. Jones, Worchel, Goethals, and Grumet (1971) included the important variation of presenting *strong or weak* essays to perceivers. Subjects presented with weak essays written under no-choice instructions actually attributed attitudes in opposition to the essay assignment. That is, the pattern of attributions correspondent with essay direction was not only not present but was reversed. In terms of the present line of reasoning, the persuasiveness of the weak (ambivalent) essays in the Jones et al. study may have been less than that expected by perceivers. They then inferred that the writer did not personally endorse the assigned position but in fact believed in the opposite position.

If we consider the many replications of the observer bias phenomenon in attitude attribution (Jones, 1979), the hypothesis that per-

ceivers may consistently be observing essays "better than expected" may be shown to have considerable explanatory power. An important research question involves documenting the expectancy that perceivers entertain regarding the relationship between a person's attitude and the quality of essay that the person is likely to produce. Observers in the present study produced attributions of attitude and ratings of persuasiveness that were highly correlated (Table 1). The causal direction is unclear, although on the basis of the experimental manipulations in the Jones et al. study, it could be assumed that the persuasiveness of the essay was assessed first and provided a rationale for the extremity or correspondence of the attitude attributed to the writer. Thus, the attitude–persuasiveness link appears to be an important element in the perceiver's implicit theory. A demonstrably weak or ambivalent essay would in this view be a much sounder reason for discounting the essay than the probabilistic or statistical implications of no-choice instructions.

Regarding the writer's self-rated freedom, actors, in their predictions, and observers both appeared to use this information to its fullest significance. While previous research in attitude attribution has not used this method of conveying constraint—that is, from the actor's definition of the situation—it is unclear that such a mode of transmitting the constraint or situational information is ill-conceived or beyond the province of theory. This information could, in principle, be underplayed or discounted, just as is the experimenter's instruction of no-choice to the essay writer. One need only remember that jurors found Patricia Hearst guilty despite her low degree of "self-rated freedom." Many professors have looked skeptically at students who mentioned situational causes for missing exams. In the present study, however, the actor's designation of his or her degree of constraint was information of considerable utility. Ratings of low freedom gave perceivers a salient and plausible reason for discounting the observed behavior as indicative of the writer's personal attitude. Actors expected observers to use the information in this way (Figures 1 and 2), and observers did in fact use it (Figure 3).

References

Borgida, E., & Nisbett, R. E. The differential impact of abstract vs. concrete information on decisions. *Journal of Applied Social Psychology*, 1977, *7*, 258–271.

Jones, E. E. How do people perceive the causes of behavior? *American Scientist*, 1976, *64*, 300–305.

Jones, E. E. The rocky road from acts to dispositions. *American Psychologist*, 1979, *34*, 107–117.

Jones, E. E., & Harris, V. A. The attribution of attitudes. *Journal of Experimental Social Psychology*, 1967, *3*, 1–24.

Jones, E. E., & McGillis, D. Correspondent inferences and the attribution cube: A comparative reappraisal. In J. H. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research* (Vol. 1). Hillsdale, N.J.: Erlbaum, 1976.

Jones, E. E., & Nisbett, R. E. The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior*. Morristown, N.J.: General Learning Press, 1971.

Jones, E. E., Worchel, S., Goethals, G. R., & Grumet, J. Prior expectancy and behavioral extremity as determinants of attitude attribution. *Journal of Experimental Social Psychology*, 1971, *7*, 59–80.

Kelley, H. H. Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 15). Lincoln: University of Nebraska Press, 1967.

Miller, A. G. Perceived freedom and the attribution of attitudes. *Representative Research in Social Psychology*, 1974, *5*, 61–80.

Miller, A. G. Constraint and target effects in the attribution of attitudes. *Journal of Experimental Social Psychology*, 1976, *12*, 325–339.

Monson, T. C., & Snyder, M. Actors, observers, and the attribution process: Toward a reconceptualization. *Journal of Experimental Social Psychology*, 1977, *13*, 89–111.

Ross, L. The intuitive psychologist and his shortcomings: Distortion in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10). New York: Academic Press, 1977.

Snyder, M., & Jones, E. E. Attitude attribution when behavior is constrained. *Journal of Experimental Social Psychology*, 1974, *10*, 585–600.

Wells, G. L., & Harvey, J. H. Do people use consensus information in making causal decisions? *Journal of Personality and Social Psychology*, 1977, *35*, 279–293.

Received October 23, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 8, 1432-1438

Inducing Biased Scanning in a Group Setting to Change Attitudes Toward Bilingualism and Capital Punishment

Patrick O'Neill and Diane E. Levings
Acadia University, Wolfville, Canada

Triads of male and female Canadian high school students were assigned to debate either capital punishment or bilingualism. Teams were given 40 minutes to prepare their arguments. Half were told in advance which side of the argument they would be debating, and half were told they would be assigned at the end of their discussion. Scoring of the taped discussions confirmed that this manipulation produced biased scanning of arguments in the predetermined condition and unbiased scanning in the later-determined condition. A postmeasure of attitudes was administered either before or after the actual debate. Biased scanning led to significant attitude change in the predicted direction, and this effect was not influenced by time of presentation of the posttest. The study provides support for the conflict theory of attitude change.

According to conflict theory as articulated by Janis and Mann (1968), attitude change is facilitated by inducing an individual to think of arguments in favor of only one side of a controversial topic. This "biased scanning" makes a set of cognitions salient and tends to shift the subject's attitude toward the position for which he or she has been developing arguments.

Girodo and Strickland (1974) suggested that the conflict theory explanation of attitude change would hold when there was high interest in the issue, when there was opportunity for the subject to think up arguments (i.e., to engage in biased scanning), and when there was incentive for the subject to do so. In their research, the largest change was produced under such conditions. It is assumed that when subjects have high interest in an issue they will have given it some thought in the past, and ideas about the issue will be stored in memory and available for scanning.

The present study attempted to test predictions from conflict theory with two issues that were expected to engage the interest of Canadian high school students: bilingualism and capital punishment. In Canada, as in a number of other countries, there is strong sentiment on the issue of the death penalty (Allmand, 1976; Fattah, 1976). Each new sensational murder revives the arguments for and against executions. The fact that Canadian legislators recently abolished capital punishment at a time when voter sentiment was strongly in favor of retention (Fattah, 1976) helps keep the debate alive.

Bilingualism is as controversial as capital punishment and more personally involves most Canadians. In response to pressures dividing the country, the federal government has actively promoted a policy of bilingualism. A much-disputed aspect of this policy is the attempt to expand the use of the French language in traditionally English-speaking areas. This has often produced a backlash from Anglophones—an outbreak of strong ethnic prejudice against Francophone Canadians. Such a situation occurred in the 1970s in the Annapolis Valley of Nova Scotia, where the Anglophone majority actively opposed the federal government's plan to move a Francophone unit to the local military base. Nova Scotia

This study was supported by a grant from the Canada Council. The authors acknowledge the assistance of Marsha Eisner, Janice Hilchey, Mary McCann, Nancy Gaudet, and Anne Toland, and the help of Susan Scanlon.

Requests for reprints should be sent to Patrick O'Neill, Psychology Department, Acadia University, Wolfville, Nova Scotia, Canada B0P 1X0.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3708-1432$00.75

newspapers reported that such an event would lead to "social deterioration" and warned that it was unsafe to bring Francophones into the area. A Conservative member of Parliament suggested that emotionalism was running so high that it might get out of hand. The French-speaking minority in the area, descendants of the early Acadian settlers in North America, noted with alarm that "the hate and bigotry, for so long dormant" among local Anglophones, was now apparent. The present study was done in the two Anglophone high schools closest to the site of the proposed Francophone "invasion," where it was anticipated that students' feeling of involvement with the issue would be very high.

Laboratory studies of attitudes, including the conflict theory of attitude change, are typically done with individuals rather than groups. However, the use of groups has several advantages. First, in our society many decisions are made, issues are considered, and biased scanning of information often occurs in a group context. Therefore, it is of some interest to know whether the results of studies with individuals generalize to group situations. Second, if a group is induced to discuss an issue, the discussion can be recorded so that biased or unbiased scanning of arguments can be measured directly rather than being merely inferred.

Several studies have used techniques in which subjects generate arguments on one side of an issue because they must defend that position in a debate (Scott, 1957, 1959; Hilchey & O'Neill, Note 1). In the present study, triads of high school students discussed either capital punishment or bilingualism in preparation for a debate on the topic with another "team." The involvement value of these issues has already been noted. Students were also offered incentives in the form of cash prizes for winning teams and further cash awards as grand prizes. The purpose of the study was to compare biased versus unbiased scanning of arguments directly to determine whether attitude change would be a function of biased scanning. Half of the subjects were told before their preparation session which side of the issue they would debate, and half were not assigned until after their preparation session. It was anticipated that the former condition would lead to relatively more biased scanning of arguments and consequently to more attitude change.

It might be argued that while biased scanning produces more attitude change initially, such persuasive effects fade away when exposed to counterarguments. This might be the case, because if biased scanning is employed, the subject is not innoculated against arguments on the other side of the issue (McGuire & Papageorgis, 1971). To check this possibility, for half of the subjects the postmeasure was given following the preparation session and before the debate. For the other subjects the postmeasure was given following the debate, when those in the biased scanning condition would have been faced with counterarguments that they had not considered.

Because no pretest was used in the present study to minimize reactivity (Campbell, 1969; Campbell & Stanley, 1966), it may be useful to comment on the initial positions of the subjects that can be deduced without a premeasure. As indicated, ethnic prejudice against French Canadians was running high in the target area at the time of the study. On capital punishment, Hilchey (Note 2) reviewed evidence to indicate that Canadians in general and Nova Scotians in particular tend to be in favor of the death penalty. She found, in a university near the site of the present research, that students were overwhelmingly in favor of capital punishment. It might be assumed that this was also the general sentiment of the high school population from which our sample was drawn.

Pretest of the Instrument

Before running the present study, a postmeasure of attitude change was designed specifically to focus on the issues of bilingualism and capital punishment. A disguised technique similar to Cook's measure of plausibility (Brigham & Cook, 1970; Waly & Cook, 1965) was used. Subjects were asked to "rate how convincing these arguments are; circle the appropriate number after each argument." The numbers ranged from 1 (not convincing) to 7 (very convincing) with no other labels.

There were 13 arguments. Only 8 were scored in this study, 4 on the bilingualism issue and 4 on capital punishment; the other arguments were on women's liberation and were "fillers." On each issue, two arguments supported one side and two supported the other. The following are examples of arguments about capital punishment:

If a person is executed for murder, we can be absolutely certain that he will not kill again. Any other sentence always leaves open the possibility that an appeal board will mistakenly free a dangerous killer, or that he may escape from prison. In cases where guilt is beyond doubt, dangerous murderers should be given the death penalty.

Many jurors are extremely reluctant to convict when the death penalty may be the result. There would certainly be more convictions if juries knew that life imprisonment was the most serious penalty that could be applied. It is better for law and order to have two murderers convicted and sent to prison than to have one convicted and executed and the other set free. Therefore, capital punishment should be abolished.

The following are 2 examples of arguments on the bilingualism issue:

People of many races, cultures and languages have played their part in the development of Canada. This is not just a country of English and French. It makes as much sense to say that we should all speek Cree and Inuit, Greek and Italian, as to say that we should all speak French and English. To expect that all Canadians will be fluent in more than their own language is simply impractical.

Canadians must deal with the fact that there are two major languages in the country. Each language is spoken by so many people that this can never be a country where everyone speaks one language, like the United States. If the country is to remain united, both major groups must make an effort to understand and communicate with one another; to do that, they must overcome the language barrier.

This measure[1] was pretested on university students in introductory psychology classes. They were given the questionnaire while being told that the experimenter planned to do research on debating and needed to know how "convincing" these arguments seemed. Scores on items were added to get separate totals on the capital punishment and bilingualism issues. Two months later, a different experimenter telephoned students who had scored highest or lowest on each of the issues. There were 10 high and 10 low on each issue. Students came to the laboratory individually. They were told that for some future research we needed to develop a number of arguments on current affairs topics. Topics assigned were feminism and either capital punishment or bilingualism. For the two target issues, the statements were (a) "We need to bring back capital punishment to lower the crime rate in Canada" and (b) "All Canadians should learn to speak both French and English to make the country fully bilingual."

Table 1
Distributions of Validation Subjects on the Questionnaire and Side Chosen in Writing an Essay

Essay	Questionnaire: In favor	Questionnaire: Against
Capital punishment issue		
In favor	8	2
Against	0	10
Bilingualism issue		
In favor	9	1
Against	0	10

The students were told that they should take one side of the issue and write an essay containing as many arguments as they wished. They were told, "It doesn't matter to us which side of the argument you pick; it's probably going to be easier if you pick the one you agree with." Of course, the side picked was exactly what interested us. Students in this validation study could be classified as in favor of or against capital punishment or bilingualism on the basis of scores on the target instrument and again on the basis of the side they chose when writing an essay 2 months later. The results are presented in Table 1. They are in the predicted direction for both issues, significant beyond the .005 level using the Fisher's exact test. The reliability of the measure was estimated using the Spearman-Brown formula. On the pretest, the reliability was .48 for the bilingualism ques-

[1] Copies of the instrument are available on request from the first author.

tions and .71 for capital punishment. In the main study the reliability was .40 for bilingualism and .64 for capital punishment.

Method

Subjects

The study involved 48 students, 24 males and 24 females, from two high schools, all from Grades 11 and 12.

Procedure

First contact with subjects was made in their classrooms, where they were told that they would be invited to take part in research on debating, that no previous experience in debating was necessary, that participation was voluntary, and that cash prizes of $5 would be given to each person on the winning side in a debate, with an additional $25 given to the three best debaters in each school. The night before each session, six males or six females were chosen at random from class lists and were telephoned, invited to take part, and given a time and place. If a student was not available, same-gender replacements were taken randomly from class lists.

Sessions were held immediately after school hours in two classrooms. When the six students arrived, they were met by two experimenters. Subjects were divided into triads alphabetically according to the first letter of the last name. Each experimenter then took three subjects to each of the rooms. The following instructions were given to the assignment-before-discussion group:

> As you know, we are doing research on debating. The procedure is not very complicated. You will be assigned a current affairs topic (and you will be told which side of the issue to debate). You will have 40 minutes to discuss the topic and consider what arguments you might present. As you know, another group is meeting in another room to discuss the same question (and they will be assigned the other side of the issue). At the end of the 40-minute discussion we will bring the two groups together for the actual debate.

The modifications in instructions made for the assignment-after-discussion condition were the omission of the phrases in parentheses and the insertion of the sentence "At the end of the discussion I will tell you which side of the issue you must present in the debate; of course the other group will be assigned to the other side of the issue." They were told that the debates would be taped and would later be evaluated by two judges at the university to determine the cash awards. Subjects were then told:

> We will also be taping your discussion now, before the debate—but that's just for our own records. The judges will *not* hear it, and of course it will have no bearing on whether you win or lose.

In the assignment-before condition, subjects were informed:

> I'll give you a slip of paper telling you what topic you are to debate and which side you are on. This has been determined randomly. Here is your topic.

The experimenter then opened a sealed envelope and handed a slip of paper to a member of the group. This slip provided both the statement of the issue and the position (in favor or against). The procedure was similar for the assignment-after condition, except that only the issue—and not the side to be debated—was given before the discussion. The issues stated were (a). "We need to bring back capital punishment to lower the crime rate in Canada and (b) "All Canadians should learn to speak both French and English to make the country fully bilingual." The discussions were taped on cassette tape recorders using small omnidirectional lapel microphones that were hung from the ceiling.

At the end of the discussion period, the posttest was administered to half of the triads. In each room, the experimenters told their groups:

> While we're waiting, I'd like to get you to do a small task for me. This will help us with another research project. We have a number of different current affairs topics, and we're not sure how convincing certain arguments are about those topics. I'd like you to help by reading over these arguments, one at a time, and rating them to indicate how convincing they are. We'll use the arguments that are judged most convincing and discard the others.

After this the two triads were brought together, and the debate took place. Each debater was given 5 minutes for initial presentation of argument, with speakers from each side alternating. Then each side had another 5 minutes for rebuttal. For half of the groups the posttest was administered following the debate, using instructions similar to those above.

Design

To determine whether time of assignment was successful in producing biased versus unbiased scanning, each triad was treated as a subject in a 2×2 factorial, with time of assignment as one factor (before or after the predebate discussion) and issue as the second factor (capital punishment or bilingualism). Testing for attitude change effects, with each individual now constituting a subject, another 2×2 factorial was employed with time of assignment as one factor and time of administration of the attitude measure as the second factor (before or after the debate.[2]

[2] It will be noted that in the attitude change portion of the study, "issue" was not used as a factor but was counterbalanced within conditions. In the

Results

There were three major questions of interest in the present study.

1. Were we successful in producing biased versus unbiased scanning in the assignment conditions?

2. If so, did biased scanning lead to more attitude change?

3. If so, was this attitude change affected by the time of presentation of the posttest (i.e., before or after the debate)?

Biased Versus Unbiased Scanning

Each 40-minute tape was scored by one of four experienced raters who had been trained in a previous study that used the same scoring procedure (Hilchey & O'Neill, Note 1). Each time an argument was raised in group discussion, the rater briefly noted the content of the argument and indicated whether it was for or against the issue. In cases where there was doubt, one of the other raters would be consulted for a second opinion. The definition of an argument, taken from Fowler and Fowler (1969), was "a reason advanced for or against." Examples of statements not counted as arguments included "I'm against capital punishment" and "we studied capital punishment in Grade 10." Generally, however, the concept of an argument was interpreted liberally, with any reason given for or against a position being scored. Each time an argument was advanced, it was counted. Thus, for example, a group member might say "bilingualism will unite our country" several times during the discussion, and this would be noted as a probilingualism argument each time it was raised. The quality of the arguments was not an issue, merely the extent to which the group considered one side or both sides of the question. When the tapes were scored, a ratio of pro and anti arguments was computed, varying between 0 and 1, with 1 indicating a perfect balance and lower numbers indicating a tendency to focus on only one side of the issue. Mean scores for the assignment-before-discussion condition were bilingualism issue = .19, capital punishment issue = .11; for the assignment-after condition, mean scores were bilingualism issue = .55, capital punishment issue = .87. In a 2×2 analysis of variance, looking at time of assignment and issue, there was a main effect for time of assignment, $F(1, 12) = 124.0$, $p < .001$, with the assignment-before condition producing more biased scanning; there was a main effect for issue, $F(1, 12) = 6.0$, $p < .05$, but this effect can best be explained in terms of the significant interaction between issue and time of assignment, $F(1, 12) = 17.0$, $p < .005$. When subjects did not know which side of the issue they would be required to argue, they were much more unbiased in their consideration of capital punishment than they were in their consideration of bilingualism. This effect was substantiated by the Newman-Keuls procedure, which showed a significantly higher score (more unbiased discussion) for capital punishment than for bilingualism within the assignment-after condition.

Attitude Change

Because we were successful in producing biased scanning, it was expected that there would be more attitude change in the biased scanning (assignment-before) condition. A possible confounding variable was the actual

---

usual laboratory study of persuasion, where issues have less social impact, a design including issue as a factor would typically be employed even when this factor is expected to have no effect. In the present study, however, ethical concerns prevented us from employing the assignment-before condition, which we expected to produce the most attitude change, to increase ethnic prejudice. The assignment-before condition was therefore always used for the "in favor" side on bilingualism and the "against" side on capital punishment. Because the assignment-after condiiton should not produce biased scanning and thus should have less impact on attitudes, this condition was used for the antibilingualism and the procapital punishment sides. This permitted a procedure where triads could be assigned to each side of the issue, and following their separate discussions they could be brought together for a debate. In summary, each cell of the second 2×2 factorial contains an equal number of males and females, an equal number arguing in favor and against an issue, and an equal number discussing bilingualism and capital punishment.

Table 2
Mean Scores on the Posttest for Assignment to Discussion and Time of Attitude Measurement

Issue	Assignment before discussion (biased scanning)		Assignment after discussion (unbiased scanning)	
	Attitude measure		Attitude measure	
	Before debate	After debate	Before debate	After debate
Bilingualism	5.25	4.79	4.29	3.92
Capital punishment	4.79	5.25	3.98	3.71
M	5.02	5.02	4.14	3.81

Note. Attitude posttest was scored in the direction of assignment.

direction of assignment, which was expected to have an effect itself. This was controlled by scoring the posttest in the direction of assignment. Thus, the mean score on the 7-point scale indicated amount of agreement with either the "in favor" position or the "against" position, whichever was assigned. Mean scores on the postmeasure are presented in Table 2. Each cell included an equal number of males and females, an equal number debating in favor and against, and an equal number debating bilingualism or capital punishment. There was a significant main effect for time of assignment, $F(1, 40) = 21.01$, $p < .001$, with biased scanning producing more attitude change. Neither the main effect of time of attitude measurement nor the interaction was significant, indicating that the persuasive effects of biased scanning were not affected by listening to opposing arguments during the debate. The analysis of group discussion indicated that in the assignment-after-discussion condition, discussion was somewhat more unbiased on the capital punishment issue than on bilingualism, and in view of this fact similar differences in attitude change might be expected. Table 2 indicates that there was a trend in this direction, although it was nonsignificant.

Discussion

Conflict theory suggests that when a person is induced to think of arguments for only one side of an issue, those arguments become salient and influence attitudes. The theory previously received support in studies with subjects run individually. The present research indicates that the effect can be produced in a small group setting.

An advantage of using groups and recording group discussion is that biased or unbiased scanning of arguments is made public; researchers can hear and score arguments and quantify the degree of bias toward one side or the other. In studies with individuals, the experimenter is typically left in the position of inferring that biased scanning has taken place, reasoning backward from the results (e.g., Girodo & Strickland, 1974). An obvious disadvantage in the use of groups is the lack of control over all the factors that influence attitude change. We cannot be sure to what extent a subject is changing his or her attitude as a function of self-generated arguments or because he or she hears persuasive arguments from others in the group. Indeed, these factors may interact so that potential debaters are more receptive to one-sided communications from others (see Greenwald, 1969, 1970). While such questions cannot be answered in an experiment such as the present one, these factors were not the focus of the research and would have no systematic influence on the results.

Half of the subjects were not given the postmeasure until after they had debated the issue and had heard counterarguments. It is known that subjects can be "immunized" against counterarguments by hearing them in

mild form before such arguments are presented with persuasive intent (McGuire & Papageorgis, 1961). In this study, however, the attitude change created by biased scanning was not affected by the time of the posttest. Still, caution should be used in generalizing this result beyond the present experimental procedure. There may be some aspects of debating per se, including the competitive aspects involved and the probable arousal of reactance (Brehm, 1966), that make counterarguments presented during a debate ineffective in persuading the opposing side.

An unexpected finding was the interaction between issue and time-of-assignment in producing biased scanning. For both issues, groups that had to prepare for the debate without knowing their eventual assignment were much more unbiased in their consideration of arguments than were groups in the assignment-before condition. But within the unbiased condition, subjects were significantly more balanced in their treatment of capital punishment than bilingualism. This study was done in a geographic area where bilingualism is a matter of considerable debate and concern; presumably capital punishment, however interesting it might be to the subjects, touched few of them directly. This provides a footnote to the point made by Girodo and Strickland (1974) that "high interest" in an issue promotes biased scanning and facilitates attitude change. It seems that high interest in the sense of personal involvement with an issue also makes it more difficult to create *unbiased* scanning of arguments, even when there is high incentive to look at both sides of the question.

Reference Notes

1. Hilchey, J. K., & O'Neill, P. *Using a debate technique to change attitudes toward capital punishment.* Paper presented at the annual meeting of the Canadian Psychological Association, Ottawa, June 1978.

2. Hilchey, J. K. *Testing theories of decision-making and attitude change with capital punishment as the target issue.* Unpublished honors thesis, Acadia University, Wolfville, Canada, 1978.

References

Allmand, W. G. *On the abolition of capital punishment.* Ottawa, Canada: Queen's Printer, 1976.

Brehm, J. W. *A theory of psychological reactance.* New York: Academic Press, 1966.

Brigham, J. C., & Cook, S. W. The influence of attitude and judgments of plausibility: A replication and extension. *Educational and Psychological Measurement,* 1970, *30,* 283–292.

Campbell, D. T. Reforms as experiments. *American Psychologist,* 1969, *24,* 409–429.

Campbell, D. T., & Stanley, J. C. Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching.* Chicago: Rand McNally, 1966.

Fattah, E. A. *The Canadian public and the death penalty: A study of a social attitude.* Ottawa, Canada: Queen's Printer, 1976.

Fowler, F. G., & Fowler, H. W. *The pocket Oxford dictionary of current English* (5th ed.). Oxford, England: Clarendon Press, 1969.

Girodo, M., & Strickland, L. H. A conflict-theory interpretation of attitude change on forced compliance situations. *Canadian Journal of Behavioural Science,* 1974, *6,* 385–397.

Greenwald, A. G. The open-mindedness of the counterattitudinal role player. *Journal of Experimental Social Psychology,* 1969, *5,* 375–388.

Greenwald, A. G. When does role playing produce attitude change? Toward an answer. *Journal of Personality and Social Psychology,* 1970, *16,* 214–219.

Janis, I. L., & Mann, L. A conflict-theory approach to attitude change and decision making. In A. G. Greenwald, T. C. Brock, & T. M. Ostrom, (Eds.), *Psychological foundations of attitudes.* New York: Academic Press, 1968.

McGuire, W. J., & Papageorgis, D. The relative efficacy of various types of prior belief-defense in producing immunity against persuasion. *Journal of Abnormal and Social Psychology,* 1961, *62,* 327–337.

Scott, W. A. Attitude change through reward of verbal behavior. *Journal of Abnormal and Social Psychology,* 1957, *55,* 72–75.

Scott, W. A. Attitude change by response reinforcement: Replication and extension. *Sociometry,* 1959, *22,* 328–335.

Waly, P., & Cook, S. W. Effect of attitude on judgments of plausibility. *Journal of Personality and Social Psychology,* 1965, *2,* 745–749.

Received October 30, 1978 ■

Editors Named for New *Journal of Personality and Social Psychology*

The Publications and Communications Board has named Melvin Manis, Ivan D. Steiner, and Robert Hogan to edit the *Journal of Personality and Social Psychology* (JPSP), which will appear in a new format in 1980. The affiliations of the new editors are, respectively, University of Michigan and Veteran's Hospital, Ann Arbor; University of Massachusetts; and Johns Hopkins University. Manis will edit the section on ATTITUDES AND SOCIAL COGNITION, Steiner the section on INTERPERSONAL RELATIONS AND GROUP PROCESSES, and Hogan the section on PERSONALITY PROCESSES AND INDIVIDUAL DIFFERENCES.

An announcement on the restructuring of JPSP appeared on p. 668 of the June 1978 issue of JPSP. As indicated at that time, the three editors will work independently with separate groups of advisory editors. Their acceptances will appear as distinct sections in monthly issues of JPSP. Detailed descriptions of each section's coverage appear on pages 468–469 of the March issue of JPSP.

Authors should submit manuscripts to the appropriate section editor as follows:

ATTITUDES AND SOCIAL COGNITION

Melvin Manis
Research Center for Group Dynamics
Institute for Social Research
University of Michigan
Ann Arbor, Michigan 48106

INTERPERSONAL RELATIONS AND GROUP PROCESSES

Ivan D. Steiner
Department of Psychology
University of Massachusetts
Amherst, Massachusetts 01003

PERSONALITY PROCESSES AND INDIVIDUAL DIFFERENCES

Robert Hogan
Johns Hopkins University
Department of Psychology
33rd and Charles Streets
Baltimore, Maryland 21218

Journal for the Theory of Social Behaviour

Edited by P. F. SECORD, (Department of Urban Studies, Queens College, CUNY) and R. HARRÉ (Linacre College, Oxford). Associate Editors. P. MANICAS (Department of Philosophy, Queens College, CUNY) and C. SMITH (Department of Sociology, Queens College, CUNY).

Volume 9, No. 1 March 1979

CONTENTS

MICHAEL GORR: Agency and Causation

CATHERINE COBB MOROCCO: The Development and Function of Group Metaphor

MARVINA C. RICH: Verbal Reports and Mental Processes: Issues of Accuracy and Awareness

WILLIAM D. GEAN: Emotion, Emotional Feeling and Passive Bodily Change

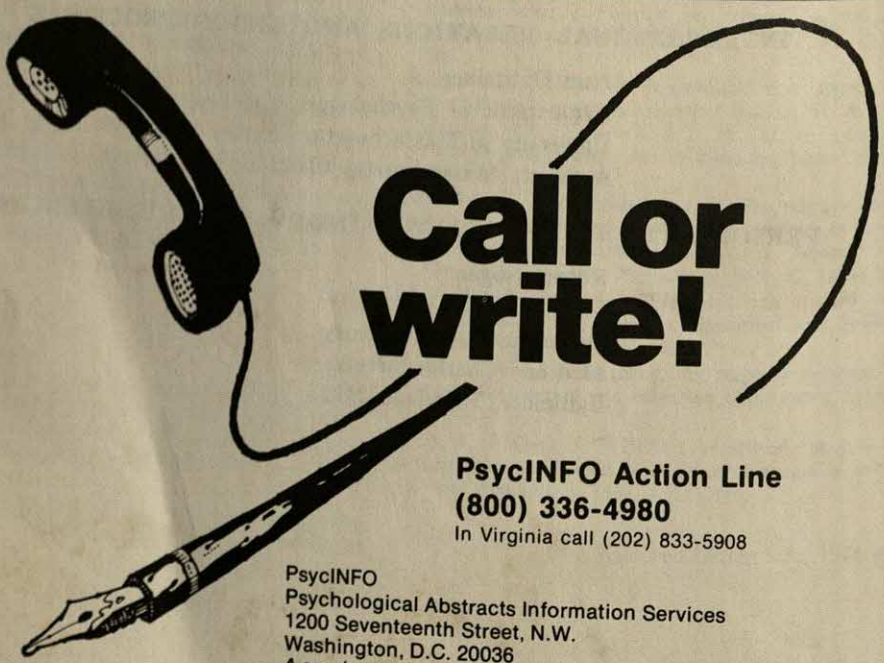
JAMES RUSSELL: The Status of Genetic Epistemology

ROBERT L. ARRINGTON: Practical Reason, Responsibility and the Psychopath

STANFORD M. LYMAN: Stewart Culin and the Debate over Trans-Pacific Migration

Annual subscriptions: Institutions £14.00 (inland) £16.80 (overseas) $35.00 (USA and Canada); Individuals £7.50 (inland) £9.00 (overseas) $18.50 (USA and Canada)

Orders to: JOURNALS DEPARTMENT, BASIL BLACKWELL & MOTT LTD., 108 Cowley Road, Oxford OX4 1JF, England.

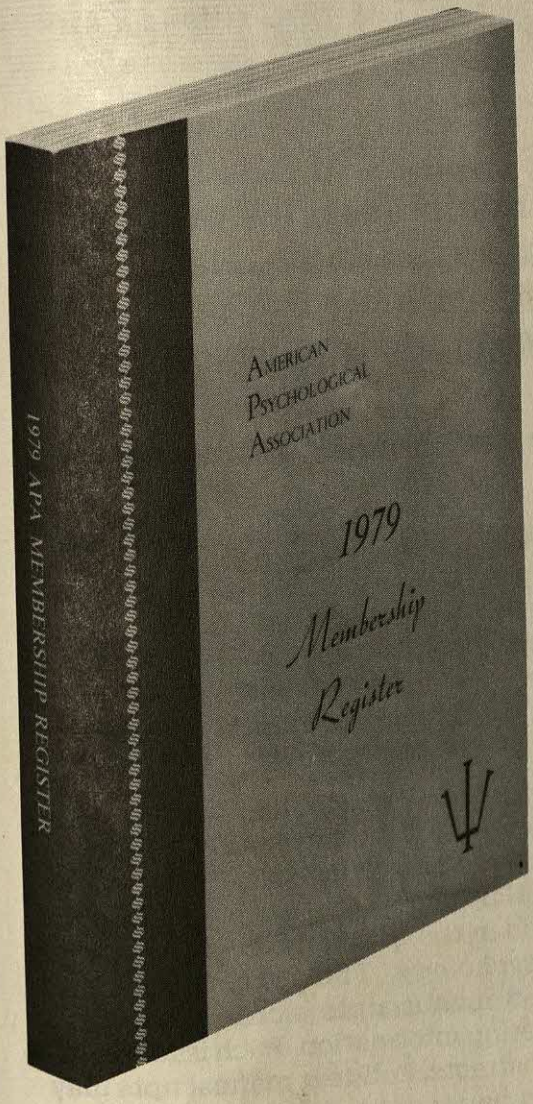


Call or write!

PsycINFO Action Line
(800) 336-4980
In Virginia call (202) 833-5908

PsycINFO
Psychological Abstracts Information Services
1200 Seventeenth Street, N.W.
Washington, D.C. 20036
A service of the American Psychological Association

PsycINFO, the database of record in psychology and related disciplines



Now Available

APA's Convenient, Alphabetical Listing of over 49,000 Fellows, Members, and Associates includes:

- **Telephone Numbers — office and/or home**
- **Current Mailing Address**
- **APA Membership and Divisional Status**
- **Divisional Membership Rosters— for each of APA's 36 divisions**
- **ABPP and ABPH Diplomates Rosters**

Ordering Information

The **1979 Membership Register** is available to APA members for $8.00 and to nonmembers for $15.00. Checks should be made payable to the American Psychological Association and must accompany all orders of $25.00 or less.

Please mail your orders to:
American Psychological Association
Order Department
1200 17th St., N.W.
Washington, D.C. 20016

The Journal Supplement Abstract Service is an information service providing ready access to psychology-related materials not available through conventional communication channels.

JSAS Selects and Disseminates:

- Educational materials
- Massive data collections
- Methodological techniques and procedures
- Major projects in progress
- Descriptions of effective techniques or programs
- Technical reports
- Fresh looks at controversial issues
- Invited lectures
- Management of psychological resources
- APA task force reports
- Bibliographies
- Literature reviews

Manuscripts are accepted by a panel of editors. They meet the same high standards set for American Psychological Association's journals, but are not bound by conventional restrictions of format, length, or editorial policy.

Available titles are listed in the quarterly JSAS *Catalog of Selected Documents in Psychology.* The catalog includes abstracts of newly accepted manuscripts along with manuscript length, price, and ordering information. Each issue of the catalog contains about 50 abstracts. Full-text manuscripts may be purchased individually in paper or microfiche.

Subscription rates for the quarterly catalog are $6 for APA members and $14 for nonmembers (foreign, $16). Please include full remittance for orders of $25 or less. To order, make check payable to APA and mail to:

American Psychological Association
Subscription Dept.
1200 17th Street, NW
Washington, DC 20036

(Continued from front cover)

A Creative Personality Scale for the Adjective Check List 1398
Harrison G. Gough

Conflict and Avoidance in the Helping Situation 1406
S. Mark Pancer, Linda M. McMullen, Randal A. Kabatoff, Kent G. Johnson, and Carole A. Pond

The Importance of Consistency of Modeling Behavior Upon Imitation: A Comparison of Single and Multiple Models 1412
Peter A. Fehrenbach, David J. Miller, and Mark H. Thelen

Self-Serving Attributional Biases: Perceptual or Response Distortions? 1418
Gifford Weary

The Bias Phenomenon in Attitude Attribution: Actor and Observer Perspectives 1421
Arthur G. Miller, Robert Baer, and Peter Schonberg

Inducing Biased Scanning in a Group Setting to Change Attitudes Toward Bilingualism and Capital Punishment 1432
Patrick O'Neill and Diane E. Levings

Editors Named for New *Journal of Personality and Social Psychology* 1439

17/9/80

Willo P. White, editor

Resources in Environment and Behavior

NEW FROM APA IN 1979

An invaluable sourcebook for students, instructors, and researchers in the new field of environment and behavior. This solid reference includes:

- Overview and history of this emerging field
- Graduate programs—both formal and informal
- Teaching innovations introduced in the United States, Canada, and Great Britain
- Funding sources
- Career opportunities
- Directory of key individuals currently working in the field
- Annotated bibliography
- Listing of relevant journals

Resources in Environment and Behavior is available from the American Psychological Association for $10 (soft cover only). To order, make your check payable to APA.

American Psychological Association
Order Department
1200 17th Street, NW
Washington, DC 20036

Please include full remittance for orders of $25 or less.

JOURNAL OF

Personality AND Social Psychology

VOLUME 37 NUMBER 9 · SEPTEMBER 1979

On Problems With the Cause–Reason Distinction in Attribution Theory 1441
John H. Harvey and Jalie A. Tucker

Causal Explanation, Teleological Explanation: On Radical Particularism in Attribution Theory 1447
Arie W. Kruglanski

On the Relationship Between Causes and Reasons 1458
Allan R. Buss

Competence and the Overjustification Effect: A Developmental Study 1462
Ann K. Boggiano and Diane N. Ruble

The Role of Social Comparison in Choice Shifts 1469
George R. Goethals and Mark P. Zanna

Some Child-Rearing Antecedents of Criminal Behavior in Adult Men 1477
Joan McCord

Are Causal Attributions Causal? A Path Analysis of the Cognitive Model of Achievement Motivation 1487
Martin V. Covington and Carol L. Omelich

Uncertainty-Reducing Properties of Achievement Tasks 1505
Yaacov Trope

The Role of Facial Response in the Experience of Emotion 1519
Roger Tourangeau and Phoebe C. Ellsworth

Deindividuation and Valence of Cues: Effects on Prosocial and Antisocial Behavior 1532
Robert D. Johnson and Leslie L. Downing

Social and Emotional Messages of Smiling: An Ethological Approach 1539
Robert E. Kraut and Robert E. Johnston

The Effects of Mediator Rewards and Suggestions Upon Negotiations 1554
James A. Wall, Jr.

(Continued on inside back cover)

Published monthly by the

AMERICAN PSYCHOLOGICAL ASSOCIATION

ISSN 0022-3514

Acting Editor

Clyde Hendrick, *University of Miami*

Action Editors

Irwin Altman, *University of Utah*
Charles S. Carver, *University of Miami*
Richard J. Harris, *University of New Mexico*
John H. Harvey, *Vanderbilt University*
William Ickes, *University of Missouri–St. Louis*
George Levinger, *University of Massachusetts—Amherst*
Brendan Gail Rule, *University of Alberta, Edmonton, Canada*
David R. Shaffer, *University of Georgia*
Cookie White Stephan, *New Mexico State University*
Barbara Strudler Wallston, *George Peabody College for Teachers*

Consulting Editors

Robert Steven Baron, *University of Iowa*
Paul A. Bell, *Colorado State University*
Ellen Berscheid, *University of Minnesota (Minneapolis)*
Michael H. Birnbaum, *University of Illinois at Urbana-Champaign*
Jack Brehm, *University of Kansas*
Philip Brickman, *Northwestern University*
J. Merrill Carlsmith, *Stanford University*
Paul Christopher Cozby, *California State University, Fullerton*
Rick Crandall, *Texas Christian University*
James M. Dabbs, Jr., *Georgia State University*
John M. Darley, *Princeton University*
Andrew R. Davidson, *University of Washington*
James H. Davis, *University of Illinois at Urbana-Champaign*
Kay Deaux, *Purdue University*
Marshall Dermer, *University of Wisconsin—Milwaukee*
Edward Donnerstein, *Iowa State University*
Ebbe B. Ebbesen, *University of California, San Diego*
Yakov M. Epstein, *Rutgers—The State University*
Irene Hanson Frieze, *University of Pittsburgh*
J. Martin Giesen, *Mississippi State University*
Lewis R. Goldberg, *University of Oregon*
William G. Graziano, *University of Georgia*
Alan E. Gross, *University of Maryland (College Park)*
Nancy Hirschberg, *University of Illinois at Chicago Circle*
Chester A. Insko, *University of North Carolina at Chapel Hill*
Alice M. Isen, *University of Maryland, Baltimore County*
Martin F. Kaplan, *Northern Illinois University*
Samuel S. Komorita, *University of Illinois at Urbana-Champaign*
Vladimir J. Konečni, *University of California, San Diego*
Ellen Langer, *Harvard University*
Mark R. Lepper, *Stanford University*
Darwyn E. Linder, *Arizona State University*
Dale T. Miller, *University of British Columbia, Vancouver, Canada*
Judson Mills, *University of Maryland (College Park)*
Warren T. Norman, *University of Michigan (Ann Arbor)*
Michael Ross, *University of Waterloo, Waterloo, Canada*
Michael F. Scheier, *Carnegie-Mellon University*
Barry R. Schlenker, *University of Florida*
Kelly G. Shaver, *College of William & Mary*
Phillip Shaver, *New York University*
Marvin E. Shaw, *University of Florida*
Steven J. Sherman, *Indiana University*
Harold Sigall, *University of Maryland (College Park)*
Mark Snyder, *University of Minnesota (Minneapolis)*
Melvin L. Snyder, *University of Texas at Austin*
Shelley E. Taylor, *Harvard University*
Stephen G. West, *Florida State University*
Steve Worchel, *University of Virginia*
Mark P. Zanna, *University of Waterloo, Waterloo, Canada*

Assistant to the Editor
Sally Miller

APA Journal Staff

Anita DeVivo, *Executive Editor*
Ann I. Mahoney, Manager, *Journal Production*
Marjorie M. Henry, *Production Supervisor*
Robert J. Hayward, *Advertising Representative*
Juanita Brodie, *Subscription Manager*

The Journal of Personality and Social Psychology publishes original reports of research, methodology, theory, criticism, and literature review in the fields of personality and social psychology of normal processes and populations. Submitted manuscripts should delineate an important question and provide a clear answer to it. In the case of research reports, this will often require demonstration that the reported results are generalizable beyond a single research setting. Consult the editorial in the January 1976 issue (Vol. 33, pp. 1–7) for a detailed statement of policies and instructions to authors.

Manuscripts: Submit manuscripts in quadruplicate, prepared according to the *Publication Manual of the American Psychological Association* (2nd ed.). Copies should be highly legible and will not be returned to authors unless they are original typewritten copies. Mail manuscripts according to instructions on p. 1439 of the August issue.

Required supplementary information: Manuscripts cannot be reviewed for publication unless the author provides supplementary information in the following categories: (a) a cover letter that indicates the relation (or the lack of relation) of the submitted manuscript to other papers by the same author(s) that are currently submitted, in press, or recently published; (b) one copy of each such related paper; (c) one copy of a statement of the procedures of the research reported in the submitted manuscript in sufficient detail to serve as a guide for conduct of an exact replication; and (d) one copy of the summary tables (e.g., correlation tables, analysis of variance summaries) for the *major* analyses reported. If it appears that the supplementary materials in categories (c) or (d) might be bulky, the author should exercise judgment and restraint in selecting what to send. It is not necessary to send supplementary material that repeats verbatim manuscript content.

After a manuscript is accepted, authors will be expected to provide assurance that their data are available in unanalyzed form for reanalyses by other investigators and that their research procedures were consistent with the American Psychological Association's published principles of research ethics, except insofar as may be detailed otherwise in the manuscript.

Copyright and Permission: All rights reserved. Written permission must be obtained from the American Psychological Association for copying or reprinting text of more than 500 words, tables, or figures. Permission is normally granted contingent upon like permission of the author, inclusion of the APA copyright notice on the first page of reproduced material, and payment of a fee of $10 per page, table, or figure. Abstracting is permitted with credit to the source. Permissions and fees are waived for authors who wish to reproduce their own material for personal use; fees only are waived for authors who wish to use their own material commercially. Permission and fees are waived for the photocopying of isolated articles for nonprofit classroom or library reserve use by instructors and educational institutions. Libraries are permitted to photocopy beyond the limits of U.S. copyright law: (1) those post-1977 articles with a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through the Copyright Clearance Center, P.O. Box 765, Schenectady, NY 12301; (2) pre-1978 articles, provided that the per-copy fee stated in the Publishers' Fee List is paid through the Copyright Clearance Center, P.O. Box 765, Schenectady, NY 12301. Address requests for reprint permission to the Permissions Office, APA, 1200 Seventeenth Street, N.W., Washington, D.C. 20036.

Subscriptions: Subscriptions are available on a calendar year basis only (January through December). Nonmember rates for 1979: $60 domestic, $64 foreign, $6 single issue. APA member rate: $20. Write to Subscription Section, APA.

Change of Address: Send change of address notice and a recent mailing label to the attention of the Subscription Section, APA, 30 days prior to the actual change of address. APA will not replace undelivered copies resulting from address changes; journals will be forwarded only if subscribers notify the local post office in writing that they will guarantee second-class forwarding postage.

Reprints: Authors may order reprints of their articles from the printer when they receive proofs.

Back Issues and Back Volumes: For information regarding back issues or back volumes write to Order Department, APA.

Microform Editions: For information regarding microform editions write to any of the following: Johnson Associates, Inc., P.O. Box 1017, Greenwich, Connecticut 06830; University Microfilms, Ann Arbor, Michigan 48106; or Princeton Microfilms, Princeton, New Jersey 08540.

Published monthly in one volume per year by the American Psychological Association, Inc., 1400 North Uhle Street, Arlington, Virginia 22201. Printed in the U.S.A. Second-class postage paid at Arlington, Va., and at additional mailing offices.

Diary No. 133
Date 3. 3. 1980
File No. Librarian
Bureau Edn Pay Research

JOURNAL OF
Personality
AND
Social Psychology

Volume 37 September 1979 Number 9

On Problems With the Cause–Reason Distinction in Attribution Theory

John H. Harvey and Jalie A. Tucker
Vanderbilt University

Buss's statement on causal and reason explanations of behavior involves a contribution regarding a necessary distinction in attribution theory. In this paper, we argue that despite his ambitious reformulation of previous analyses of lay attribution, Buss fails to provide substantive guidance for empirically testing the major assumptions of his hypothesized cause–reason distinction. Most important, Buss's analysis does not yield unambiguous theoretical criteria for operationally distinguishing intentional actions and unintentional occurrences and for judging the content of actors' and observers' explanations as causal or reason in nature. Without such specifications, we conclude that Buss' central ideas may be empirically untestable.

> We are now in a position to . . . steer a truer course that will lead us out of the conceptual fog with regard to causes and reasons in the attribution process. (Buss, 1978, pp. 1319–1320)

When a theorist makes such an assertion, however metaphorical, about integral concepts in a domain of work, that theorist's ideas invite and deserve careful scrutiny. In this paper, we seek to provide such scrutiny of the major aspects of Buss's (1978) reconceptualization of previous attributional analyses of lay explanation of behavior.

We are indebted to Gary Wells for comments on an earlier version of this paper.

Requests for reprints should be sent to John H. Harvey, Department of Psychology, Vanderbilt University, Nashville, Tennessee 37240.

To summarize Buss's position, he argued that attribution theorists (e.g., Heider, 1958; Jones & Nisbett, 1972) have failed to distinguish both conceptually and empirically between two logically distinct types of lay explanation (i.e., causal and reason explanations) and that this lack of distinction has greatly impeded an accurate understanding of this major area of social psychological inquiry. As an alternative to previous attributional analyses of lay explanation (especially the Jones & Nisbett, 1972, actor–observer hypothesis), which he considered exclusively causal in nature, Buss proposed (a) that an actor's behavior may be dichotomized according to whether it was intentionally or unintentionally initiated (termed an "action" and an "occurrence," respectively) and (b) that actors and

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3709-1441$00.75

observers differ in the types of explanations they may give for actions and occurrences. More specifically, Buss hypothesized that occurrences were explained exclusively with causes by both actors and observers and that actions were explained solely with reasons by actors, whereas observers could give reason or causal explanations of actions.

We readily acknowledge the potentially important contribution of Buss's statement in making this conceptual distinction between causal and reason explanations of actions and occurrences. However, although Buss is correct in concluding that this distinction has not been *formally* addressed in relevant attribution writings, his assertion that theorists have uniformly neglected reason explanations in favor of a causal analysis represents an overly generalized, polemical criticism. For example, see Heider (1958, p. 100) and Jones and Davis (1965, pp. 222, 263) for prior discussions of the role of reason explanations and actor intentionality in lay attributions. More importantly, we find Buss's statement most problematic in terms of its cogency and clarity as a reformulation of actor–observer phenomena. It will be argued below that while Buss's position may eventually contribute to a better understanding of lay explanations of behavior, ambiguities and oversights inherent in his analysis, especially in relation to the empirical testability of his ideas, greatly reduce the potential importance of his conceptualization.

Issues Involved in an Empirical Separation of Actions and Occurrences

Fundamental to Buss's conceptual distinction between causal and reason explanations is an assumption that an actor's behavior can be unambiguously dichotomized as either intentional action or unintentional occurrence. Thus, an adequate empirical test of Buss's hypothesized cause–reason attributional distinction rests on the operational separation of actions and occurrences, which he perceives as having generally been confounded in earlier research on actor–observer differences. As detailed below, we maintain that Buss's statement does not provide either a clear logical or empirical basis for operationalizing actions and occurrences.

In an attempt to clarify the nature of this error in previous research, and to illustrate his definitions of actions as consciously controlled intentional behaviors and occurrences as unwilled behaviors that the actor "suffers," Buss cited a number of examples of actions and occurrences from relevant experimental work. Examples of actions were intelligence test performance, expression of attitudes and opinions, translation of sentences, decision behavior, and carrying parcels (p. 1318); examples of occurrences were headache, blushing, perspiring, emotional arousal, galvanic skin response, friendliness, talkativeness, nervousness, dominance, and looking happy (pp. 1317–1318). Unfortunately, for an investigator seeking operational definitions of actions and occurrences, the criteria implied by these examples for distinguishing the two types of behavior present substantial difficulties. First, Buss's examples imply that physiological responses and overt displays of emotion are occurrences that operate completely outside the actor's conscious control, while actions encompass all deliberate behaviors, including those dependent on innate abilities such as intelligence. On both logical and empirical grounds, one must seriously question the implicit assumptions (a) that physiological responses and overt manifestations of emotionality cannot be at least partially controlled consciously (cf. Miller, 1969) and (b) that behaviors dependent on innate abilities can be completely controlled and deliberately manipulated (cf. Kruglanski, 1975).

Similarly, Buss's classification of the expression of attitudes and opinions as intentional actions is questionable on the grounds that attitude expressions are often affected by various experimental manipulations and/or self-perception processes (Bem, 1972) that do not appear to be readily controllable by the actor; thus in these instances it may be argued that the expression of an attitude represents an occurrence, not an action, in Buss's scheme. When one further considers that Buss classified expression of opinions as an action but regarded "looking happy" as an occurrence, one is left with the curious conclusion that if we say we like someone (an opinion), it is an intentional action; but if we smile in the

resence of this same person ("look happy"), is an unwilled occurrence. Although these ssues raised as a result of Buss's examples are robably not fundamentally damaging to his asic action–occurrence distinction, they evertheless demonstrate the difficulty in lassifying behaviors according to the actor's apacity to control various behaviors inten- ionally and consciously.

A second, highly related issue in determin- ng whether an actor's behavior should be lassified as an action or an occurrence con- erns specification of criteria for imputing ntention to an actor's behavior in ambiguous ituations where the behavior in question may r may not have been controlled by the actor (e.g., as in the case of overt displays of emo- ion). Despite Buss's contention that inten- ionality is a crucial differentiator of actions and occurrences, he did not address potential problems in imputing intentionality, thereby implying that it is usually possible to establish clearly whether or not an actor's behavior is intentional. One can envision ambiguous circumstances, however, where the actor, observer, and investigator fail to agree on the presence or absence of the actor's intentionality and/or conscious control over a given behavioral sequence. To use Buss's example (pp. 1317–1318) of such a situation, the experiments by Storms (1973), Regan and Totten (1975), and Taylor and Fiske (1975) involved actors and observers giving explanations of the actor's degree of friendliness, talkativeness, nervousness, and dominance in a "getting acquainted" situation. On the one hand, Buss classified these behavioral dimensions as occurrences that should evoke causal explanations from both actors and observers. Buss also allowed, however, that if the actor consciously controls his/her self-presentation along these dimensions, the actor would provide a reason, rather than a causal, explanation of behavior. Alternatively, the actor might adamantly deny any such willful manipulations, whereas the observer and investigator might perceive the actor as consciously and intentionally controlling his/her behavior.

In light of potential disagreements between the actor, observer, and investigator concerning the presence or absence of intention on the actor's part, the investigator would seemingly be forced either (a) to rely on the actor's self-report of his/her behavioral intentionality or lack thereof or (b) to judge the content of the actor's explanation to determine whether it comprises a reason or causal explanation of an intended action or unintended occurrence, respectively. The first possibility depends on the somewhat questionable validity of the actor's self-report, especially in relation to willingness and ability to admit intention, whereas the second depends on the existence of independent criteria that can be used to classify the content of explanations as either primarily causal or primarily reason in nature. As will be discussed in the following section, Buss's analysis does not readily yield such criteria.

Overall, we must conclude that Buss's analysis is weak in clarity of definitions or empirical implications for concepts such as intention and conscious control of behavior, which are central to his action–occurrence distinction. Presumably, these concepts refer to qualities of actions, not occurrences, but is it necessary for both to exist for an action to be perceived? For instance, would it not be possible for a behavior to be consciously and deliberately taken, but not intentionally initiated? Considerations such as these are totally lacking in Buss's analysis and require further explication if his statement is to be of much value to researchers.

Empirical Implications of the Cause–Reason Distinction

Buss's second major assumption, that actors and observers differ in the types of explanations they give for actions and occurrences, also entails major difficulties in its empirical implications. In line with his criticism that existing actor–observer research has typically confounded action and occurrence behaviors, Buss seems to suggest that actor–observer differences should be investigated by having actors and observers give explanations of behaviors that can unambiguously be classified as actions and occurrences. One thus can formulate a basic experimental design to test the validity of Buss's cause–reason distinction;

this design would involve comparisons of the explanations offered by actors and observers of both actions and occurrences.

Under these experimental conditions, Buss's model predicts that no differences should be expected in the content of actors' and observers' (causal) explanations of occurrences, but that differences may be expected in their explanations of actions. That is, actors may give only reason explanations for actions and observers may give reason, causal, or reason plus causal (mixed) explanations. Buss further contended that the types of explanations elicited from observers to explain actions may be manipulated by instructing the observer (a) to explain the action using reasons as he/she perceives the actor would, (b) to give a reason explanation based on the observer's personal interpretation of the action, or (c) to give a strictly causal explanation (p. 1318). Conceivably, actors could also be instructed to give explanations from the perspectives of both the self and an observer. Buss did not consider this latter possibility, however, apparently because his model precludes the possibility that actors might give causal explanations of actions by adopting the perspective of an observer. Only when both actors and observers give explanations of the same logical form (i.e., when actors and observers explain occurrences with causes and actions with reasons) did Buss consider the resulting explanations as directly comparable. Despite this assertion about the logical comparability of explanations under these conditions, Buss also maintained the seemingly inconsistent position that actors and observers "are engaged in *fundamentally* (i.e., logically) *different* situations to explain the same action" (p. 1316). Overall, it would seem that an adequate empirical test of Buss's cause–reason distinction would involve having both actors and observers explain actions and occurrences from the perspective of both an actor and an observer.

Although this experimental design seems to assess adequately the validity of Buss's cause–reason distinction, his statement is obscure regarding meaningful a priori criteria for distinguishing and classifying the content of the explanations that would be produced. For instance, Buss merely states that causal and reason explanations may be distinguished by the greater degree of justification "vis-à-vis society's norms for 'proper' conduct" (p. 1315) in reason than causal explanations, and by the greater degree of "such 'connotative attributes' as lawfulness, determinism, antecedent–consequent relationship, predictability and replicability" (p. 1314) in causal than reason explanations. He fails, however, to elaborate on how these rather vague distinctions might be measured in explanations derived in an experimental situation. This difficulty in empirically classifying explanations as causal or reason is compounded further by Buss's contention that actors give reason explanations of actions that are couched in causal terminology (pp. 1315, 1318), and that observers may make justifications for the actor (p. 1318).

It may prove possible, of course, to establish empirically those situations in which content differences in actors' and observers' explanations of actions are either maximized or minimized, in a manner similar to findings of divergence and convergence in actor–observer explanations conducted within the actor–observer research paradigm (e.g., Wells, Petty, Harkins, Kagehiro, & Harvey, 1977). Notwithstanding this possibility, however, Buss's failure to articulate well-developed a priori criteria for establishing the causal or reason nature of explanations must be considered a fundamental shortcoming of his theoretical conceptualization. Without better a priori criteria, one is generally left in the scientifically indefensible position of defining, in a circular manner, reason explanations as those given by actors for actions and causal explanations as those given by actors and observers for occurrences.

A final issue, which indirectly relates to problems in investigating Buss's conceptualization, concerns his implicit characterization of the information sources and perceptive capacities of actors and observers. Specifically, in comparison to Jones and Nisbett's hypothesis, Buss implicitly appears to accord the observer more and the actor fewer information sources and perceptive powers for deriving attributions. Whereas Jones and Nisbett em-

dasized that the actor often possesses more
levant information regarding his/her be-
avior than the observer, who often merely
impses short segments of the actor's ongoing
ehavior, Buss instead presented the observer
possessing the capacity to perceive motives
nd forces influencing the actor's behavior
at operate outside the actor's awareness (p.
316). The observer and also the investigator,
ho qualifies as an observer, may then use
is information to give a causal explanation
f the actor's behavior. In contrast, Buss de-
icted the actor's perceptual and attributional
pacities as substantially limited by his/her
eed to provide society with a justification for
ersonal behavior.[1]

Hence, in Buss's conception, the observer is
ssentially characterized as capable of omni-
iently perceiving unconscious forces influ-
ncing the actor's behavior, which form the
asis of the observer's truly causal explanation
f the actor's behavior. This implied view is
rprisingly consistent with the disinterested-
cientist efficient-causality model that Buss
eems to reject. Buss's implicit characteriza-
ion of the perceptual and attributional capac-
ies of actors and observers does not neces-
arily invalidate the basic cause–reason dis-
nction advanced in his paper. Nonetheless,
s questionable foundation, which involves
ttle consideration of extant evidence and
nalysis regarding actor–observer inference
rocesses, does lessen the intuitive appeal of
uss's conceptualization.

Conclusions

The foregoing discussion highlights some of
he issues and problems involved in attempt-
ng to incorporate Buss's ideas in empirical
nvestigations of lay explanations of behavior.
We acknowledge that his criticism regarding
he previous failure of attribution investiga-
ors to distinguish actions and occurrences em-
pirically is well taken. However, we must
onclude that Buss's analysis offers little sub-
tantive guidance in establishing theoretical
criteria for distinguishing between these two
types of behavior or in testing his hypothesis
regarding logical differences in the types of ex-
planations given by actors and observers.
Without better criteria for distinguishing actions and occurrences and for judging the content of explanations as causal or reason in nature, the impact of Buss's contribution approaches that of an untestable armchair analysis.

---

[1] Although Buss may reasonably take issue with Jones and Nisbett's assumptions about possible mediators of actor–observer differences, he fails to elaborate the psychological processes that may mediate actors' presumed preoccupation with justifying, reason explanations. If, as Buss implies, motivations associated with self-esteem maintenance or enhancement, or with self-presentation concerns, underlie actors' reason attributions, we view Buss's analysis as inadequate in its failure to address several works relevant to the role of self-serving bias mediators of attributions, such as Miller and Ross's (1975) and Bradley's (1978) influential discussions of information processing alternatives to a self-esteem position or Monson and Snyder's (1977) important reconceptualization of the psychological processes involved in actor–observer attributional phenomena.

References

Bem, D. J. Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6). New York: Academic Press, 1972.

Bradley, G. Self-serving biases in the attribution process: A reexamination of the fact or fiction question. *Journal of Personality and Social Psychology*, 1978, *36*, 56–71.

Buss, A. R. Causes and reasons in attribution theory: A conceptual critique. *Journal of Personality and Social Psychology*, 1978, *36*, 1311–1321.

Heider, F. *The psychology of interpersonal relations*. New York: Wiley, 1958.

Jones, E. E., & Davis, K. E. From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2). New York: Academic Press, 1965.

Jones, E. E., & Nisbett, R. E. The actor and the observer: Divergent perceptions of the causes of behavior. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior*. Morristown, N.J.: General Learning Press, 1972.

Kruglanski, A. W. The endogenous–exogenous partition in attribution theory. *Psychological Review*, 1975, *82*, 387–406.

Miller, D. T., & Ross, M. Self-serving biases in the attribution of causality: Fact or fiction? *Psychological Bulletin*, 1975, *82*, 213–225.

Miller, N. E. Learning of visceral and glandular responses. *Science*, 1969, *163*, 434–435.

Monson, T. C., & Snyder, M. Actors, observers, and the attribution process: Toward a reconceptualization. *Journal of Experimental Social Psychology,* 1977, *13,* 89–111.

Regan, D. T., & Totten, J. Empathy and attribution: Turning observers into actors. *Journal of Personality and Social Psychology,* 1975, *32,* 850–856.

Storms, M. D. Videotape and the attribution process: Reversing actors' and observers' points of view. *Journal of Personality and Social Psychology,* 1973, *27,* 165–175.

Taylor, S. E., & Fiske, S. T. Point of view and p ceptions of causality. *Journal of Personality Social Psychology,* 1975, *32,* 439–445.

Wells, G. L., Petty, R. E., Harkins, S. G., Kagehi D., & Harvey, J. H. Anticipated discussion of int pretation eliminates actor–observer differences the attribution of causality. *Sociometry,* 1977, 247–253.

Received September 14, 1978

New JPSP Format Delayed Until April

Recent issues of this journal carried an announcement that the *Journal of Personality and Social Psychology* would appear in January 1980 under new policies and in a new sectioned format. However, in order to publish the backlog of manuscripts accepted under present policies, the first three issues of 1980 will continue to carry articles accepted under Clyde Hendrick's Acting Editorship.

This journal will appear in April as a sectioned journal edited by Melvin Manis, Ivan D. Steiner, and Robert Hogan, as previously announced.

Robert S. Daniel, Chair, APA
Publications and Communications Board

Journal of Personality and Social Psychology
1979, Vol. 37, No. 9, 1447–1457

Causal Explanation, Teleological Explanation: On Radical Particularism in Attribution Theory

Arie W. Kruglanski
Tel-Aviv University, Tel-Aviv, Israel

The cause–reason distinction in attribution theory is examined. The author notes conceptual difficulties associated with previous interpretations of cause as opposed to reason and offers a reformulation that shows how *reason* may refer to a specific type of explanation (notably, teleological) and *cause* may refer (a) to the general case of explanation (the inclusive sense) *or* (b) to the general case of nonteleological explanation (the exclusive sense). From this perspective he considers Buss's recent comments regarding teleological and nonteleological (causal) modes of naive explanation. It is noted that teleological (reason-type) explanation does indeed fall within the purview of attribution theory but is merely one among a vast number of possible explanatory types that all properly belong within the attributional domain. Hence, the reason concept is of no unique significance for the theory of attribution. The foregoing argument culminates in the conclusion that a theory of naive epistemology might well rid itself of concern with lay concepts (such as cause and reason) that have figured frequently in attributional formulations. Such concepts may be thought of as the *contents* of knowledge, of which there may be an infinite variety. Instead, an epistemic theory should restrict itself to the *process* of knowledge acquisition assumed to be invariant across the diverse epistemic contents of possible interest to the layperson.

My object in the following pages is to consider the cause–reason distinction in attribution theory and to comment on Buss's (1978) recent discussion of this topic. Despite my disagreement with some of its contents, I find Buss's article of considerable interest, particularly as it calls attention to the relevance of British philosophy of the ordinary language (e.g., Anscombe, 1957; Austin, 1962; or Ryle, 1949) for contemporary research on attribution (Harvey, Ickes, & Kidd, 1976, 1978; Jones, Kanause, Kelley, Nisbett, Valins, & Weiner, 1972; Kelley, 1967).

The main task that the ordinary-language philosophy set for itself was the explication of concepts in everyday discourse. Buss (1978) takes two such concepts, cause and reason, and proclaims them of central importance to the theory of attribution. By contrast, I have felt for some time (cf. Kruglanski, 1977; Kruglanski, Hamel, Maides, & Schwartz, 1978; Kruglanski & Bar-Tal, Note 1) that the preoccupation with lay concepts threatens to ensnare attribution theory in endless particularities that might ultimately divest it of scientific value. In the remainder of this article I present my arguments for this position, with special reference to the cause–reason issue.

The present essay is composed of four major parts. In the first, some difficulties are noted with common conceptions of cause and reason, and a reinterpretation is offered clear of those difficulties. In the second part, teleological and nonteleological modes of naive explanation are examined and are contrasted with Buss's views on these matters. The third part deals with Buss's claim that conceptual confusion in attribution research may have resulted from a failure to draw the distinction between cause and reason. In the fourth part the cause–reason issue is cast into broader per-

Requests for reprints should be sent to Arie W. Kruglanski, Department of Psychology, Tel-Aviv University, Tel-Aviv, Israel.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3709-1447$00.75

spective and an evaluation is made of the place that any lay concepts may be hoped to assume within the theory of attribution.

The Cause–Reason Distinction Reconceptualized

Some Problems With Current Conceptions of Cause Versus Reason

Let us see how Buss characterizes *cause*, *reason*, and the differences that set them apart. As he expressed it (Buss, 1978, p. 1311), "causes are that which brings about a change," whereas "reasons are that for which a change is brought about (e.g., goals, purposes, etc.)." In Aristotle's classification, Buss's *cause* is *efficient cause*, *reason* is *final cause*, and they are both presumed to differ from *material cause*, or "that in which the change comes about," and *formal cause*, or "the pattern or shape of that which is changed" (p. 1312). We are further informed that causal explanation is associated with "such connotative 'attributes' as lawfulness, determinism, antecedent–consequent relationship, predictability and replicability" (p. 1314).

But the foregoing clarifications notwithstanding, the distinction between *cause* and *reason* still seems possessed of a "will-o'-the-wisp" quality. For instance, characterization of *cause* as "that which brings about change" would appear equally applicable to reason. Surely an actor has a reason for an action before it occurs (a person may feel thirsty and then proceed to drink), and action would not have occurred without the reason, so the reason may well be said to bring about the change (i.e., the action). This conclusion is hardly altered if *reason* is conceptualized as the end state that the actor strives to attain (e.g., the quenching of one's thirst). Anticipation of the end state must antedate the action and, in a sense, bring it about. Needless to say, the actual end state (vs. mere anticipation thereof) does not bring about the action, and many a behavior may run its course without eventual attainment of the underlying purpose.

The above considerations suggest that reason may bring about action (change) and that the antecedent–consequent relation that Buss (p. 1314) reserved for causal explanation is shared by teleological explanation. Furthermore, it appears that lawfulness, predictability, replicability, and even a certain sense of determinism may well be predicated on teleological explanations, so they may not differentiate it from causal explanation. To explain a person's action in terms of a reason means that, all else being constant, given the reason, the action would follow. This of course means that reason and action are *lawfully* (hence *replicably*) related and that action may be *predicted* from reason. As has just been noted, actual success in predicting an action from a reason requires the *constancy* of all remaining conditions including possible alternative goals and beliefs that the actor might have as to the best means for accomplishing his/her particular end. That such psychological constancy across occasions might sometimes be unlikely *in practice* does not in the least detract from the *principle* of predictability and replicability intended by teleological explanation.

To illustrate the essential lawfulness conveyed by a teleological explanation, consider a brigand accosting a victim with a "money or life" dilemma. Most people would agree that the brigand is attempting to elicit in the victim the "life-saving" end as well as the belief that yielding the money would be the surest way of reaching this end. Note that the brigand's expectation of being obeyed (or this person's prediction that he/she would be) betrays the assumption that purposes systematically lead to actions (that may best serve the purposes involved).

Furthermore, should the brigand's prediction fail to materialize (should the victim refuse to comply), it seems unlikely that this would be explained by an unlawfulness of the purpose–action relation. Rather, the actor's unexpected behavior would probably be taken to reflect a shift in circumstances, that is, a violation of the constancy assumption, that may render the situation irrelevant to the original purpose–action hypothesis. For example, one might postulate that a different purpose dominated the victim's conduct (a

sudden penchant for heroics) or that the actor deemed a different action to serve the original purpose best (having become a black belt in karate, the victim could feel that a side kick to the assailant's chin might best rid him/her of the threat involved).

Thus, contrary to implying that the relation of purpose to action is helter-skelter, social interaction seems fundamentally to depend on a belief that the latter may be predicted from the former, that is, to depend on the essential lawfulness intended by teleological explanation (for a similar view see, e.g., Schlenker, 1974). The preceding discussion makes plain that the demarcation of teleological explanation from causal explanation is a largely unaccomplished task.

An Asymmetrical View of Cause Versus Reason

I believe that the cause–reason question may be neatly resolved (or dissolved) when reformulated *asymmetrically* rather than *symmetrically*, as it has been formulated by several contemporary commentators (e.g., Buss, 1978). By a symmetrical interpretation of cause versus reason I mean their treatment as equally definitive concepts of unique significance for the problem of explanation. However, while most people seem to have little difficulty with the concept of reason (usually defined in such familiar terms as purpose or goal), explication of the notion of cause has required considerable analytic efforts. For example, Buss approvingly cites the view that "causal explanation of human action would be in terms of bodily movements" (Buss, 1978, p. 1313). In addition, Buss equates causality with "that which makes things happen, or brings about change" (p. 1312). I admit to finding such clarifications not entirely satisfactory and to being left with a gnawing sense of ignorance regarding the intended meaning of cause as a concept.

According to the present asymmetrical view, reason is interpreted as a specific type of explanation (notably, teleological), whereas cause is interpreted as explanation[1] in the generic sense. More precisely, the cause notion has apparently been employed in two broad senses: one sense (henceforth labeled *inclusive*) has denoted any explanation, including the teleological. The second sense (henceforth *exclusive* has denoted any explanation excluding the teleological. The inclusive sense seems to underlie Aristotle's classification of causes into efficient, material, formal, and final, for example. This line of reasoning seems now to reflect simply four types of explanation, of which teleological explanation (based on final causes) may be one. Similarly, in attribution theory, frequent references to causal explanation connote (somewhat redundantly) the theory's concern with naive explanation of whichever type rather than with special kinds of naive explanation. Finally, the pervasive use of the because preposition with all types of explanation is readily comprehensible if we assume that cause simply means explanation, no more, no less.

The exclusive interpretation of cause (as any nonteleological explanation) seems prominent in those treatments where it is explicitly opposed to reason, as in Buss's (1978) recent analysis. In those instances, cause seems defined negatively as a "nonreason" but is not further characterized more positively. In the remainder of this article, too, cause will be contrasted with reason and will exclusively denote any nonreason-type explanatory category. Regardless of whether the inclusive or the exclusive view of cause is adopted, the assymmetrical interpretation of the cause–reason issue has this fundamental implication: Reason is merely one type of explanation, whereas cause may include an infinite variety of explanatory types. Thus Aristotle's fourfold classification of possible causes is considered a wholly arbitrary taxonomy. Besides efficient, material, formal, and final explanations, one can have mathematical, mechanical, geographic, demographic, or social explanations,

---

[1] Consistent with common usage, *explanation* is used here to mean a statement constructed around the connective *because* (as in "Bill stayed home because it rained"). It is generally accepted that such statements presume a universal conditional (if–then) proposition ("if it rains Bill stays home") and an instantiating clause affirming the antecedent of the above conditional ("it did rain"; see, e.g., Popper, 1961, p. 19; Turner, 1965, p. 272).

and so on. Similarly, cause and reason may not now be regarded as *two* equivalent explanatory types of special significance. Rather, reason is seen to represent a particular type of explanation, whereas cause is seen to represent explanation in general or nonreason-type explanation in general. Thus, we seem to have an infinity of explanatory types of which teleological explanation is but one. In the following section this type of explanation is considered in further detail.

Teleological and Nonteleological Modes of Naive Explanation

Two major questions with which Buss (1978) deals at length are (a) whether nonteleological (in his terms, *causal*) explanations of actions and/or teleological explanations of occurrences may be possible and (b) *who* (actors, observers, or both?) may be capable of nonteleological versus teleological accounts of actions or occurrences. In his answers Buss implies that nonteleological accounts of actions are in principle possible, although they may be given by observers only, not by actors: "*action* . . . is explained by the actor with reasons. The observer may use causes and/or reasons in explaining action" (p. 1311). By contrast, occurrences (i.e., events devoid of intentionality) are assumed to be explicable only nonteleologically: "*occurrence* . . . is explained by both actors and observers with causes" (p. 1311). In the forthcoming section I closely examine Issues a and b above and comment on Buss's proposed answers.

Are Nonteleological Accounts of Actions and/or Teleological Accounts of Occurrences Possible?

The answers to these questions depend on how one chooses to define teleological and nonteleological explanations. If by nonteleological explanation is meant explanation that denies the behavior's purposive character, a nonteleological explanation of actions is impossible by definition. I suspect, however, that for most people a teleological explanation is merely one that identifies the actor's reason for a behavior. For example, one might (teleologically) account for Mark's application to Harvard Law School by mentioning his hope of joining the family law firm, thus explicitly identifying the behavior's underlying reason. Similarly, one might (teleologically) explain Sara's departure for a skiing trip by labeling it an end in itself, which again implicates the action's instigating reason (Sara's desire to ski; see, e.g., Kruglanski, 1975, Derivation 9, p. 302).

Granting the above definition of teleological explanation, a nonteleological explanation would be one that is mute with respect to the actor's specific reason for a behavior. Further, because actions alone may have underlying reasons (purposes, etc.), if a teleological explanation is given, it follows logically that the event explained was an action. The obverse does not follow, however: Given that an action was explained, the explanation must not necessarily have been teleological. For example, one might nonteleologically account for John's application to Harvard in terms of his upbringing, which does not clarify John's specific reason for the action. Similarly, one might explain Paul's decision to undertake some job in terms of the situation (most people in it would have done the same) or in terms of Paul's personality (only someone like him would have done so), again without illuminating Paul's specific reason for the decision. Assuming the interpretation of cause as nonreason (the exclusive interpretation given earlier), Buss seems correct in implying that nonteleological accounts of actions are in principle possible.[2]

Finally, if the essence of teleological explanation consists of identifying the actor's reason, such explanation may not logically be possible with reference to nonintentional behaviors. One cannot identify something that is not there, that is, a purpose in a nonintentional behavior. Assuming the exclusive interpretation of the cause concept, Buss's asser-

[2] This is contrary to Postulate 1 of my endogenous–exogenous article (Kruglanski, 1975, p. 389), which implies that nonteleological explanation of actions may not be possible. Assuming the present definition of teleological and nonteleological explanation, the above implication of Postulate 1 must now retracted.

tion that occurrences must be explained causally by all actors as well as observers simply means that they may not be explained teleologically. This follows tautologically from the definition of occurrences (as nonintentional behaviors devoid of reasons) and must be considered quite indisputable. Let us consider now an issue on which Buss's claims may not seem as cogent: the hypothesis of actor–observer differences in the rendition of teleological and nonteleological explanations.

Are Observers Alone Capable of Nonteleological Explanations for Actions?

Buss provides two separate arguments for the contention that observers alone, and not actors, can explain action nonteleologically. According to one argument,

> The actor's "because" statements consist of giving his or her reasons for an action, that is, matters that weighed in on his or her deliberation. By definition, in giving one's reasons for an action, only matters that are *consciously* available to the actor can occur in such explanations. In contrast, the observer who is asked to explain "why" in regard to an actor's action can include causal attributions that operate *out* of consciousness of the actor. Thus, while an observer can make causal attributions in explaining the action of an actor . . . , the actor's explanation could not include such kinds of attributions. . . . (Buss, 1978, pp. 1315–1316)

According to the second argument: "In asking an actor to explain his or her action, the 'why' is a request to justify, or to make rational or intelligible, his or her action vis-à-vis society's norms for 'proper' conduct . . ." (p. 1315). I now examine the two foregoing arguments in turn.

The differential-consciousness hypothesis. Turning to the former argument, it may be readily conceded that only matters available in one's consciousness can figure in one's explanations and that the contents of actors' consciousness may systematically differ from those of observers' consciousness. Such differences have usually been interpreted as an informational advantage of actor over observer, as the former person more than the latter may be cognizant of his or her own bodily reactions or of the details of his or her own personal history. But it is difficult to see how an excess of information may prevent the actor from formulating explanations of which an observer may be capable.

More generally, granting that a nonteleological explanation is merely one that stays incommunicative with respect to the actor's reason for a behavior, it still remains to specify *what* in the actor's consciousness might prevent this person from arriving at explanations in this class. In fact nonteleological explanations of a person's own actions seem quite commonplace; criminal activity is often explained with reference to the adverse social conditions during the criminal's childhood, and a hostile act may be accounted for in terms of the perpetrator's tension or irritability.

The justification hypothesis. Buss's second argument for the actor's presumptive inability to formulate nonteleological accounts of his or her own behavior assumes that (a) the actor's accounts of personal actions are necessarily justificatory (Buss, 1978, p. 1315) and (b) all justificatory accounts of actions are necessarily teleological. On close inspection, both of the foregoing premises appear questionable. The first appears unwarranted if *justificatory* is taken in the sense of normatively excusable (this seems implied in Buss's reference on p. 1315 to society's norms and rules for "proper" social behavior). Surely an actor may occasionally admit to socially unwholesome motives (greed, cowardice, vengeance) or to reasons that are normatively neutral, as when one explains a trip to the grocery store by reference to the empty cupboard at home.

If by justificatory is meant, more broadly, intelligible or plausible accounts (excusable or not; cf. Buss, p. 1315) then Premise b would still seem doubtful. Many intelligible accounts of one's own behavior may be nonteleological, as in the earlier example of the criminal who may account for deviant actions in terms of the deleterious social circumstances in his or her neighborhood. Finally, if by *justificatory* is meant *rational* (Buss, p. 1315) and by *rational* is meant *teleological*, then accounting for the actor's presumptive need to give teleological explanations by invoking the need to give justificatory explanations would be circular, hence devoid of

clarifying value. The preceding discussion shows that Buss's hypotheses of actor–observer differences, relying on teleological (vs. nonteleological) explanations, are largely unsupported.

The Cause–Reason Distinction in Attribution Research

According to Buss, the failure of attribution theorists to draw the distinction between causes and reasons has resulted in much conceptual confusion. This is illustrated in reference to two research topics: (a) Jones and Nisbett's (1971) hypothesis of actor–observer differences in attribution and (b) Kruglanski's (1975) partition between endogenous and exogenous attributions. Let us examine Buss's specific comments on these issues.

Jones and Nisbett's (1971) Actor–Observer Hypothesis

In a classic attributional essay, Jones and Nisbett (1971) speculated, "actors attribute cause to situations while observers attribute cause to dispositions . . ." (p. 82). Buss (1978, p. 1316) regarded the above formulation as objectionable, for it contradicts his aforementioned assumption that actors (unlike observers) are fundamentally incapable of causal (i.e., nonteleological) accounts of their own behaviors. Furthermore, Buss viewed as categorically mistaken statements (occasionally made by researchers in the actor–observer paradigm) conjoining the *actor* term with nonintentional behaviors like nervousness and physiological reactions to music or emotional responses.

But from the present perspective the problems identified by Buss are not as serious as has been alleged. The statement "actor attributes cause for own behavior" may be taken in the inclusive sense of cause to mean simply that the actor explains his or her own behavior (indeed, I strongly suspect this to have been Jones and Nisbett's intent in the first place). But even if cause is interpreted exclusively as a nonteleological explanation, it follows from the preceding section that actors should be capable of attributing cause (for their own behavior) if observers are capable of such attributions.

As for the linkage of the *actor* term with nonintentional behaviors (occurrences), the contradiction identified by Buss exists only if actor is narrowly interpreted as producer of intentional behaviors. If actor is taken in the broader (if somewhat Pickwickian) sense of target person whose responses (voluntary or not) are being explained (an interpretation that I again believe better captures Jones and Nisbett's intent), however, the apparent contradiction exists no longer. Generally, the hypothesis of Jones and Nisbett (1971) implicates the differential preference of behaving individuals (actors, target persons) versus the observers of such persons, for situational versus dispositional explanations of the behaviors at issue. In present terms, both situational and dispositional explanations are nonteleological: They do not name specific reasons, if such reasons are indeed involved in the behaviors observed. As we have seen, actions as well as occurrences may be explained nonteleologically, and actors (target persons) as well as observers seem capable of nonteleological explanations. Hence, contrary to Buss's reasoning, Jones and Nisbett's hypothesis seems conceptually coherent, and its case would have to rest ultimately with empirical research.

Finally, to the extent that neither the person nor the situation categories are classifications of reasons, the cause–reason distinction is not particularly relevant to the paradigm of Jones and Nisbett. The distinction's relevance hinges entirely upon relevance of the reason concept to a given domain of phenomena, the residual concept of cause signifying merely a nonreason, that is, containing no unique meaning independent of the reason notion.

The Endogenous–Exogenous Partition

Buss's second example of confusion due to failure to differentiate between cause and reason recalls the theory of endogenous (vs. exogenous) attributions (Kruglanski, 1975) concerned with the case in which action is assumed to be an end in itself (endogenous

attribution) rather than a means to a further end (exogenous attribution). As Buss (1978, p. 1318) correctly points out, Postulate 1 of the theory implies a distinction between causal and teleological explanation in stating that "the lay explanation of actions is teleological rather than causal" (Kruglanski, 1975, p. 389). But subsequent portions of the same article contradict the distinction in speaking about "causally attributing actions . . . " and in implying the interchangeability of causes and reasons (p. 389).

The foregoing criticisms by Buss seem quite valuable: (a) They demonstrate a clear inconsistency in my usages of the causality notion throughout the 1975 article, and (b) they highlight an authentic shift of views between my 1975 work and my current reasoning. While Postulate 1 of the former article denies the possibility of explaining actions nonteleologically, such a possibility has been explicitly affirmed throughout the present essay.

Finally, note that the endogenous and exogenous categories refer to teleological explanations, so the notion of reason would seem quite pertinent: Implications drawn from an endogenous attribution (Kruglanski, 1975, pp. 391–392) follow without exception from its being a classification of reasons. In agreement with Buss (p. 1320), we may therefore conclude that the reason concept is of considerable significance to the endogenous–exogenous paradigm.

In sum, the concept of reason (and the derivative explanation type called *teleological*) could be of relevance to some domains of attributional research (like that of endogenous vs. exogenous attributions) while being irrelevant to alternative attributional domains (like that of actor–observer differences). What then can be safely concluded about the concept's significance for attribution theory at large? This issue is specifically addressed in the following section.

The Reason Notion and the Content Problem in Lay Epistemology

The main thrust of Buss's analysis has been for the inclusion of teleological (reason-type) explanations within the purview of attribution theory. I readily concur with the suggestion: The recognized province of attribution theory is lay explanation; a teleological account is one type of lay explanation, so it should properly fall within that province.

But it is one thing to concede the applicability of attribution theory to reason-type explanations and quite another to pronounce the reason–cause distinction as of key importance (see Buss, 1978, p. 1311) within the theory. Such centrality would certainly be implied in a scheme that regards the reason and cause categories as the unique classificatory scheme for explanations. But according to the present view there is nothing unique about classifying explanations into causes and reasons. As noted earlier, the concept of cause as distinct from reason seems best interpretable as a nonreason, that is, as representing an open class encompassing a potential infinity of explanation types. In other words, reason-type or teleological explanation is merely one of an infinite number of possible explanations. Furthermore, viewed now as it separates reason from nonreason, the reason–cause partition may be compared to the possible distinctions between (any of the infinite) explanatory categories and their complementary classes, for example, mechanical versus nonmechanical explanations, biological versus nonbiological explanations, and so forth. In sum, attribution theory deals with the *general* case of explanation, so any distinction (like that between reason and cause) based on a *specific* type of explanation (out of the infinity that are possible) must be of a vanishing significance with the theory.

But matters will now become even more extreme. We have already seen that the class of teleological statements (reason-type explanations) is a subset of the class of all explanatory statements and that the latter class defines the recognized purview of attribution theory. But obviously the class of explanatory statements (those constructed around the because preposition) is a subset of the class of all statements, and I shall now argue that it is the latter class that defines the domain of a general epistemological theory, of which attribution theory is merely a special case.

Within this general epistemic theory *any* classification of statements (as explanatory or as teleological) turns out to be of vanishing importance. Therefore, the attribution framework, constrained as it may be to explanatory statements, deals with an extremely small segment of the domain to which general epistemological laws are assumed to apply.

The epistemic theory presently referred to (and elaborated in Kruglanski et al., 1978 and in Kruglanski & Bar-Tal, Note 1) rests on the partition between the contents of knowledge (epistemic contents) and the process of knowledge acquisition (epistemic process). Epistemic contents are all propositions whose validity a person might, on occasion, wish to ascertain—in principle, all propositions. The epistemic process is the sequence of cognitive operations performed by an individual on the way to a given bit of knowledge. The terms *process* and *content* figured prominently in the recent debate about transhistorical laws in social psychology (see, e.g., Gergen, 1973, 1976; Hendrick, 1976; Manis, 1975, 1976; Schlenker, 1974, 1976; Thorngate, 1975), although their meaning in that context differed from the present one. To clarify matters, the distinction between the two usages is now considered at some length. As presently employed, process refers to systematic change and content to the object of such change. In this sense, the distinction between epistemic process and content parallels the distinction between the metabolic process and the biological system and between the evolutionary process and the evolving species, to name only two examples.

The foregoing interpretation must be sharply distinguished from that in which process denotes the abstract or the general level and content the concrete or the specific level. It is the latter meaning of the process–content partition that has characterized the social-psychology-as-history debate just mentioned. For example, in denying the feasibility of process invariance in social psychology, Thorngate (1975, p. 487) referred to "*local* organization" (meaning contents) and "*general* disorganization" (meaning process, italics mine). Gergen (1976, p. 376) similarly asserted that in speaking about "processes underlying social behavior we are dealing with a hypothetical structure suitable for encapsulating a series of selected observations . . . ." (i.e., a series of social contents). In the same vein, Manis (1976) proposed that "the development of a scientific theory . . . involves the construction of some hypothetical mechanism or process that lends coherence to an otherwise chaotic set of particulars" (p. 431). Thus, reinforcement is cited as a theoretical term representing process, and a specific classroom event, as a particular instance of reinforcement, that is, content.

It is stressed that the two foregoing interpretations of process versus content are quite orthogonal. Both change and the object of change may be conceptualized on a variety of abstraction levels, and neither may be exclusively identified with the abstract or the specific level. For example, one might particularistically describe a sequence of physiological changes transforming a given child into an adult at a specific time and place, or one might refer to the very same event sequence more abstractly as an instance of the maturation process. Similarly, the child (the object of change) might be particularistically identified as Johnny Smith or might more abstractly be referred to as child, mammal, or physical body.

The last example serves to underscore the relativity of abstraction levels; mammal is an abstract class of which human child is a particular instance, but mammal is itself a particular instance of the organism class, which in turn is a particular instance of the physical object category, and so forth. Interpreting process as the abstract and content as the specific, it is indeed correct to assert that "one man's process is another's content" (cf. Hendrick, 1976, p. 395). But the relativity hardly extends to the alternative interpretation of process and content mentioned earlier. To confuse change and the object of change would represent a gross mistaking of category (Ryle, 1949), of which few persons would be guilty.

While in the present usage the generic categories of process and content are not presumed to differ in level of abstraction, a particular content could still be the object of numerous particular processes, and a par-

ticular process could still apply invariably across multiple particular contents. Thus the (same) human body may be the object of multiple different processes: growth, maturation, aging, disease. Conversely, the process of evolution may invariably apply to all biological species, and the process of metabolism to all living organisms. In precisely this last sense it is presently assumed that the same epistemic process discussed below underlies the acquisition of all epistemic contents:[3] The sequence of cognitive operations that a knower may perform is assumed to be invariant regardless of the kind of knowledge that this knower may seek.

In light of the foregoing assumption it is now possible to appreciate the sense in which any classification of statements may be of a vanishing significance for the general epistemic theory: (a) Statements may be classed in a vast number of ways, based, for example, on their inclusion of any of the vast number of concepts that man may be competent to construct. Thus, the designation of statements as explanatory derives from their inclusion of the concept of the conditional (if-then) relation (see Footnote 1), but it is possible to devise numerous alternative classifications based on alternative included relations (e.g., the various geometric, social, or biological relations). (b) In the present framework any classification of statements is a classification of epistemic contents to which the same epistemic process is assumed to apply; for example, the process of explanation (the acquisition of so-called causal knowledge) would be no different from that involved in the acquisition of any knowledge, including nonexplanatory (noncausal) knowledge. It therefore follows that any classification of statements (any categorization of epistemic contents) may be of extremely limited import for the general theory of knowledge and that the theory might best focus instead on the epistemic process.

A detailed discussion of epistemic process is well outside the scope of the present essay (the interested reader is referred to Kruglanski, 1977; Kruglanski et al., 1978; or Kruglanski & Bar-Tal, Note 1), but passing consideration of its major attributes is necessary. Briefly, it is hypothesized that (a) the epistemic process is purposive, that is, it is initiated by the individual's purpose to acquire a given bit of knowledge. Furthermore, the process involves (b) the formulation of an epistemic problem the solution of which may advance the knower's purpose, (c) the generation of conjectures regarding the identity of such a solution, and (d) the deductive assessment of such conjectures by checking their implications for (logical) consistency with available evidence.

The epistemic problem just referred to is conceptualized as a set of mutually exclusive propositions on some topic, among which a knower might wish to choose. Only some epistemic problems may be explanatory, that is, revolve about because-type propositions; other epistemic problems may not, in this sense, be explanatory. For an example of an explanatory problem, consider someone choosing among the propositions "John laughed at the comedian, (only) because of himself," "John laughed at the comedian (only) because of the comedian," and "John laughed at the comedian (only) because of the circumstances."[4] For an example of a nonexplanatory problem, consider someone choosing among the propositions "The way to the market lies to the left," ". . . to the right," or "directly ahead."

The same epistemic process is assumed to be involved in both examples and in all other cases of epistemic behavior. For instance, an individual would not usually wonder about John's laughter or about the way to the market unless he/she had a purpose for wanting to know those things. Furthermore, in tackling any epistemic problem the lay inquirer would be likely to form conjectures about the correct resolution. For example, one might conjecture that John laughed at the comedian because of

---

[3] As with any scientific proposition, the assumption of process invariance in this case is not intended as an a priori truth but as a hypothesis open, in principle, to empirical test. In this connection it is of interest to note that at least some suggestions regarding the presumptive diversity of epistemic processes (e.g., Kelley's, 1971, 1972, proposals for separate attributional principles and schemata) are interpretable as instances of the same epistemic process applied to different contents of knowledge (cf. Kruglanski & Bar-Tal, Note 1).

[4] Roughly modeled after problems employed in McArthur's (1972) research.

himself or that the way to the market lies to the left. Such conjectures would then be tested via a deduction of some of their implications and the assessment of these against relevant evidence. Thus, the hypothesis that John laughed because of himself implies that he would also laugh on other similar occasions, that other people would not laugh similarly, and so forth. Data consistent with such implications would augment the knower's confidence regarding the correctness of the conjecture involved (as has been demonstrated empirically by McArthur, 1972). Similarly, with a nonexplanatory problem, the hypothesis that the way lies to the left implies among other things that if one pursued the path to the left one would eventually wind up at the market. Evidence consistent with such an implication would again bolster the knower's confidence regarding the validity of the hypothesis at stake.

The above discussion suggests that the same epistemic process may be involved in the acquisition of diverse epistemic contents. From a standpoint of scientific generalization it would thus seem that we should have a unitary theory of process applicable alike to divergent contents of knowledge. Still, an individual's epistemic behavior in a specific instance should vary considerably as a function of the contents investigated. Someone wishing to account for John's laughter would thus care very little about the consequence of turning left at the crossroads, similarly, someone interested in reaching the market would be indifferent, for example, to John's behavior toward various comedians. If so, the question arises whether the theory of epistemic process may not be profitably supplemented by various theories of content, that is, by the specification of various content categories that the lay person may occasionally employ in the course of epistemic behavior. Indeed, extant attributional formulations have frequently included various such categories, for example, those of internal versus external causality (cf. Kelley, 1971), of endogenous versus exogenous causality (cf. Kruglanski, 1975), of ability, luck, effort, and task difficulty (cf. Weiner, 1974). In this sense Buss's proposal to include the reason category within the attribution framework jibes well with an already popular trend in attributional theorizing.

Still, an immense difficulty with any program of constructing content-bound epistemic models (those constructed around specific concepts or statements in the layman's repertory) is their potentially infinite variety. For exactly the same reason, the program of the ordinary-language philosophers, to explicate the terms in everyday discourse, can hardly be expected to get off the ground. The everyday discourse may contain an infinity of terms, so the task of their explication may only begin but never end.

Furthermore, people across cultures and historical periods may have vastly different conceptual repertories and so must deal in epistemic contents that may vastly differ. Hence, an epistemological theory constrained to specific elements of content (e.g., to explanatory propositions or to special kind of explanatory propositions) seems condemned to the throes of cultural and historic relativism (see, e.g., Gergen, 1973, 1976). Finally, to the extent that a normal person may pose divergent epistemic problems across varying occasions, an epistemic (or attributional) framework restricted to a narrow set of contents is perforce situationally specific and particularistic. Thus, we may not have content elements in an epistemological theory for the very same reason that we must have a unitary theory of epistemic process: the fundamental scientific commitment to generalization. The contextually shifting contents of relevant knowledge, while undoubtedly essential to the full understanding and prediction of human conduct, must be divined specifically for each new situation rather than derived from an invariant set of principles.

Reference Note

1. Kruglanski, A. W., & Bar-Tal, D., *Emerging conceptions of attribution: An integrative review*. Unpublished manuscript, Tel-Aviv University, 1979.

References

Anscombe, G. E. M. *Intention*. London: Blackwell, 1957.

Austin, J. L. *How to do things with words*. Oxford, England: Oxford University Press, 1962.

Buss, A. R. Causes and reasons in attribution theory: A conceptual critique. *Journal of Personality and Social Psychology*, 1978, *36*, 1311–1321.

Gergen, K. J. Social psychology as history. *Journal of Personality and Social Psychology*, 1973, *26*, 309–319.

Gergen, K. J. Social psychology, science and history. *Personality and Social Psychology Bulletin*, 1976, *2*, 373–383.

Harvey, J. H., Ickes, W. J., & Kidd, R. F. (Eds.). *New directions in attribution research* (Vols. 1 and 2). Hillsdale, N.J.: Erlbaum, 1976, 1978.

Hendrick, C. Social psychology as history and as traditional science: An appraisal. *Personality and Social Psychology Bulletin*, 1976, *2*, 392–403.

Jones, E. E., Kanouse, D. E., Kelley, H. H., Nisbett, R. E., Valins, S., & Weiner, B. *Attribution: Perceiving the causes of behavior*. Morristown, N.J.: General Learning Press, 1972.

Jones, E. E., & Nisbett, R. E. (Eds.). *The actor and the observer: Divergent perceptions of the causes of behavior*. Morristown, N.J.: General Learning Press, 1971.

Kelley, H. H. Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 15). Lincoln: University of Nebraska Press, 1967.

Kelley, H. H. *Attribution in social interaction*. Morristown, N.J.: General Learning Press, 1971.

Kelley, H. H. *Causal schemata and the attribution process*. Morristown, N.J.: General Learning Press, 1972.

Kruglanski, A. W. The endogenous–exogenous partition in attribution theory. *Psychological Review*, 1975, *82*, 387–406.

Kruglanski, A. W. The place of naive contents in a theory of attribution: Reflections on Calder's and Zuckerman's critiques of the endogenous–exogenous partition. *Personality and Social Psychology Bulletin*, 1977, *3*, 592–605.

Kruglanski, A. W., Hamel, I. A., Maides, S. A., & Schwartz, J. M. Attribution theory as a special case of lay-epistemology. In J. H. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research* (Vol. 2), Hillsdale, N.J.: Erlbaum, 1978.

Manis, M. Comment on Gergen's "Social psychology as history." *Personality and Social Psychology Bulletin*, 1975, *1*, 450–455.

Manis, M. Is social psychology really different? *Personality and Social Psychology Bulletin*, 1976, *2*, 428–437.

McArthur, L. A. The how and what of why: Some determinants and consequences of causal attribution. *Journal of Personality and Social Psychology*, 1972, *22*, 171–193.

Popper, K. R. *The logic of scientific discovery*. New York: Wiley, 1961.

Ryle, G. *The concept of mind*. London: Hutchinson, 1949.

Schlenker, B. R. Social psychology and science. *Journal of Personality and Social Psychology*, 1974, *29*, 1–15.

Schlenker, B. R. Social psychology and science: Another look. *Personality and Social Psychology Bulletin*, 1976, *2*, 384–390.

Thorngate, W. Process invariance: Another red herring. *Personality and Social Psychology Bulletin*, 1975, *1*, 485–488.

Turner, M. B. *Philosophy and the science of behavior*. New York: Appleton-Century-Crofts, 1965.

Weiner, B. *Achievement motivation and attribution*. Morristown, N.J.: General Learning Press, 1974.

Received August 28, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 9, 1458–1461

On the Relationship Between Causes and Reasons

Allan R. Buss
University of Calgary, Calgary, Canada

Criticisms (Harvey & Tucker, Kruglanski) of my recent conceptual critique of attribution theory are criticized. Further analysis of the distinction between cause and reason is undertaken, where their relationship is set forth in terms of the growth of knowledge (at both the individual and the collective level). The model involves converting causes into reasons and makes use of the psychoanalytic therapy session as an analogue for purposes of explicating the process.

Reply to Harvey and Tucker

Although Harvey and Tucker (1979) would seem to agree that the cause–reason distinction is valid and important for attribution theory, they are somewhat concerned about what they regard as certain logical and empirical problems inherent in my analysis (Buss, 1978).

In regard to logical matters, Harvey and Tucker believe that the distinction between "action" and "occurrence" is not an easy one to make. Thus they point out that physiological responses are often susceptible to conscious and deliberate control, thereby rendering them actions rather than occurrences. I quite agree, and am glad to see that my earlier example of just such a situation (Buss, 1978, p. 1318, footnote 6) did not go unnoticed.

A second example in their attempt to demonstrate some logical problems in this area concerns intelligent behavior. Their point here is that because intelligence rests in part upon uncontrollable innate factors, intelligent behavior is somehow less an action. I would like to know what behavior, including any conceivable action, is not dependent upon or affected by innate factors (as well as environmental factors)? The point is that the person engaged in intelligent behavior *freely* forms an intention to act. The limits and/or constraints surrounding the action do not make it less an action. Even if the test giver holds a gun to the test taker's head and orders him/her to complete the latest IQ test, such behavior would constitute action, since the test taker still does have the option to refuse. Thus, the distinction between action and occurrence is *conceptually* and *logically* relatively straightforward, given sufficient analysis.

When we move into the realm of empirical matters, defined here as experimental social psychology, then I quite agree with Harvey and Tucker that there are problems. Why shouldn't there be, considering that the distinctions that I have stressed are *conceptual* rather than empirical? How could one operationally define a reason or a cause, an action or an occurrence? I cannot think of any simple way. I do think that an experimenter sensitive to the conceptual issues I have raised can arrange the task demands such that there is an avoidance of category errors. Reason and logic will guide the experimenter here. I cannot offer a pat formula. My whole point was to alert attribution theorists to the conceptual confusion in their area. If they themselves cannot remedy such confusion in their concrete researches, as Harvey and Tucker would seem to imply, then they should abandon this type of research. My challenge cannot be thrown back to me, since it rests with those who consider themselves attribution theorists and who believe that their ex-

Requests for reprints should be sent to Allan R. Buss, Department of Psychology, The University of Calgary, 2920 24th Avenue, N.W., Calgary, Canada T2N 1N4.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3709-1458$00.75

perimental program has some potential to enlighten us about lay explanation.

Reply to Kruglanski

Kruglanski (1979) has argued that causal explanation (efficient cause) is a generic form of explanation and that reason explanation (final cause) as well as Aristotle's other two types (formal cause and material cause) are "merely" special instances of efficient causality. In other words, Kruglanski is a monist, a scientific monist in the grand tradition of Hempel (1966), who attempted to reduce Dray's (1957) use of reasons in historical explanation to a special case of efficient causality. The point, however, is that reasons and causes are truly distinct logically. Perhaps the best explication of this distinction available to psychologists is Harré and Secord's (1973), especially chapters 1, 2, and 8:

> In accounts of human action reasons appear in a justificatory context, and logically imply the propriety of a happening, as opposed to its existence, which would follow from the description of its cause. (p. 148)

Thus the connection of reasons to an event is *logical*, whereas that between causes and an event is *physical*, that is, space/time. This analysis is, I think, the best thus far, and it avoids the necessity to totalize *either* causal explanation à la Hemple (and most recently Kruglanski) *or* reason explanation à la Winch.

In the same article, Kruglanski states that there are an infinite number of kinds of explanation. He cites mathematical, mechanical, geographic, demographic, social, and other types. Now, any of these types of explanations *can* be classified or reduced to one or more of Aristotle's four basic types. Kruglanski has confused form and content.

Kruglanski does make the valid point that actors are not necessarily required to give a reason account of their action and in fact may offer a causal explanation. I think my initial position, claiming that actors can only give a reason account of their action, was too extreme. It would seem that both causes and reasons could figure in an actor's self-explanation. However, I think that Harré and Secord (1973) make an important point in this connection when they state, "For 'propaganda' purposes some justificatory reasons are passed off as causes, as e.g., in the substitution in common speech of 'need' for 'want'" (p. 149). The example they give concerns the alcoholic who states that he/she "needed" a drink (implying uncontrollable causal factors) rather than that he/she "wanted" a drink (implying reasons that fail to justify the action). To the extent that taking a drink is difficult to justify, the actor will try to offer a causal explanation the validity of which is suspect.

On the Relationship Between Causes and Reasons

Let me now attempt a more explicit analysis of the distinction between cause and reason by elaborating upon their relationship. In maintaining a logical distinction between causes and reasons, I would not claim that there is no relationship between them—in fact, quite the contrary. The first point to note in this regard is that a cause *may* be cited as a reason, that is, it may figure in one's justificatory explanation. However, *not* all reasons are causes, since not every reason would have the space/time or antecedent–consequent connection to an event. This view is an extension of my previous position rather than being inconsistent with it.

The second matter to consider in the relationship between causes and reasons is more complex and involves the process of actually converting causes into reasons. A model that illustrates just such a relationship between causes and reasons has been developed and described by Apel (discussed in Radnitzky, 1970), Habermas (1971), and Ricoeur (1970), as well as Taylor (1973), where the context is arriving at increased knowledge of self. Essentially, the model involves the dialogue between patient (actor) and therapist (observer) in the psychoanalytic encounter. As long as the actions of the patient can be understood by his/her offering reasons (thereby arriving at valid intersubjective meanings, i.e., the *point* of the action is intelligible to both patient and therapist), then there is no need for the therapist (observer) to introduce

hypostatized causal forces. When there is a breakdown in a reason account of action, however, it is then necessary to resort to causal analysis, where the therapist (observer) attempts to formulate a causal explanation involving unconscious or hidden motives. To the extent that these causal factors are brought into the patient's (actor's) consciousness (through the technique of appropriately timed interpretations—thereby facilitating "insight"), then, according to the above-mentioned individuals, they *cease functioning as causes* and become matters for deliberation, that is, reasons. Thus the road toward greater self- (reason) understanding and emancipation from hypostatized causal forces lies in *converting causes into reasons*. In this way, causal analysis mediates, or is in the service of, greater (reason) understanding.

The psychoanalytic encounter and the growth of self-knowledge operate at the individual level, and Habermas especially has generalized this basic paradigm to the collective enterprise of the growth of knowledge in the social or human sciences. Revealing the underlying social forces or causes of nonliberation requires criticism grounded in the emancipatory interest. At the social level, bringing into consciousness previously unknown hypostatized forces—causal forces that hinder greater self- (reason) understanding at a social level—will, of necessity, take the form of a critique of ideologies. When previously hidden social forces are brought into consciousness, they become matters of deliberation, that is, are converted into reasons, and are thereby made not to function causally.

We can be somewhat critical of the boldness of the claims of Apel, Habermas, Ricoeur, and Taylor regarding the conversion of causes into reasons and the supposed neutralization of the causal efficacy of the cause in question as a result. If only matters were that simple and straightforward! Unfortunately, as critics of the role of insight in psychotherapy have pointed out (e.g., Hobbs, 1962), there is often no change in maladaptive behavior even after sufficient insight has been demonstrated by the patient, that is, after previously unconscious causes are brought into consciousness. According to what we might aptly dub the "conversion model," the causes should cease to function once they enter consciousness and become matters for deliberation (reasons). Since this is often not the case in the psychoanalytic encounter and most certainly does not apply to becoming conscious of previously hidden social forces, it is necessary to add a very important qualification to the conversion model.

The necessary qualification of the conversion model is that the bringing of previously hidden causes into consciousness—and the institution or creation of new reasons by this means—does not ipso facto neutralize or destroy the original cause. Rather, when a cause is converted into a reason, the cause may still operate, but now the awareness of that cause may *potentially* lead to eventual elimination of the cause qua cause. Thus, the conversion of causes into reasons may (but not necessarily) provide for deflating previously unconscious causes of their causal power.

The modified conversion model proposed here permits going beyond the implications of attribution theory for psychotherapy that have thus far been considered. Storms (1973), Storms and Nisbett (1970), and Valins and Nisbett (1972) have all discussed both the desirability of and the means for having the patient (actor) adopt more of an observer perspective with respect to his/her action. In this way, the patient (actor) would tend to give more dispositional or self-attributions in explaining his/her inappropriate or maladaptive action and would thereby assume greater responsibility for it (and thus, presumably, greater responsibility for changing it). The modified conversion model permits us to be a little more sophisticated in understanding how, in fact, a greater frequency of internal attributions made by the patient (actor) can work toward the goal of greater self-understanding and emancipation. That is to say, as a person becomes more self-conscious of previously unconscious internal causes vis-à-vis maladaptive behavior, such causes may (but not necessarily) be made causally impotent through conversion into reasons.

However, it should be noted that a higher frequency of internal causal attributions on the part of the patient (actor) is not in and of itself desirable if the attributions are not valid. Storms (1973; see also Galper, 1976) did recognize that increasing the number of self-attributions may actually lead to an exacerbation of the pathological behavior if the causes of such behavior are external. The modified conversion model also goes beyond abstract individual liberation by incorporating unconscious social causes of nonliberation within its framework. Thus one must guard against the reification of abstract internal causes (dispositions), as well as abstract external causes (situations), and adopt more of a dialectical view of subjective (internal) and objective (external) forces. Viewed in this way, the successful conversion of causes into reasons, and thus the achievement of greater human emancipation, is two dimensional rather than one dimensional.

Conclusion

Causes and reasons are both (a) logically distinct and (b) intimately connected to each other. In order to understand truly the nature of lay explanation, it is necessary to take cognizance of both of these conclusions. If attribution theorists will not or cannot do this (because of the inherent limitations of their experimental methodology and philosophy of science), then they should give up any pretense that they are better able to explain lay explanation than are the lay explainers.

References

Buss, A. R. Causes and reasons in attribution theory: A conceptual critique. *Journal of Personality and Social Psychology,* 1978, *36,* 1311–1321.

Dray, W. *Laws and explanation in history.* Oxford, England: Oxford University Press, 1957.

Galper, R. E. Turning observers into actors: Differential causal attribution as a function of "empathy." *Journal of Research in Personality,* 1976, *10,* 328–335.

Habermas, J. *Knowledge and human interests.* Boston: Beacon Press, 1971.

Harré, H., & Secord, P. F. *The explanation of social behavior.* Totowa, N.J.: Littlefield, Adams, 1973.

Harvey, J. H., & Tucker, J. A. On problems with the cause–reason distinction in attribution theory. *Journal of Personality and Social Psychology,* 1979, *37,* 1441–1446.

Hempel, C. G. Explanation in science and history. In W. H. Dray (Ed.), *Philosophical analysis and history.* New York: Harper & Row, 1966.

Hobbs, N. A new cosmology. *American Psychologist,* 1962, *17,* 18–34.

Kruglanski, A. W. Causal explanation, teleological explanation: On radical particularism in attribution theory. *Journal of Personality and Social Psychology,* 1979, *37,* 1447–1457.

Radnitzky, G. *Contemporary schools of metascience* (Vols. 1 and 2). Stockholm: Scandinavian University Books, 1970.

Ricoeur, P. *Freud and philosophy: An essay on interpretation.* New Haven, Conn.: Yale University Press, 1970.

Storms, M. D. Videotape and the attribution process: Reversing actors' and observers' points of view. *Journal of Personality and Social Psychology,* 1973, *27,* 165–175.

Storms, M. D., & Nisbett, R. E. Insomnia and the attribution process. *Journal of Personality and Social Psychology,* 1970, *16,* 319–328.

Taylor, C. Peaceful coexistence in psychology. *Social Research,* 1973, *40,* 55–82.

Valins, S., & Nisbett, R. E. Some implications of attribution process for the development and treatment of emotional disorders. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior.* Morristown, N.J.: General Learning Press, 1972.

Received December 15, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 9, 1462–1468

Competence and the Overjustification Effect: A Developmental Study

Ann K. Boggiano and Diane N. Ruble
Princeton University

This study examined the conditions under which information regarding competence would mitigate the negative side effects of rewards on the intrinsic interest of preschool and middle elementary school children. Children engaging in a task of high initial interest anticipated a reward made contingent either upon meeting a standard based on absolute performance level or upon task engagement alone, or they were not rewarded. In addition, they were provided with direct information concerning competence presented in terms of social comparison. The pattern of results indicated that the preschool children were primarily affected by information about meeting the absolute standard but not the social comparison information. That is, the overjustification effect did *not* occur when attaining a reward was made contingent on meeting an absolute standard of performance. In contrast, social comparison information superseded the effect of the contingency of the reward on subsequent interest in the target task for the older children. These findings suggest the importance of research from a developmental perspective for attempts to delineate the conditions under which rewards may avert the undermining of intrinsic interest.

A number of researchers have demonstrated that offering a reward to a person for engaging in an otherwise enjoyable activity will undermine that person's subsequent interest in that activity when the reward is discontinued—a phenomenon typically explained by means of an attributional analysis (Bem, 1972; Nisbett & Valins, 1971) and termed the *overjustification effect* (Lepper, Greene, & Nisbett, 1973). That is, if the justification for having undertaken an interesting task is perceived to be an overly sufficient reason for task engagement (e.g., an extrinsic incentive), the person would then infer that behavior was motivated by the reward itself rather than by enjoyment, thereby decreasing interest in the target task when the reward is no longer available. A substantial number of studies have shown the reliability of this effect across both types of reward and population (cf. Condry, 1977, Deci, 1975; Lepper & Greene, 1978; Ross, 1976).

The overjustification effect, however, is not assumed to be an inevitable result of reward attainment. Based on theories postulating that intrinsic interest stems from perceptions of competence and self-determination (Deci, 1975; White, 1959), several investigators have hypothesized that rewards conveying information regarding competence should sustain rather than undermine subsequent intrinsic interest (Arkes, 1978; Lepper & Greene, 1978). Specifically, it has been argued that level of intrinsic interest should vary directly with information regarding competence or incompetence that is conveyed by means of reward systems.

A few lines of research have provided some support for the competency hypothesis. Recent research has demonstrated that a reward made contingent on meeting a performance

The authors are grateful to Mark Lepper, Michael Ross, Thane Pittman, and Nina Feldman for many helpful suggestions on an earlier draft of this paper, and to Jeanne Courtright, Barbara Hoyler, and Paget Gross for their invaluable assistance in collecting the data.

Requests for reprints should be sent to Diane N. Ruble, Department of Psychology, Princeton University, Princeton, New Jersey 08540.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3709-1462$00.75

standard produces more interest than a reward made contingent on task engagement alone, as shown in both behavioral (Karniol & Ross, 1977) and attitudinal measures (Enzle & Ross, 1978). These results were interpreted as suggesting that rewards made contingent on successful performance provide cue value regarding competency at the task, thereby maintaining intrinsic interest. Similarly, research using a social reward paradigm has shown that task-related praise sustains or even enhances intrinsic interest (Anderson, Manoogian, & Reznick, 1976; Deci, 1975; Pittman, Davey, Alafat, Wetherill, & Wirsul, in press; Swann & Pittman, 1977). This body of research, then, has provided indirect support for the competency hypothesis by demonstrating that intrinsic interest is not undermined when reward is conveyed by means of verbal feedback or is made contingent on successful performance.

The primary purpose of the present study was to test the competency hypothesis by manipulating the availability of *direct* information about competence. That is, although intrinsic interest should be maintained when reward attainment is made contingent on an absolute performance standard, the competency hypothesis would predict that if more direct information about competence or incompetence is additionally provided (in this case, performance level relative to others), subsequent level of interest should vary directly with the type of information regarding competence that is presented.

The competency hypothesis was also assessed by means of a developmental analysis of the effect of different kinds of competency information on intrinsic interest. According to Veroff's (1969) developmental theory of achievement motivation, it is not until 7–8 years of age that children derive feelings of competence from *comparative* standards of excellence; however, even very young children (e.g., preschoolers) are assumed to make evaluative judgments regarding their level of competence based on *absolute* performance standards. Evidence supporting these propositions is provided in several recent studies (e.g., Ruble, Parsons, & Ross, 1976; Ruble, Boggiano, Feldman, & Loebl, in press). These developmental findings suggest an alternative explanation for the results of one recent study where rewards made contingent on a comparative standard were *not* found to sustain subsequent interest in nursery school children (Greene & Lepper, 1974). Specifically, the findings described above would suggest that this failure to support the competency hypothesis may have been because young children do not use comparative information in making self-evaluations of competence. Thus, if the findings of the present study supported the predictions that competency information conveyed in relative terms mitigates the undermining effect of reward for older but not for younger children, additional support for the competency hypothesis would then be obtained.

In the present study, children at two age levels engaged in a task that pilot testing revealed to be highly interesting for both groups. Upon completion of the target task, children were provided with competence information in terms of relative and/or absolute level of performance, which was predicted to avert the undermining effect of reward on intrinsic interest. However, the two types of information concerning competence (absolute vs. relative) were expected to affect the children differentially at the two age levels in the manner described above.

To test these hypotheses, children in the experimental conditions anticipated a reward made contingent either upon task engagement alone or upon meeting a performance standard (i.e. competence information based on absolute performance level). In addition, the children were either provided with information regarding their performance relative to peers (i.e. competence/incompetence information conveyed directly in terms of social comparison) or were not provided with this information. Finally, a no-information-regarding-competence/no-reward group was included as a baseline control.

Method

Subjects

Subjects were 147 children of predominantly middle-class backgrounds drawn from two private nursery schools and two public elementary schools in

central New Jersey. Three children were excluded from the analyses for reasons explained below. The nursery school children ranged in age from 3–5 to 5–8, with a mean age of 4–6, whereas the mean age of the elementary school children from grades 3 through 5 was 9–11. An attempt was made to distribute subjects equally by sex across each cell. Since preliminary analyses revealed no interaction of sex or experimenter with experimental conditions on any measure, the data were collapsed across these variables for purposes of analyses.

Procedure

The children were escorted individually into a mobile trailer parked near their school. After one of four female experimenters (three of whom were blind to the hypothesis) demonstrated to the children the different activities that she used with children from different schools at various times, she then told the children that they would have the opportunity to play the "find the hidden pictures" game.

The hidden picture game consisted of a stack of eight embedded figure tasks with five to eight embedded figures on each of the eight papers. The children were told that the object of the game was to find and circle the *one* particular thing on each of the eight papers which they would be instructed to circle after they turned on a tape recorder. They were also told that they should work as quickly as possible since they would have to proceed to the next picture when the person on the recorder instructed them to turn the page. Subjects were told, however, that they could find and circle other hidden pictures on each of the pages if they wanted to and had sufficient time. The children were given two practice trials on an additional embedded-figure task to ensure comprehension of the task and were also given the opportunity to practice using the recorder to facilitate recognition of the different switches (the "on" button was painted green and the "off" button was painted red). The colors and game context of the task (i.e. finding and circling one hidden picture on each page) were all designed to enhance the interest value of the task. Pilot testing did, in fact, reveal that this task was highly interesting for the two age groups. The interest value is also indicated by the high percentage of time the subjects in the baseline control condition played with the task.

After the practice session, subjects were shown the scoreboard that displayed numbers ranging from 0 to 8. Children in the performance-contingent reward conditions were told that they could earn two Hershey's Kisses *only* if they obtained a score of at least 3 out of a possible 8. Stars were painted above the numbers 3 through 8 to ensure comprehension of the absolute performance criterion, which they were told meant that they did "okay" at the game. Children in the task-contingent reward conditions were given the same information as children in the performance-contingent reward conditions; however, they were instructed that they would be given two Hershey's Kisses for simply *playing* the game, regardless of their score.

The scoreboard for children of both age groups in the social-comparison information conditions displayed the scores of same-age others: children in the relative incompetency condition were shown a scoreboard indicating that most of the other children had obtained a score of 7 out of the maximum 8, whereas some of the others had scored 6 or 8; children in the relative competency condition were shown a scoreboard indicating that most of the others had obtained a score of 1 out of the maximum 8, while some of the others had scored 0 or 2. Children in the control condition were simply told that when they completed the task, they could put their score under the number on the board indicating on how many pages they found the particular figure they had been instructed to circle.

After the experimenter was confident that the children understood all aspects of the task and after the children had been given the choice of whether or not to play the game, they turned on the recorder and began the task. (One fourth-grade boy in the control condition stated that he preferred not to play the target game, and his data were excluded from the analyses.) To make performance level comparable across age groups, the actual difficulty of the task increased with grade level, so that all children obtained a score of 4 (excluding one third-grade and one fourth-grade boy, whose data were excluded from the analyses). After tabulating their scores and indicating to the children the hidden pictures on each of the pages, all children except those in the control group received the reward and put their scores on the appropriate scoreboard.

The experimenter then told the children that she had another task for them to do but needed some time to prepare the task. The children were told that they could play with any of the games that were on the table during the 5–10 minute waiting period. The alternative games, which had been pretested for their interest value, were: marbles, crayons with a game that involved following the lines to complete a picture, two mazes, and an additional set of hidden pictures. Although the same tasks were used for the two age groups, the difficulty level of the tasks was matched to age level. After the experimenter placed the tape in the recorder that corresponded to the additional set of embedded figures tasks, she stressed that the children should feel free to play with any of the toys on the table and made the children aware that they would not be given any further candy. The experimenter then went to the back room, unobtrusively watched through a one-way mirror, and recorded the amount of time during a 6-minute interval that each of the children spent with the target activity.

Results

Following the procedure of previous research (Greene & Lepper, 1974; Lepper et

Table 1
Transformed Percent of Time Spent Playing With Target Task

Reward group	Social comparison condition: Others better (indicating incompetence)	Others worse (indicating competence)	No information
	Younger children		
Performance contingent	.33 (.44)	.29 (.41)	.45 (.62)
Task contingent	.17 (.23)	.20 (.27)	.19 (.27)
Control			.52 (.73)
	Older children		
Performance contingent	.21 (.28)	.49 (.67)	.39 (.55)
Task contingent	.25 (.32)	.50 (.68)	.21 (.29)
Control			.49 (.68)

Note. Raw percentiles appear in parentheses.

al., 1973), a log transformation, $Y = \log [Y + 1]$, on the proportion of time subjects spent with the target task was performed to produce homogeneity of variance (Winer, 1971, p. 400). The major hypotheses of the study were examined by a 2 (Nursery/Elementary School Children) × 2 (Performance-Contingent Reward/Task-Contingent Reward) × 3 (Relative Competence/Relative Incompetence/No Information Regarding Competence) analysis of variance on the data. These means are presented in Table 1.

The results concerning the undermining effect of reward were expected to be revealed in two comparisons. First, it was predicted that subsequent interest in the task would be greater when the reward was made contingent on meeting a performance standard than when reward was made contingent on mere task engagement. As expected, the main effect of contingency was significant, $F(1, 111) = 4.45$, $p < .05$. However, this effect was primarily due to the data of the younger subjects, as described below in the within-grade analyses and as shown in Table 1. Additional analyses between the combined control groups and the contingency conditions revealed that interest was undermined only in the task-contingent conditions, as predicted, $t(130) = 2.0$, $p < .05$; $t(130) = 1.16$, *ns*, for task-contingent and performance-contingent groups, respectively.

The second major prediction concerned the competency information. As expected, a significant Grade × Social Comparison interaction, $F(2, 111) = 3.28$, $p < .05$, showed that relative competency information affected the subsequent interest of the older but not the younger children. Within-grade analyses were performed to examine further the pattern of data obtained. The analyses indicated that for the younger age group, reward made contingent on task engagement produced less interest in the task- than performance-contingent reward, as expected, $F(1, 53) = 4.45$, $p < .05$. Contrasts between the control and the combined contingency conditions indicated that only task-contingent reward undermined interest for the nursery-school-aged children, $t(130) = 2.67$, $p < .01$; $t(130) = 1.31$, *ns*, task- and performance-contingent reward groups, respectively. Moreover, comparative information had no effect on their later interest in the task, as expected, $F(2, 53) < 1$.

The data pattern for the older children differed sharply from that of the younger age group. For this group, intrinsic interest differed as a function of the comparative information in the predicted direction, $F(2, 57) = 5.61$, $p < .01$. Children given information about their comparatively excellent performance showed more interest in the target task than those who thought they performed worse than others, $t(57) = 3.33$, $p < .01$, or than children who had been given no information regarding performance relative to others, $t(57) = 2.33$, $p < .05$. However, contingency

Table 2
Performance as a Function of Age Level and Condition

Reward group	Social comparison condition: Others better (indicating incompetence)	Others worse (indicating competence)	No information
	Younger children		
Performance contingent	10.9	12.9	10.6
Task contingent	12.2	11.8	9.8
Control			11.3
	Older children		
Performance contingent	12.0	11.6	11.4
Task contingent	10.9	8.6	11.2
Control			11.2

of reward did not independently affect level of interest or interact with the comparative information, both Fs < 1. A specific test of the overjustification hypothesis by comparing task-contingent reward with the control did produce the expected undermining effect of reward on interest for this age group only when *no* information regarding competency was provided, $t(130) = 2.17$, $p < .05$. Furthermore, information regarding relative incompetence was found to decrease interest relative to the control, as predicted, $t(130) = 2.13$, $p < .05$. Intrinsic interest was sustained, as can be seen in Table 1, only when subjects were given information that they were relatively competent or that they had met an absolute standard of competency when no relative information was provided.

Since advocates of competing response theory have argued that performance differences resulting from distraction in anticipation of reward can explain overjustification studies(e.g., Reiss & Sushinsky, 1975; 1976), analyses were performed to test the viability of this alternative explanation for the present data. A 2 (Grade) $\times$ 7 (Condition) analysis of variance on the children's overall level of performance did not show a main effect of condition or grade nor a significant interaction, all Fs < 1. The means for the performance level of children as a function of grade and condition can be seen in Table 2. Within-cell correlations between performance level and time spent at the target task were not significant. Thus, the competing response hypothesis appears unable to provide a satisfactory account of the present data, since no differences in level of performance were found.

Discussion

The results of the present study provide support for the hypothesized effect of different types of competency information on intrinsic interest. The current experiment demonstrated that reward based on successfully meeting an absolute performance standard did not undermine interest, thereby replicating previous research (Karniol & Ross, 1977). However, when more direct information regarding competency was presented by means of social comparison, the pattern of results obtained depended on developmental level.

Consistent with previous developmental findings (e.g., Veroff, 1969; Ruble et al., in press) and with present predictions, competence information sustained the intrinsic interest of the preschool children when it was presented in terms of an absolute but not a relative standard. In contrast, although the older children were not impervious to the absolute standard indicating competence, the differential social comparison information superseded the reward contingency in its effect on subsequent interest for this age group. Specifically, task-contingent reward decreased interest relative to the control only when no information about relative compe-

tency was provided or when the comparative information indicated relative incompetence. Task-contingent reward did not produce diminished interest in the older children, as it had in the younger children, when information could be gleaned about their comparative excellence.

Thus, these results provide strong support for the competency hypothesis by showing that social comparison, providing highly direct and unambiguous information about level of competence, superseded reward contingency in its effect on intrinsic interest in older children. Furthermore, when social comparison was not provided or was not used (i.e., by the younger children) the contingency of the reward upon absolute performance level (a different kind of competence information) had the predicted effect upon later interest.

The developmental findings in the present study are relevant to recent hypotheses about the effect of rewards on intrinsic interest in children. First, they do not support Harter's (1978) hypothesis that the presentation of extrinsic incentives undermines intrinsic interest less at earlier developmental stages than at later stages. Second, the present data suggest important qualifications to Arkes' (1978) argument that young children's use of the competence but not attribution principles (e.g., discounting) renders an attributional explanation of overjustification less plausible and makes the former principle the key factor in overjustification research. While the importance of the competence principle is obviously not disputed here, our findings demonstrate that young children are not affected by certain types of competency information, in contrast to Arkes' suggestions. Moreover, the present findings suggest that the type of information that will most effectively sustain interest in children depends on the developmental level of the child.

The present results are also relevant to interpreting the findings of previous research examining the relationship between information regarding competence and the effects of rewards on intrinsic interest. First, it appears that Greene and Lepper's (1974) failure to support the competency hypothesis was probably due to their attempt to manipulate preschooler's perceptions of competence by means of social comparison information. Second, it appears that some important aspects of the competency hypothesis were obscured in the study by Karniol and Ross (1977), since they grouped children varying in age from 4 to 9 years and conveyed information about competence by varying absolute and relative standards simultaneously. Specifically, although the older children's intrinsic interest may have been influenced by both the absolute and relative information provided, the younger children may have focused on the absolute standard and may have been unaffected by the relative information.

The mediating factor assumed to account for the present findings is differential perceptions of competence resulting from information in the various competency conditions. Perceptions of competence were not assessed here, however, because they would have been contaminated by the behavioral measure or, if taken before the behavior, would have potentially confounded this measure. Nevertheless, several lines of previous research provide support for the mediating variable hypothesized here. First, a number of studies have demonstrated that information regarding absolute and relative level of performance affects children's perceptions of competence in the manner suggested here (Ruble et al., 1976; in press). Second, a considerable amount of research in the achievement literature has demonstrated a relationship between perceived competence and various task-approach indices, including persistence (cf. Ruble & Boggiano, in press; Weiner, 1974; in press). Thus, both of the mediating links necessary to support the competency hypothesis (i.e., information → self-reported competence, and competence → persistence) are well demonstrated in previous related studies.

In summary, the results of the present study provide the strongest support to date for the hypothesis that information about competence can mitigate the undermining effect of reward on intrinsic interest. The availability of direct information regarding competence presented by means of social comparison appeared to override completely the effect of

reward on behavior. This finding raises the intriguing issue of the link between overjustification phenomena and the standard achievement motivation literature, as noted by several investigators (e.g., Deci & Porac, 1978; Lepper & Greene, 1978). Research is presently underway to explore the nature of this link.

References

Anderson, R., Manoogian, S., & Reznick, J. The undermining and enhancing of intrinsic motivation in preschool children. *Journal of Personality and Social Psychology*, 1976, *34*, 915–922.

Arkes, H. R. Competence and the maintenance of behavior. *Motivation and Emotion*, 1978, *2*, 201–211.

Bem, D. J. Self-perception theory. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 6). New York: Academic Press, 1972.

Condry, J. Enemies of exploration. *Journal of Personality and Social Psychology*, 1977, *35*, 459–477.

Deci, E. L. *Intrinsic Motivation*. New York: Plenum Press, 1975.

Deci, E. L., & Porac, J. Cognitive evaluation and human motivation. In M. R. Lepper & D. Greene (Eds.), *The hidden costs of rewards*. Hillsdale, N.J.: Erlbaum, 1978.

Enzle, M. E., & Ross, J. M. Increasing and decreasing intrinsic interest with contingent rewards: A test of cognitive evaluation theory. *Journal of Experimental Social Psychology*, 1978, *14*, 588–597.

Greene, D., & Lepper, M. R. Effects of extrinsic rewards on children's subsequent intrinsic interest. *Child Development*, 1974, *45*, 1141–1145.

Harter, L. Effectance motivation reconsidered. *Human Development*, 1978, *21*, 34–64.

Karniol, R. & Ross, M. The effect of performance-relevant and performance-irrelevant rewards on children's intrinsic motivation. *Child Development*, 1977, *48*, 482–487.

Lepper, M. R., & Greene, D. *The hidden costs of rewards*. Hillsdale, N.J.: Erlbaum, 1978.

Lepper, M. R., Greene, D., & Nisbett, R. E. Undermining children's intrinsic interest with extrinsic rewards: A test of the "overjustification hypothesis." *Journal of Personality and Social Psychology*, 1973, *28*, 129–137.

Nisbett, R. E. & Valins. *Perceiving the causes of one's own behavior*. Morristown, N.J.: General Learning Press, 1971.

Pittman, T. S., Davey, M. E., Alafat, K. A., Wetherill, K. V., & Wirsul, N. A. Informational vs. controlling verbal rewards, levels of surveillance, and intrinsic motivation. *Personality and Social Psychology Bulletin*, in press.

Reiss, S., & Sushinsky, L. Overjustification, competing responses, and the acquisition of intrinsic interest. *Journal of Personality and Social Psychology*, 1975, *31*, 1116–1125.

Reiss, S., & Sushinsky, L. The competing response hypothesis of decreased play effects. A reply to Lepper and Greene. *Journal of Personality and Social Psychology*, 1976, *33*, 233–244.

Ross, M. The self-perception of intrinsic motivation. In J. H. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research*. Hillsdale, N.J.: Erlbaum, 1976.

Ruble, D. N., & Boggiano, A. K. Optimizing motivation in an achievement context. In B. Keogh (Ed.), *Advances in special education* (Vol. 1). Greenwich, Conn.: JAI Press, in press.

Ruble, D. N., Boggiano, A. K., Feldman, N. S., & Loebl, J. H. A developmental analysis of the role of social comparison in self-evaluation. *Developmental Psychology*, in press.

Ruble, D. N., Parsons, J. E., & Ross, J. Self-evaluative responses of children in an achievement setting. *Child Development*, 1976, *47*, 990–997.

Swann, W. B., & Pittman, J. S. Initiating play activity of children: The moderating influence of verbal cues on intrinsic motivation. *Child Development*, 1977, *48*, 1128–1132.

Veroff, J. Social comparison and the development of achievement motivation. In C. P. Smith (Ed.), *Achievement-related motives in children*. New York: Russell Sage Foundation, 1969.

Weiner, B. *Achievement motivation and attribution theory*. Morristown, N.J.: General Learning Press, 1974.

Weiner, B. A theory of motivation for some classroom experiences. *Journal of Educational Psychology*, in press.

White, R. W. Motivation reconsidered: The concept of competence. *Psychological Review*, 1959, *66*, 297–333.

Winer, B. J. *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill, 1971.

Received July 7, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 9, 1469-1476

The Role of Social Comparison in Choice Shifts

George R. Goethals
Williams College

Mark P. Zanna
University of Waterloo, Waterloo, Canada

Research was conducted to test social comparison predictions regarding influence processes related to risk taking in groups. It was shown that shifts toward risk were as likely to occur in groups where subjects exchanged information about their positions on the Kogan and Wallach Choice Dilemma Questionnaire (CDQ) *and* information about their self-ratings of ability as in ordinary group discussions of the CDQ items. Subjects in groups where only information about CDQ positions was exchanged showed far fewer shifts to risk. These findings are discussed in terms of a social comparison analysis of the social influence processes involved in risky shifts, which assumes that comparison processes can be engaged fully only when comparability is established by knowledge of other group members' standing on traits thought to be related to risk taking.

In a comprehensive review of the risky shift literature, Dion, Baron, and Miller (1970) suggested that the existing research seemed to lend more support to Brown's (1965) "risk as a value" explanation of the risky shift than to accounts emphasizing diffusion of responsibility, leadership, or familiarization. However, Dion et al. noted that the existing data did not seem to offer strong support for the social comparison mechanism that was implicit in Brown's argument. Subsequently, Burnstein and Vinokur (cf. Burnstein & Vinokur, 1975, 1977; Burnstein, Vinokur, & Trope, 1973) proposed an alternative "persuasive arguments" version of the risk as a value idea. This account holds that subjects discussing the items in the Kogan and Wallach (1964) Choice Dilemma Questionnaire (CDQ) have available a preponderance of arguments favoring risk, presumably because risk is a cultural value. As a result of this preponderance of arguments, in any discussion about the CDQ items, subjects are likely to encounter more good arguments favoring risk that they had not considered before than arguments favoring caution. These arguments are likely to persuade subjects and produce genuine attitude change in the direction of increased risk.

Burnstein and Vinokur (1975) did acknowledge a "few embarrassing studies" that produced an effect constituting a "puzzle for the persuasive-arguments explanation" of the risky shift. These were the studies showing that a shift was sometimes produced when subjects were simply given information about others' positions (e.g., Teger & Pruitt, 1967). The explanation for these studies seemed to be social comparison, that is, that subjects shifted when they discovered the positions of other people, because they learned that they themselves were not as risky as they had imagined relative to others, and they wanted to correct that state of affairs by shifting. Although this effect has not been consistently obtained,[1] it did suggest that a shift could sometimes be produced without discus-

The authors thank Joan Hall, Thomas Martin, Linda McKend, Sandra Ryan, and Betsy Zanna for their help in collecting and analyzing the data, and John Holmes, David Myers, and Neil Vidmar for comments on an earlier draft of the manuscript.

Requests for reprints may be sent to either George R. Goethals, Department of Psychology, Williams College, Williamstown, Massachusetts 01267, or Mark P. Zanna, Department of Psychology, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1.

[1] Pruitt (1971a, 1971b, has cited several articles that have included one or more information exchange conditions. It is important to note that of the 13 replications only 4 actually resulted in significant risky shifts.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3709-1469$00.75

sion and therefore without opportunity for argumentation. However, Burnstein and Vinokur suggested that persuasion could account for even those shifts obtained without discussion. Their rationale was that knowledge of other people's positions inspires a self-generation of reasons for those positions. It is these reasons, not comparison pressures, that produce the shift. Burnstein and Vinokur argued, then, that persuasion was necessary and sufficient to account for all risky shifts. (This position has been disputed in reviews of the literature that are not necessarily sympathetic to social comparison theory, cf. Myers & Lamm, 1976. Also, Baron & Roper, 1976, have questioned whether Burnstein and Vinokur's 1975 data unequivocally support the self-generation of arguments idea.)

Recently, however, two new lines of research have been reported that revitalize the social comparison account of the risky shift. Blascovitch and his colleagues (Blascovitch & Ginsburg, 1973; Blascovitch, Ginsburg, & Veach, 1975) have conducted several studies showing shifts toward risk in groups placing bets in blackjack games in gambling establishments in the state of Nevada. They argued for a social comparison explanation for their findings. Baron and his colleagues (Baron & Roper, 1976; Sanders & Baron, 1977) have also proposed an account of risky shifts involving social comparison. This view holds that subjects want to be as risky as other group members or slightly more so because risk is a value, and that there is a tendency for the group as a whole to shift toward risk as a result of this desire. Countervailing values for being sensible or not reckless presumably place some kind of limit on this competition to be relatively risky. The most recent exposition of this approach, that of Sanders and Baron (1977), argues that social comparison plays a large but not exclusive role in choice shifts. They believe that both social comparison and persuasion operate to produce risky shifts and that neither is capable of accounting for all the data. They suggest that the two processes are not mutually exclusive and may even be complementary. For example, it is possible that subjects are receptive to arguments favoring risk that emerge in group discussion because such arguments justify a shift toward risk that they want to make because of social comparison concerns. The arguments can provide a justification for opinion changes that subjects want to make anyway.

The purpose of the present paper is to discuss several studies designed to explore more thoroughly the role of social comparison processes in choice shifts. The starting point for these investigations was Goethals and Darley's (1977) recent restatement of Festinger's (1954) social comparison theory within an attribution theory framework. Their formulation suggests that people will compare their opinions with others who should have similar opinions, given their standing on attributes related to and predictive of those opinions, and that people will feel that they have selected appropriate opinion positions if people who are similar on these attributes agree. A recent experiment by Zanna, Goethals, & Hill (1975) supports this analysis by showing that subjects do compare with others who are similar on related attributes.

What kind of attributes do people feel are related to risk taking? A study by Jellison and Riskind (1970) suggested that ability is perceived as related to risk. They contended, in fact, that risk is a cultural value, as Brown (1965) and others had suggested, because of its link with ability. Risk and ability are linked in that the probability that a person will succeed if he engages in a risky behavior is directly dependent on his ability—the more ability, the more probable success. Given the risk–ability link, social comparison theory implies that people will feel it is appropriate to take as much risk as others of equal ability but less risk than those who possess greater ability.

In order to test this reasoning, one initial and two follow-up experiments were conducted in which subjects were given information indicating that others of either extremely high or moderate ability had recommended a high degree of risk on CDQ items. It was predicted that subjects would feel comparable to others of moderate ability and would want to move toward risk themselves when exposed

to their risky recommendations, but would cease comparison with the superior-stimulus persons and would not be influenced by their risky recommendations. The first experiment was only partially successful. In neither experimental condition did subjects' recommendations differ significantly from those in a control, nor did they differ from each other. However, manipulation check data showed that many subjects in the high-ability conditions perceived themselves to be equal in ability to the stimulus persons. An internal analysis was performed where subjects were assigned to one of three groups depending on whether they had indicated on the manipulation check that they saw the stimulus persons as having more, less, or equal ability compared to themselves. Subjects in the group perceiving themselves as equal to the stimulus persons were more risky than subjects in the control condition and more risky than those who perceived the stimulus persons to have more ability. This finding suggests that being influenced toward risk in a group is dependent on a perception of similarity and comparability on the dimension of ability and on social comparison tendencies being fully engaged as a consequence.

This conclusion, however, is clearly an extrapolation from the data and is only supported by an internal analysis. Consequently two replications were conducted in which great emphasis was placed on embellishing the high-ability manipulations so that it would be impossible for the subjects to perceive themselves as being even remotely as talented as the stimulus persons. Unfortunately, the results for both replications were very similar to the results of the initial experiment. As in the initial study, many subjects perceived the stimulus persons in the high-ability conditions to be equal in ability to themselves, and in each case an internal analysis was performed. For both follow-up studies this analysis indicated, as before, that subjects shifted to risk more when they perceived the stimulus persons to be equal in ability than when they perceived the stimulus persons to be more able, and that only in the equal ability group was there a significant shift to risk.

Given the limited success of these experiments, a new approach was devised to test more directly the hypothesis that the extent to which groups will shift to risk is dependent on the extent to which social comparison tendencies are fully engaged in those groups. This approach grew out of the previously discussed finding that experimental procedures using information exchange, where subjects are given information about other group members' CDQ positions but not their reasons for adopting those positions, show shifts of lesser magnitude than those found following full discussions of CDQ items. These data suggest that social comparison is not sufficient to account for large choice shifts following group discussion and that persuasive arguments are needed fully to influence group members toward risk. How can these studies be reconciled with a social comparison analysis of choice shifts? The key would seem to be a careful consideration of what is and what is not exchanged in information exchange experiments.

The three experiments reported above suggest that ability information interacts with information about the stimulus person's positions on CDQ items and that for social comparison processes to operate fully in a group discussing risk, the group members need to know more than the other members' recommendations regarding risk; they need to have information about others' abilities as well. It seems reasonable to argue, then, that the traditional information exchange studies may not fully engage social comparison because they fail to give subjects ability information that is critical in making any comparison of recommendations regarding risk meaningful. That is, because subjects have no information about the group members' abilities, their recommendations regarding risk have ambiguous implications for the level of risk that is appropriate for the subject to recommend. Hence, we should not expect consistent shifts in traditional information exchange conditions (see Footnote 1).

Suppose subjects were placed in an information exchange situation where information about both position and ability is being conveyed. If each person in the group indicates

that he believes himself to be somewhat more able than the average person, as one might expect from Jellison and Riskind's (1970) results, then each person in the group should come to feel that other group members are comparable to themselves on the relevant background characteristic, ability (i.e., they are all "above average"). Social comparison should be fully engaged, and we should expect risky shifts of greater magnitude to occur.

But will shifts in the more complete information exchange condition be as large as those that result from group discussion? If the result of group discussion, in addition to exchange of information concerning position, is primarily exchange of information concerning ability, then our answer would be yes. If relevant arguments make a difference over and above the reassessment due to social comparison, however, then our answer would be no. At any rate, the purpose of our experiment is to give social comparison theory a fair test; the prediction is that a more complete information exchange condition will produce a shift toward risk of greater magnitude than a "traditional" information exchange condition. Whether the shift in this new information exchange condition will be as large as that following a group discussion is, for us, an empirical question.

Method

Subjects

Two replications of the experiment were conducted. In one replication, conducted at Rider College (Trenton, New Jersey), 38 male and 26 female summer students served as subjects; in the other replication, conducted at the University of Waterloo, 73 female students participated. In each case, care was taken to assure that the subjects participating in a given group were not close acquaintances. In the Rider replication each subject was paid for his or her participation; in the Waterloo replication, subjects were drawn from a voluntary subject pool associated with their introductory psychology course.

Procedure

Subjects were run in groups of four, except for nine groups in the Waterloo replication, which contained five subjects. Each group was composed of subjects of the same sex who did not know each other well. Four groups of subjects were run in each experimental condition in each replication.[2] The experimenter began by explaining that the study was concerned with decision making. He or she then described the nature of the choice dilemma problems, giving subjects a sample problem (the football captain). Next he or she asked subjects to read, think about, and answer the same four choice dilemmas as employed in the initial experiment (items dealing with the prospective chemistry graduate student, the prospective concert pianist, the businessman considering running for political office, and the research physicist). After subjects had given their initial responses (and the booklets had been collected), they were asked to fill out an anonymous background questionnaire that asked, "Compared to the average person of your age and sex, rate yourself in terms of your overall talent, creativity, and ability" (answered on a 7-point scale with points labeled "a great deal above [below] the average," scored 7 and 1, respectively, "moderately above [below] the average," "slightly above [below] the average," and "just at the average," scored 4).

Group discussion condition. At this point the procedures for the three experimental conditions diverged. Calibrating the time parameters employed by Teger and Pruitt (1967), subjects in the group discussion condition were given 30 minutes to discuss the set of four choice dilemmas. New booklets containing the problems were handed out to facilitate these discussions.

Information exchange of position condition. This condition closely resembled that conducted by Teger and Pruitt (1967). In addition to the new booklets of choice dilemmas, subjects were given a set of cards indicating the various odds depicted in the choice dilemma problems and were asked merely to indicate their current positions by holding up the appropriate card. In this condition each subject in turn indicated his or her current position three times for each problem.

Information exchange of position and ability condition. This condition resembled the previous information exchange condition except that subjects were also handed a set of cards that matched the alternative answers to the question for which they rated their own ability. Just prior to indicating their positions for the first time on the first choice dilemma, subjects were asked, again in turn, to indicate their answers to the ability question. In the Waterloo replication subjects were asked to indicate their ability positions again after the second of the four choice dilemmas. No talking was allowed in either information exchange condition.

Control condition. Subjects in the control condition were given only 10 minutes to reconsider their

[2] Actually, in the Rider replication, subjects in the control condition were run in groups of size 2, 3, or 4. For purposes of analysis subjects were randomly assigned to four groups of size 4.

recommendation. No information whatsoever about others' responses was provided.

In the Rider replication, subjects in all conditions were asked to make their final private recommendations after the entire discussion, information exchange, or reconsideration phase; in the Waterloo replication, final judgments were made after each item was processed. Finally, subjects were debriefed and, in the Rider replication, were paid for their participation.

Results

Preliminary analyses indicated no significant effect of sex of subject or interaction of sex with experimental conditions within the Rider replication; hence, data were collapsed across this dimension within this replication. In addition, preliminary analyses indicated no significant effect of replication or interaction of replication with experimental conditions; hence, data were also collapsed across this dimension.

Ratings of Ability

As expected subjects did rate themselves as more able than "the average person of (their) age and sex." The overall mean on the scale, which ranged from "a great deal below the average," scored 1, to "a great deal above the average," scored 7, was 5.36. Twelve subjects (or 9%) felt they were "a great deal above," 51 (or 37%) felt they were "moderately above," 50 (or 36%) felt they were "slightly above," and only 23 (or 17%) felt they were "just at the average." One subject in the sample (or 1%) felt he was "moderately below average." This distribution was similar for each condition.

Postmanipulation Shifts

A preliminary analysis indicated that the variation in the magnitude of shifts was considerably greater in the group discussion condition whether the unit of analysis was the group ($F_{max} = 28.15$, $p < .01$) or the individual subject ($F_{max} = 6.82$, $p < .001$).[3] Therefore, to best describe the data the primary analyses are nonparametric in nature.

Table 1 presents the number of groups in which more subjects shifted in the risky than in the cautious direction and the number of groups in which an equal or greater number shifted in the cautious direction. The overall chi-square analysis indicates that the treatment effect is highly reliable, $\chi^2(3) = 10.42$, $p < .02$. As can be seen in Table 1, this overall effect is primarily due to the fact that a greater proportion of groups in the group discussion and information exchange of position and ability conditions (81% combined) had more subjects shifting to risk than to caution; this percentage contrasts with that for the information exchange of position and control conditions (25% combined), $\chi^2(1) = 8.03$, $p < .01$; Yates' correction for continuity was applied to this statistic. This difference accounts for 97.5% of the overall differences among the four conditions. Individual comparisons indicated that (a) the proportion of groups with more shifts to risk than caution was not reliably greater in the group discussion condition than in either the information exchange of position condition or the control condition ($p = .13$, in each case, by Fisher's exact test) but (b) the proportion of groups with more shifts to risk than caution was significantly greater in the information exchange of position and ability condition than in either the information exchange of position condition or the control condition ($p < .05$, in each case, by Fisher's exact test). Finally, there was no difference between the group discussion and information exchange of position and ability conditions.[4]

---

[3] The direct comparisons of variances between the group discussion and the information exchange of position and ability conditions were highly reliable in both cases, $F(7, 7) = 15.28$, $p < .001$, and $F(34, 34) = 6.82$, $p < .001$, for the by-group and by-individual subject analyses, respectively.

[4] There are other possible nonparametric analyses. One could consider the total number of subjects who shifted to risk or caution or who did not change. As in the analysis above, a virtually identical proportion of subjects shifted to risk in the group discussion and exchange of position and ability conditions (59% combined), and a virtually identical proportion shifted to risk in the exchange of position and control conditions (36% combined). These proportions differ significantly, ($\chi^2(1) = 6.22$, $p < .02$. Another approach is to consider the proportion of groups shifting to risk, using as the criterion whether the final mean score of the subjects in a group is

Table 1
Number of Groups in Which More Subjects Did or Did Not Shift to Risk Than to Caution

Direction in which subjects in groups shifted	Experimental condition			
	Group discussion	Information exchange of position and ability	Information exchange of position	Control
More to risk than caution	6	7	2	2
The same or more to caution	2	1	6	6

Discussion

The central hypothesis of the present study was that there would be more shifts to risk resulting from exchanging information on CDQ items if subjects also exchanged information that could establish that they were similar on attributes related to risk taking. This hypothesis was confirmed. It is interesting to recall (see footnote 1) that information exchange experiments have yielded highly inconsistent data. Sometimes they produce a risky shift; sometimes not. It could be that whether or not a shift occurs depends on the degree to which subjects can infer important unknowns about the other person's position on related attributes. If they know absolutely nothing about the other persons, then little comparability can be assumed and no choice shift should follow information exchange. If, on the other hand, subjects are familiar with each other and similar levels of ability can be assumed, then social comparison will be more strongly engaged and a choice shift is more likely to follow.

The results support the general proposition that social comparison on a particular attribute is only meaningful when something is known about how you and the others should compare on that attribute, given your relative standings on related attributes. This point can easily be seen in many instances of opinion formation and change. Whenever a person expresses an opinion on any issue, the meaning of that expression must be interpreted in terms of that person's identity and his standing on various related attributes. An American congressman appraises the president's performance. We would like to know to what political party he belongs. A Canadian citizen signs a petition opposing confederation. We would like to know whether the person is a Francophone or Anglophone and in what province he or she lives. The point is equally true with ability comparison. A person has just run the hundred-yard dash in 11 seconds. How old was the person? Male or female? What kind of coaching, equipment, and so forth, was made available to the person? What were the track and weather conditions? The meaning of another's opinion or performance can only be calibrated by knowing the person's standing on related attributes. Knowledge of environmental factors might also be necessary in many instances.

How, then, do present findings relate to the controversy in the literature regarding the various explanations for the risky shift and other choice shifts? The present results show that when information about one's self-estimate of ability and one's recommendations about risk are exchanged, groups shift as often toward risk as they do following group discussion. The mechanism producing the shift in the information exchange of position

riskier or more cautious than the mean of their initial scores. The proportion of groups shifting to risk is 75% in both the group discussion and exchange of position and ability conditions. The proportion of groups shifting to risk is 50% in both the exchange of position and control conditions. With the small number of groups, these proportions do not differ significantly. Overall, however, the findings from these analyses are similar to those from the analysis above.

and ability condition cannot be persuasion. There is no argumentation in this condition. Even if one accepted a self-generation-of-arguments explanation of those studies where information exchange of position has produced a risky shift, it is difficult to see how such an explanation could account for the clear difference between the two information exchange conditions reported here.

While it seems clear that social comparison produces the shift in the information exchange of position and ability condition, there is considerable doubt as to what is producing the risky shift in the group discussion condition. This uncertainty is not new. There has for many years been doubt about the cause of the shift in that condition. It seems that the answer would well be social comparison. The same kind of social comparison that is engaged in our exchange of position and ability condition could occur in the group discussion condition. Through discussion, participants could quickly see that the others were socially comparable. In order to explore the possibility that subjects would come to view their fellow group members as more able (and thus more similar to themselves in ability) following group discussion, subjects in the experimental conditions were asked at the very end of the experiment to indicate how they felt about each person in their group "in terms of his (or her) overall talent, creativity, and ability" (using same 7-point scale on which they had previously indicated their own ability). In the information exchange of position condition, subjects in 6 (of the 8) groups, on the average, believed they were at least a third of a unit more able than the "average" group member; in contrast, this was the case in only 3 groups in the information exchange of position and ability condition and in only 1 group in the group discussion condition, $\chi^2(2) = 6.51$, $p < .05$. The difference between the group discussion and the information exchange of position conditions, which is reliable at the 5% level by Fisher's exact test, indicates that group discussion apparently did lead subjects to think more highly of their fellow group members. Moreover, there was a relationship between the perceived ability of the "average" group member and risk taking. Of the 10 groups feeling that fellow group members were at least one-third of a unit less able than they themselves, only 5 had more subjects shifting in the risky than in the cautious direction, 1 had an equal number of risky and cautious shifters, and 4 had more subjects shifting in the cautious than in the risky direction; of the remaining 14 groups, who essentially saw others as similar in ability, 10 had more subjects shifting in the risky than in the cautious direction, 4 had equal numbers of risky and cautious shifters, and no group had more subjects shifting in the cautious direction, $\chi^2(2) = 6.97$, $p < .05$. (It might be noted that the one group in the group discussion condition, which for some reason did not come to believe that fellow group members were basically similar in ability, was the only group in that condition that had more cautious than risky shifters.)

In arguing favorably concerning a social comparison analysis of the risky shift, we do not mean to imply that persuasion does not occur or that it does not occur in groups discussing risk. The processes of persuasion and comparison, again, are not mutually exclusive. Many things happen in group discussions. That both comparison and persuasion are occurring at the same time, in different amounts for different individuals, seems perfectly plausible. The goal of the present research has not been to rule out persuasion. It has simply been to show that influence produced by social comparison can be substantial but that social comparison is not fully engaged unless particular kinds of important information are made available to subjects, making the fact of comparability evident to them. Once comparability has been appropriately established, comparison can produce opinion shifts of the magnitude found in studies of the risky shift. In real life, discussion may be necessary to establish comparability. The present study shows, however, that even if comparability is established by other means, similar opinion shifts occur.

References

Baron, R. S., & Roper, G. A reaffirmation of a social comparison view of choice shifts, averaging, and extremity effects in autokinetic situations. *Journal*

of Personality and Social Psychology, 1976, *33,* 521–530.

Blascovich, J., and Ginsburg, G. Blackjack and the risky shift. *Sociometry,* 1973, *36,* 42–55.

Blascovich, J., Ginsburg, G., & Veach, T. A pluralistic explanation of choice shifts on the risk dimension. *Journal of Personality and Social Psychology,* 1975, *31,* 422–429.

Brown, R. *Social psychology.* New York: Free Press, 1965.

Burnstein, E., & Vinokur, A. What a person thinks upon learning he has chosen differently from others: Nice evidence for the persuasive-arguments explanation of choice shifts. *Journal of Experimental Social Psychology,* 1975, *11,* 412–426.

Burnstein, E., & Vinokur, A. Persuasive argumentation and social comparison as determinants of attitude polarization. *Journal of Experimental Social Psychology,* 1977, *13,* 315–332.

Burnstein, E., Vinokur, A., & Trope, Y. Interpersonal comparison versus persuasive argumentation: A more direct test of alternative explanations for group-induced shifts in individual choice. *Journal of Experimental Social Psychology,* 1973, *9,* 236–245.

Dion, K. L., Baron, R. S., & Miller, N. Why do groups make riskier decisions than individuals? In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 5). New York: Academic Press, 1970.

Festinger, L. A theory of social comparison processes. *Human Relations,* 1954, *7,* 117–140.

Goethals, G. R., & Darley, J. M. Social comparison theory: An attributional approach. In J. M. Suls & R. L. Miller (Eds.), *Social comparison processes: Theoretical and empirical perspectives.* Washington, D.C.: Hemisphere, 1977.

Jellison, J. M., & Riskind, J. A social comparison of abilities interpretation of risk-taking behavior. *Journal of Personality and Social Psychology,* 1970, *15,* 375–390.

Kogan, N., & Wallach, M. A. *Risk taking: A study in cognition and personality.* New York: Holt, Rinehart & Winston, 1964.

Myers, D. G., & Lamm, H. The group polarization phenomenon. *Psychological Bulletin,* 1976, *83,* 602–627.

Pruitt, D. G. Choice shifts in group discussion: An introductory review. *Journal of Personality and Social Psychology,* 1971, *20,* 339–360. (a)

Pruitt, D. G. Conclusions: Toward an understanding of choice shifts in group discussion. *Journal of Personality and Social Psychology,* 1971, *20,* 495–510. (b)

Sanders, G. S., & Baron, R. S. Is social comparison irrelevant for producing choice shifts? *Journal of Experimental Social Psychology,* 1977, *13,* 303–314.

Teger, A. I., & Pruitt, D. G. Comparison of group risk taking. *Journal of Experimental Social Psychology,* 1967, *3,* 189–205.

Zanna, M. P., Goethals, G. R., & Hill, J. F. Evaluating a sex-related ability: Social comparison with similar others and standard setters. *Journal of Experimental Social Psychology,* 1975, *11,* 86–93.

Received July 26, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 9, 1477-1486

Some Child-Rearing Antecedents of Criminal Behavior in Adult Men

Joan McCord
Department of Psychology and Sociology, Drexel University

Records collected during childhood and coded prior to knowledge of adult behavior provided information about the childhood homes of 201 men. Thirty years later, information about criminal behavior was collected from court records. Multiple regression and discriminant function analyses indicate that six variables describing family atmosphere during childhood—mother's self-confidence, father's deviance, parental aggressiveness, maternal affection, parental conflict, and supervision—have an important impact on subsequent behavior.

Despite a massive literature emphasizing the importance of child rearing, conscientious critics (e.g., Clarke & Clarke, 1976; Yarrow, Campbell, & Burton, 1968) have raised legitimate doubts regarding the impact of parental behavior on personality development. Many of the studies that link parental behavior with personality development rely upon a single source of information for both sets of variables; systematic reporting biases could thus cause obtained relationships. Most of the remaining studies have depended upon concurrent measurements, leaving doubt as to the direction of influence between parents' behavior and characteristics of the child. Questions about interpreting the results of both types of studies serve to highlight the importance of longitudinal research.

A few researchers have gathered information through longitudinal studies, using independent sources for measuring child rearing and for measuring personality. Robins (1966) analyzed information from clinic records gathered during childhood and related that information to data gathered when the subjects were adults. Robins pioneered assessment of long-term effects of child rearing, and her study raises doubts about the validity of retrospective reports on family socialization. Nevertheless, since predictor models combined variables describing child rearing with other types of variables (empirically linked with outcome), the research fails to provide convincing evidence that child-rearing differences affected adult behavior.

Block (1971) evaluated character development among subjects in the Berkeley longitudinal studies. Dividing 63 subjects into five types and checking differences in their backgrounds, Block reached the conclusion: "What comes through, for both sexes and without exception in viewing the various types, is an unequivocal relationship between the family atmosphere in which a child grew up and his later character structure" (p. 258). Although Block reports many statistically reliable differences, his analyses do not permit the reader to evaluate the strength of relationships between family atmosphere and character structure.

In 1973-1974, Werner and Smith (1977) retraced 88% of the children born on Kauai Island in 1955. Interviews with the mothers provided evidence about the family environ-

This study was supported by U.S. Public Health Service Research Grant 2 RO1 MH26779, National Institute of Mental Health (Center for Studies of Crime and Delinquency). It was conducted jointly with the Department of Probation of the Commonwealth of Massachusetts. The author wishes to express appreciation to the Division of Criminal Justice Services of the State of New York and to the Maine State Bureau of Identification for supplemental data from criminal records, though they are responsible neither for the statistical analyses nor for the conclusions drawn from this research.

Requests for reprints should be sent to Joan McCord, Department of Psychology and Sociology, Drexel University, Philadelphia, Pennsylvania 19104.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3709-1477$00.75

ment of subjects when they were newborn infants, age 2, and age 10. Although combined measures tended to account for a relatively high proportion of variance in several problem areas, the authors did not assess specific child-rearing models as predictors of outcome behavior.

Lefkowitz, Eron, Walder, and Huesmann (1977) used a main effects model in stepwise multiple regression for their longitudinal study of aggression. Only two of the six variables that together accounted for about a quarter of the variance in male aggressiveness at age 18 were related to child rearing at age 8. Since the model included both redundant measures of child rearing and heterogeneous variables (e.g., child's preference for girls' games, parents' religiosity, and ethnicity of family), effects of differences in child rearing may have been masked by collinearity (Blalock, 1963; Gordon, 1968; Mosteller & Tukey, 1977).

The paucity of evidence to support a view that child rearing affects personality has led some authors (e.g., Clinard, 1974; Jessor & Jessor, 1977) to the conclusion that home atmosphere during childhood has a negligible effect upon personality development. Such authors present the challenge to which the present research is addressed: if parental behavior has an important impact upon personality development, differences in child rearing ought to contribute to variations in subsequent behavior.

Method

Subjects for this study were selected from a treatment program designed to prevent delinquency. The youths ranged in age, at the time of their introduction to the program, from 5 to 13 ($M = 10.5$, $SD = 1.6$).

Counselors visited 253 boys twice a month, for an average 5-year period between 1939 and 1945. With the exception of one who was a nurse, the counselors had been trained as social workers. After each visit, the counselor recorded observations about the family as well as the child.[1]

Case records from the treatment program described, in detail, whatever activities the counselors had observed on their visits to the homes. The records included reports of conversations with parents, friends, neighbors, and teachers as well as with the boys. Counselor turnover (a potential problem from a treatment perspective) produced a benefit for research: most of the families were visited by more than one counselor.

To justify treatment of family backgrounds as independent units for analyses, only one subject from a family was included. Boys not reared by their natural mothers were also excluded. After eliminating brothers ($n = 21$) and those not reared by their natural mothers ($n = 36$), 201 cases remained for analysis.[2]

In 1957, coders read each case thoroughly in order to form judgments about the home and family interaction. These coders had no access to information about the subjects other than that contained in treatment records. A 10% random sample of the records was read independently by a second coder to yield an estimate of the reliability of the coding.[3] Variables from the coded case records were used in the present study.

The mother's attitude toward her son had been classified as actively affectionate (if there had been considerable interaction, without continual criticism, between mother and child, $n = 95$), passively affectionate (if there had been little interaction between mother and son, though the mother had shown concern for her child's welfare, $n = 51$), ambivalent or passively rejecting (if there had been marked alternations in the mother's attitude toward her son so that she had seemed sometimes to be actively affectionate and sometimes rejecting, or if the mother had seemed unconcerned about the child's welfare, $n = 43$), or actively rejecting (if the mother had appeared to be constantly critical of the boy, $n = 11$). Independent reading of 25 cases resulted in the same ratings for 80%.

Two ratings from the original codes were combined to evaluate effects of supervision. One described whether the child's activities outside of school were governed by an adult. This scale was divided to indicate whether supervision was generally present, occasionally present, or absent. Independent coding yielded identical ratings for 84% of the 25 randomly selected cases. The second rating described parental expectations regarding the boys' activities. Coders were instructed to rate expectations as "high" if the

---

[1] The project included a matched control group. Since records on family life, for the control group, were limited to information gathered during the intake interviews supplemented by information from secondary sources, the control group was not used in this study. Originally, 325 boys were included in both the treatment group and the control group. By January 1942, 253 boys remained in each group. (See Powers and Witmer, 1951, for details regarding selection of cases and a description of the treatment program.)

[2] The criteria are not mutually exclusive. Five men were eliminated through both of the selection criteria.

[3] See McCord and McCord (1960) for a complete description of the coding.

child was given responsibility for care of his younger siblings, for preparation of meals, for contributing to the financial support of the family, or for doing "extremely well" in school. Independent ratings yielded agreement for 76% of 25 cases. The scales from the 1957 codes were combined to classify subjects into one of four categories: supervision generally present and high expectations for the child to accept responsibilities ($n = 40$), supervision generally present without evidence that high expectations were placed on the child ($n = 78$), occasional supervision ($n = 60$), and supervision absent ($n = 23$).[4]

A rating of parental conflict was based on counselors' reports of disagreements between the parents. Raters were instructed to look for conflicts about the child and conflicts about values, about money, about alcohol, and about religion. Parental conflict was coded into one of four categories: no indication, apparently none, some, or considerable. For the present research, cases were divided into those whose parents evidenced considerable conflict ($n = 68$) and those coded in alternative categories. Independent readers agreed, for this division, on 80% of the cases checked for reliability.

Three measures from the 1957 codes were combined to identify aggressive parents. Coders classified the aggressiveness of each parent by looking for evidence that the parent "used little restraint" when angry. Case records included reports on parents who threw things (e.g., one father threw a refrigerator down the stairs in the midst of an argument with his wife), hit people, broke windows, and shouted abuses. Independent coders agreed on 84% of the fathers and 92% of the mothers in classifying parents as aggressive. The coders described paternal discipline; the category "consistently punitive" identified fathers who regularly used physical force (e.g., beating a child) or very harsh verbal abuse. Independent coders agreed on 92% of the cases for ratings regarding this classification. If a parent was coded as aggressive or the father was coded as consistently punitive, the child was classified as having an aggressive parent ($n = 75$).[5]

The 1957 codes included a measure of the mother's self-confidence. A rating as self-confident was assigned if that mother showed signs of believing in her own abilities ($n = 55$). Other possibilities for this rating were "no indication of general attitude; evidence that mother saw herself as a victim or pawn in a world about which she could do nothing; and neutral, that is, generally seemed merely to accept things as they came." For this variable, independent raters agreed in classifying 84% of the 25 cases used to estimate reliability.

In 1957, coders rated a father as alcoholic if the case record indicated that he had lost jobs because of repeated drinking, if marital problems were attributed primarily to his excessive drinking, if welfare agencies had repeatedly pointed to the father's drinking as grounds for family problems, or if the father had received treatment specifically for alcoholism. Independent coders agreed for 96% of the ratings on this variable. In 1948, after termination of the treatment program, criminal records on the family members of subjects were collected; these records were locked in a file separate from the case histories. In 1975, after names had been replaced by numerical identifiers, an assistant unfamiliar with the case records coded these criminal records. For the present study, a father was considered "deviant" if the case record indicated that he was an alcoholic, if the criminal record showed that he had been convicted at least three times for drunkenness, or if his criminal record showed that he had been convicted for a serious crime (i.e., theft, burglary, assault, rape, attempted murder, or murder). These criteria led to identification of 86 fathers as deviant.[6]

Case records included information about family structure. A father was considered "absent" if his residence was not with the subject's mother. Independent coding of 25 cases yielded agreement on 96% regarding whether or not the boy was living with both natural parents.[7] The 71 boys having absent fathers ranged in age, at the time when the loss occurred, from birth to 16 ($M = 7.01$, $SD = 5.03$). The father-absent subjects were subclassified to identify those whose natural fathers had been present during their first 5 years ($n = 48$) and those for whom the absence had occurred prior to the age of 5 ($n = 23$).

These seven variables (mother's affection, supervision, parental conflict, parental aggression, mother's self-confidence, father's deviance, and paternal absence) were used to depict the home atmosphere of subjects during childhood. The first three are regarded as directly related to child rearing. Relationships among these measures are shown in Table 1.

Subjects had been selected from congested urban neighborhoods. Nevertheless, differences in social status could contribute to subsequent differences in behavior. Two measures of social status were available. The case records supplied information about the father's occupation. Coders classified these occupations as white-collar (9.6%), skilled tradesmen (32.8%), or unskilled workers (57.6%). The reliability check yielded agreement on 96% of the ratings. A second measure of social status was provided by a rating of the neighborhoods in which the boys were raised. These ratings had been made, in 1938 and 1939, as part of the selection procedures. The ratings took into account delinquency rates, availability of recreational facilities, and proximity to bars, railroads, and junkyards. These ratings were coded on a 4-point scale from better to worst neighborhoods. The two measures tended to covary, Cramer's $V(6) = .218$, $p = .0044$.

---

[4] Only nine boys exposed to high expectations had not been rated as generally supervised.

[5] Fifty of the subjects were classified as having aggressive parents by the direct description of parental aggression.

[6] Forty-nine had been convicted for serious crimes.

[7] Of the 86 deviant fathers, 40 were also absent fathers.

Table 1
Relationships Among Variables Describing Home Atmosphere (Cramer's V)

	Supervision	Parent conflict	Parent aggression	Mother's self-confidence	Father's deviance	Father's absence
Mother's affection	.241***	.209*	.184*	.199*	.209*	.110
Supervision		.267**	.106	.308***	.230*	.187*
Parent conflict			.188**	.109	.381***	.375***
Parent aggression				.289***	.144*	.024
Mother's self-confidence					.125	.087
Father's deviance						.206*

* $p < .05$. ** $p < .01$. *** $p < .001$.

Between 1975 and 1978, the subjects were retraced. Among the 201 men included in the study, 153 (76%) were alive and in Massachusetts at least until the age of 40;[8] 16 (8%) had died prior to their 40th birthdays; 29 (14%) had migrated from Massachusetts; and 3 (1%) remained to be found.

During 1975 and 1976, the names (and pseudonyms) of all the men who had been in the program were checked through court records in Massachusetts.[9] These criminal records were traced and coded by different people from those who coded other records. Coders of the criminal records (and those who traced them) had no access to other information about the subjects. The court records showed the dates of court appearances and the crimes for which the subjects had been convicted. They were coded to show the type of crime and the age of the person when he was convicted. Convictions for serious property crimes (larceny, auto theft, breaking and entering, arson) and serious personal crimes (assault, attempted rape, rape, attempted murder, kidnapping, and murder) were used as dependent measures for this study.

Among the 201 men, 71 had been convicted for at least one serious crime; 53 had been convicted for property crimes and 34 for personal crimes (including 15 convicted for both types). Their ages when first convicted ranged from 8 to 38, with a mean of 18.7 ($SD = 8.7$) and a median of 20. Those convicted prior to their 18th birthdays were classified as juvenile delinquents ($n = 43$); those convicted after reaching the age of 18 were classified as adult criminals ($n = 48$, including 20 who had been juvenile delinquents).

After analyzing the relationship to crime of each of the childhood variables separately, multiple regression analyses (General Linear Model Procedure, Barr, Goodnight, Sall, & Helwig, 1976) were used to ascertain the contribution of child rearing to the variance in number of serious property and personal crimes. To test the degree to which knowledge of home atmosphere could enable accurate prediction of subsequent behavior, the six central variables describing home atmosphere were used in discriminant function analyses to identify criminals.

As a more stringent test of the contribution of home atmosphere to subsequent crime, the discriminant function analyses were also used to predict criminals among the subsample whose criminal records provided the most complete histories of convictions: those men living in Massachusetts at least until the age of 40. If this function identified criminality more accurately for the total group than for those living in Massachusetts, there would be grounds for suspecting an interaction effect between home background and unmeasured variables. If this function identified criminality at least as accurately for those alive in Massachusetts at the age of 40, there would be additional support for a conclusion that home atmosphere during childhood contributes to criminality.

Results

As a first step toward learning whether parental behavior contributes to subsequent differences in criminality, the seven scales describing home atmosphere and the two scales describing social status were individually analyzed for their contributions to the variance in number of serious crimes against property and persons. (See Table 2.)

With the exception of father's absence, each of the scales describing home atmosphere accounted for a statistically significant ($p < .05$) proportion of the variance in number of crimes against property, persons, or both.[10]

[8] Among them, 147 were in Massachusetts through their 45th birthdays.

[9] These records were supplemented by court records from the states of New York, Maine, Michigan, Nebraska, and Florida, where some of the men had resided.

[10] The Duncan multiple range test, modified for unequal groups (Kramer, 1956), indicated that boys without supervision, reliably ($p < .05$) more than boys in the other three categories, were convicted

Table 2
Relationships Between Variables Describing Home Background and Crimes

		Property crimes			Personal crimes		
	DF	R^2	*F*	*PR > F*	R^2	*F*	*PR > F*
Mother's affection	3, 196	.092	6.60	.0003	.029	1.92	.1261
Supervision	3, 197	.152	11.77	.0001	.071	5.04	.0023
Parent conflict	1, 199	.008	1.76	.1866	.035	7.14	.0081
Parent aggression	1, 199	.012	2.38	.1242	.036	7.36	.0073
Mother's self-confidence	1, 199	.022	4.50	.0350	.024	4.98	.0268
Father's deviance	1, 199	.024	4.81	.0295	.000	0.00	.9569
Father's absence	2, 198	.005	0.47	.6286	.026	2.67	.0715
Neighborhood	3, 197	.028	1.87	.1335	.016	1.10	.3513
Father's occupation	2, 195	.008	0.74	.4798	.012	1.15	.3181

Neither of the measures of social status was significantly related to these types of crimes.

Supervision and mother's self-confidence were related to both crimes against property and crimes against persons; mother's affection and father's deviance were related to property crimes (though not to personal crimes); conflict and parental aggression were related to personal crimes (though not to property crimes). The boys who lacked maternal affection, who lacked supervision, whose mothers lacked self-confidence, and whose fathers were deviant were more often subsequently convicted for property crimes. The boys who lacked supervision, whose mothers lacked self-confidence, and who had been exposed to parental conflict and to aggression were subsequently more often convicted for personal crimes.

The relationships to criminality of individual variables describing home atmosphere, though statistically significant, each accounted for a relatively small proportion of the variance. More important, since "criminogenic" conditions tended to be related to one another, these relationships could not be taken as evidence that the differences in home background that they represented resulted in differences in subsequent behavior.

To evaluate the contribution of parental behavior to subsequent behavior, the six central variables describing home atmosphere were divided into two sets. The first set included those variables that described characteristics of the parents, characteristics that might be viewed as antecedent to child-rearing practices: parental aggressiveness, paternal deviance, and mother's self-confidence. The second set included the three variables that described interpersonal behavior: parental conflict, supervision, and mother's affection; these were considered to be direct measures of child rearing. The effect of this division was to classify families in two ways. The first classification took account of relationships among the variables describing the parents; the second took account of relationships among the variables describing child rearing.

Sequential multiple regression models were used. They introduced the measure of social status (the interaction of father's occupation and neighborhood) as the first variable. The regression procedure next evaluated the sequential contribution to explained variance of parental characteristics (the interaction of paternal deviance, maternal self-confidence, and parental aggression). After controlling effects of both social status and parental characteristics, the procedure evaluated effects of child rearing (i.e., the interaction of supervision, parental conflict, and the mother's affection).

Child rearing, as measured in this longitudinal study, clearly accounts for a signif-

for both property and personal crimes. This a posteriori test showed that boys rejected by their mothers were most likely to be convicted for property crimes, and boys who had affectionate mothers were least likely.

Table 3
Home Environment and Subsequent Criminality

Sequential contribution	*DF*	R^2	*F*	*PR > F*
Predicting property crimes[a]				
Social status	11	.0688	1.55	.1185
Parental characteristics	7	.0610	2.16	.0404
Child rearing	27	.2612	2.40	.0005
Predicting personal crimes[b]				
Social status	11	.0541	1.16	.3222
Parental characteristics	7	.0588	1.97	.0616
Child rearing	27	.2444	2.13	.0023
Predicting total crimes[c]				
Social status	11	.0620	1.40	.0620
Parental characteristics	7	.0691	2.45	.0209
Child rearing	27	.2601	2.39	.0005

[a] Model: $R^2 = .391$, $F(45, 151) = 2.15$, $p = .0003$. [b] Model: $R^2 = .3573$, $F(45, 151) = 1.87$, $p = .0028$. [c] Model: $R^2 = .3912$, $F(45, 151) = 2.16$, $p = .0003$.

icant proportion of the variance in subsequent criminality. Table 3 describes the decomposition of the regression models.

As predictors of property crimes, the model accounts for 39.1% of the variance, $F(45, 151) = 2.15$, $p = .0003$. Parental aggression, paternal deviance, and maternal self-confidence account for 6.1% of the variance after controlling social status, $F(7, 151) = 2.16$, $p = .0404$. Parental conflict, supervision, and maternal affection contribute significantly to the variance after effects of social status and parental characteristics have been controlled, $R^2 = .261$, $F(27, 151) = 2.40$, $p = .0003$.

As predictors of personal crimes, the model accounts for 35.7% of the variance, $F(45, 151) = 1.87$, $p = .0028$. The three more direct measures of child rearing contribute significantly to the variance after effects of social status and parental characteristics have been removed, $R^2 = .244$, $F(27, 151) = 2.13$, $p = .0023$.

As predictors of the total number of serious crimes for which the men had been convicted, the model accounts for 39.1% of the variance, $F(45, 151) = 2.16$, $p = .0003$. Parental characteristics account for 5.9% of the variance after effects of social status have been controlled, $F(7, 151) = 2.45$, $p = .0209$. The child-rearing variables account for 26.0% of the variance, $F(27, 151) = 2.39$, $p = .0005$, after removing effects of both social status and parent characteristics.

Adding information about whether or not a man had been reared in a home marked by paternal absence did not reliably increase the accuracy of any of the predictions.[11]

Within the (relatively restricted) range of social class represented in the study, the contribution of social status to the variance in crimes was not statistically reliable. On the other hand, both parental characteristics and child-rearing practices were reliably related to the number of crimes for which the subjects had been convicted.[12]

Approximately a third of the 200 men coded on all six variables describing home atmosphere had been convicted for at least one serious crime. A discriminant function based on the variables describing home atmosphere for these 200 men correctly identified 147

[11] R^2 was increased by .002 for property crimes, .007 for personal crimes, and .001 for total crimes.

[12] Without controlling for social status, parental characteristics and child-rearing variables accounted for 36.7% of the variance in property crimes, $F(34, 162) = 2.76$, $p \leq .0001$, 30.8% of the variance in personal crimes, $F(34, 162) = 2.12$, $p = .0010$, and 36.3% of the variance in total number of serious crimes, $F(34, 162) = 2.71$, $p \leq .0001$.

Table 4
Results of Discriminant Function Analyses

Dependent and independent variables	Correct as criminals		Correct as noncriminals		Overall accuracy		% improvement over chance	$z > p$
	n	%	*n*	%	*n*	%		
All subjects								
Ever criminal								
Home atmosphere	71	67.6	129	76.7	200	73.5	19.3	.0001
Adult criminal								
Home atmosphere	48	56.3	152	87.5	200	80.0	16.5	.0001
Men living in Massachusetts through the age of 40								
Ever criminal								
Home atmosphere	60	81.7	92	70.7	152	75.0	22.8	.0001
Adult criminal								
Home atmosphere	42	71.4	110	84.6	152	80.9	20.9	.0001
Juvenile delinquency record	42	45.2	110	83.6	152	73.0	13.0	.0011

(73.5%) as criminals or noncriminals; random predictions based on prior probabilities would be expected to identify only 54.2% correctly,[13] $z = 5.48$, $p < .0001$. The function based on parental aggression, maternal self-confidence, paternal deviance, supervision, maternal affection, and parental conflict correctly identified 76.7% of the noncriminals and 67.6% of the criminals. (See Table 4.)

Forty-eight men had been convicted for serious crimes after their 18th birthdays. The six variables describing home atmosphere provided a discriminant function which correctly identified 27 (56.3%) of the 48 adult criminals and 133 (87.5%) of the 152 men without records for convictions as adults. Predictions based on the descriptions of home atmosphere provided a 16.5% improvement over the 63.5% expected from random predictions based on prior probabilities, $z = 4.85$, $p < .0001$. (See Table 4.)

After discarding men who had died before the age of 40, migrated from Massachusetts, or who had not yet been found, 152 men who were living in Massachusetts at least until the age of 40 and whose case records had been coded for all six variables describing home atmosphere remained for discriminant function analyses. Among these men, the discriminant function correctly identified 75.0%, a slight improvement over the rate of correct identification among the total group of men and a 22.8% improvement over random procedures based on prior probabilities, $z = 5.63$, $p < .0001$. (See Table 4.) This discriminant function correctly identified 81.7% of the 60 criminals and 70.7% of the 92 noncriminals.

A breakdown of the results shows that the discriminant function had correctly classified as criminal 78.1% of those convicted only for property crimes, 78.6% of those convicted only for personal crimes, and 92.9% of those

[13] The model used to estimate predictions based on chance assumes that the number of predictions as criminal would be proportional to the actual distribution of criminals among subjects. Alternative models that might be considered range from assuming that each individual is as likely to be convicted as not (which would result in an expectation for correct predictions among half the noncriminals and half the criminals) to assuming that all or no individuals would be convicted. Although a "rational bet" would maximize correct predictions by predicting that all individuals would fall into the larger class, this model is inappropriate when the interest is in correct identification of those in the smaller class. An equiprobability model for the discriminant function analysis based on family atmosphere resulted in correct sorting of 68% of the men (62% of the noncriminals and 79% of the criminals) in terms of whether or not they had been convicted for serious crimes.

convicted for both property and personal crimes.

Among the 60 men convicted for serious crimes and still living in Massachusetts at the age of 40, 18 (30.0%) had been convicted only as juveniles, 23 (38.7%) had first been convicted after the age of 18, and 19 (31.7%) had been convicted both as juvenile and as adults. Were one to predict that only and all juvenile delinquents would be convicted as adults, the prediction would be correct for 73.0%, an improvement over an expectation of 60.0% from random procedures based on prior probabilities, $z = 3.27$, $p = .0011$. This prediction would be correct for 45.2% of the adult criminals and for 83.6% of the men not convicted as adults. Predictions based on juvenile records would, of course, be right for *none* of the men first convicted as adults (54.8% of the adult criminals) and for only 51.4% of the juvenile delinquents.

Among men living in Massachusetts at the age of 40, the discriminant function analysis based on home atmosphere during childhood correctly identified 80.9% as criminal or noncriminal after the age of 18. (See Table 4.) Use of the variables describing home atmosphere during childhood resulted in a 20.9% improvement over chance identification, $z = 5.26$, $p < .0001$, and a 7.9% improvement over predictions based on the subjects' juvenile criminal histories, $z = 2.19$, $p = .0282$. This function correctly identified 71.4% of the adult criminals and 84.6% of the noncriminals. The discriminant function based on home atmosphere correctly identified as criminals 65.2% of the men who had first been convicted as adults. In terms of their subsequent criminal records, this discriminant function correctly sorted 78.4% of the juvenile delinquents and 81.7% of those who had not been juvenile delinquents.[14]

Summary and Discussion

Recent criticism of the assumption that child-rearing practices have an important impact on personality development posed the issue addressed in this research. In order to evaluate the assumption, records describing home atmosphere during childhood, recorded during childhood, were linked with records of subsequent criminality, gathered when the subjects were middle-aged. The two sources of information were independent: data collection had been separated by several decades, the data had been coded by different people, and the coders had no access to information other than that which they were coding. Therefore, measures of home atmosphere were uncontaminated by retrospective biases and measures of subsequent behavior were uncontaminated by knowledge of home background.

Records describing home atmosphere had been written between 1939 and 1945. These records were case reports of counselors' repeated home visits to the 201 boys included in this study. The case records were coded, in 1957, to provide descriptions of home atmosphere.

Information about criminal behavior was gathered from court records, 30 years after termination of the program from which descriptions of home atmosphere had been collected. Subjects were considered criminals if they had been convicted for serious crimes (those indexed by the Federal Bureau of Investigation).

In preliminary analyses, six of seven variables describing home atmosphere were reliably related to criminal behavior. Only father's absence failed to distinguish criminals from noncriminals. Considering the emphasis given to broken homes as a source of subsequent criminality (e.g., Bacon, Child, & Barry, 1963; Glueck & Glueck, 1951; Wadsworth, 1979; Willie, 1967), this finding is worthy of note.

Multiple regression analyses indicated that six variables describing home atmosphere in childhood account for a significant proportion of the variance in number of convictions for serious crimes. After controlling effects of differences in social status, parental characteristics and child-rearing variables ac-

[14] The "rational bet" that men not convicted as juveniles would not be convicted as adults would be correct for 80% of the men not convicted as juveniles. Since this bet would be correct only for the noncriminals, the prediction would fail to identify correctly any of the critical group: men first convicted as adults.

counted for 32.2% of the variance in number of convictions for property crimes and 30.3% of the variance in number of convictions for personal crimes. The three most direct measures of child rearing (supervision, mother's affection, and parental conflict) accounted for approximately a quarter of the variance in number of convictions for serious crimes—after effects of both social status and parental characteristics had been removed.

Discriminant function analyses based on the six variables describing home atmosphere correctly identified 73.5% of the men as either subsequently criminal or noncriminal; further, these six variables provided a function that for 80% of the men correctly discriminated between those convicted and those not convicted for serious crimes as adults.

As compared with analyses for the total sample, the discriminant function analyses were (slightly) more accurate when used to predict behavior among the men whose criminal records provided the most complete histories of convictions. Among men living in Massachusetts at least to the age of 40, these functions correctly identified 75% as ever criminal or as noncriminal and 80.9% as criminal or noncriminal after the age of 18. Limiting analyses to men living in Massachusetts controlled any differences contributing to migration or early death; therefore, the accuracy of discriminant functions among this subsample is interpreted as supporting the view that home atmosphere during childhood contributes to criminality.

When used to identify men convicted as adults, the discriminant function identified as criminals almost two-thirds of the men first convicted after the age of 18. This function also correctly sorted more than three-quarters of the juvenile delinquents, distinguishing between those who were and those who were not adult criminals.

Although the discriminant functions based on home atmosphere were surprisingly successful in identifying men who were to become criminals, it would be a mistake to conclude that the longitudinal design of this research has led to recognition of the causes of crime. This research is limited not only by its subjects (all of whom were reared in congested urban areas during the thirties and early forties) but also by the hypotheses considered. In this research, parental aggression, paternal deviance, maternal self-confidence, supervision, mother's affection, and parental conflict indexed home atmosphere; unconsidered variables might better describe the features in the child's home that affect his behavior. In this research, the possibly confounding variable of social status was considered; other conditions might account for the apparent link between home atmosphere and crime. Nevertheless, the evidence from this study suggests that parental behavior does have an important impact on subsequent behavior: predictions of adult criminality based on knowledge of home atmosphere were not only markedly more accurate than chance—they were also more accurate than predictions based on the individuals' juvenile criminal records.

References

Bacon, M. K., Child, I. L., & Barry, H. A Cross-cultural study of correlates of crime. *Journal of Abnormal and Social Psychology*, 1963, *66*, 291–300.

Barr, A. J., Goodnight, J. H., Sall, J. P., & Helwig, J. T. *A user's guide to SAS 76*. Raleigh, N.C.: SAS Institute, 1976.

Blalock, H. M. Correlated independent variables: The problem of multicollinearity. *Social Forces*, 1963, *42*, 233–237.

Block, J. *Lives through time*. Berkeley, Calif.: Bancroft Books, 1971.

Clarke, A. M., & Clarke, A. D. B. *Early experience: Myth and evidence*. New York: Free Press, 1976.

Clinard, M. B. *Sociology of deviant behavior*. (4th ed.) New York: Holt, Rinehart & Winston, 1974.

Glueck, S., & Glueck, E. *Unraveling juvenile delinquency*. Cambridge, Mass.: Harvard University Press, 1951.

Gordon, R. A. Issues in multiple regression. *American Journal of Sociology*, 1968, *73*, 592–616.

Jessor, R., & Jessor, S. L. *Problem behavior and psychosocial development*. New York: Academic Press, 1977.

Kramer, C. Y. Extension of multiple range tests to group means with unequal numbers of replication. *Biometrics*, 1956, *12*, 307–310.

Lefkowitz, M. M., Eron, L. D., Walder, L. O., & Huesmann, L. R. *Growing up to be violent: A longitudinal study of aggression*. New York: Pergamon Press, 1977.

McCord, W., & McCord, J. *Origins of alcoholism*. Stanford, Calif.: Stanford University Press, 1960.

Mosteller, F., & Tukey, J. W. *Data analysis and regression.* Reading, Mass.: Addison-Wesley, 1977.

Powers, E., & Witmer, H. *An experiment in the prevention of delinquency: The Cambridge-Somerville youth study.* New York: Columbia University Press, 1951.

Robins, L. N. *Deviant children grown up.* Baltimore, Md.: Williams & Wilkins, 1966.

Wadsworth, M. E. J. *Roots of delinquency.* New York: Barnes & Noble, 1979.

Werner, E. E., & Smith, R. S. *Kauai's children come of age.* Honolulu: University Press of Hawaii, 1977.

Willie, C. V. The relative contribution of family status and economic status to juvenile delinquency. *Social Problems,* 1967, *14,* 326–335.

Yarrow, M. R., Campbell, J. D., & Burton, R. V. *Child Rearing.* San Francisco: Jossey-Bass, 1968.

Received August 7, 1978 ■

Instructions for Submission of Manuscripts

Papers submitted to the *Journal of Personality and Social Psychology* after April 1, 1979 will be considered under new policies approved by the Publications and Communications Board. Effective with the 1980 volume of the journal, the following policies will apply:

The *Journal of Personality and Social Psychology* publishes original papers in all areas of personality and social psychology. It emphasizes empirical reports but may include specialized theoretical, methodological, and review papers. Beginning in April 1980, the journal will be divided into 3 independently edited sections: ATTITUDES AND SOCIAL COGNITION, INTERPERSONAL RELATIONS AND GROUP PROCESSES, and PERSONALITY PROCESSES AND INDIVIDUAL DIFFERENCES. Authors should submit manuscripts to the appropriate section editor according to the section definitions and manuscript preparation instructions provided on page 468 of the March 1979 issue of the journal.

Section editors' addresses are as follows: ATTITUDES AND SOCIAL COGNITION, Melvin Manis, Editor, Research Center for Group Dynamics, Institute for Social Research, University of Michigan, Ann Arbor, Michigan 48106; INTERPERSONAL RELATIONS AND GROUP PROCESSES, Ivan D. Steiner, Editor, Department of Psychology, University of Massachusetts, Amherst, Massachusetts 01003; PERSONALITY PROCESSES AND INDIVIDUAL DIFFERENCES, Robert Hogan, Editor, Johns Hopkins University, Department of Psychology, 33rd and Charles Streets, Baltimore, Maryland 21218.

Journal of Personality and Social Psychology
1979, Vol. 37, No. 9, 1487–1504

Are Causal Attributions Causal? A Path Analysis of the Cognitive Model of Achievement Motivation

Martin V. Covington and Carol L. Omelich
University of California, Berkeley

The causal assumptions of Weiner's cognitive reinterpretation of the traditional theory of achievement motivation are tested. This research asks the question, "Do achievement motivation groups differ in their affective reactions, expectancy of future success, and subsequent test performance as a consequence of attributions made for a previous test outcome?" Subjects ($N = 206$) were college students varying in resultant achievement motivation who experienced feelings of failure in a course test and chose to take the exam a second time under a mastery learning system. After feedback on first test performance, students made attributions for their initial failure, indicated degree of shame, and rated their expectancy for success on the second test opportunity. Treatment of this system of nonmanipulated variables by path analysis techniques provides little support for the contention that variations in expectancy and retest performance depend on attributions made for a previous failure. Affect depends in part on internal attributions, but in a direction opposite to predictions. An alternative interpretation of the role of cognitive attributions in the achievement process is explored.

The most widely known theory of achievement motivation (Atkinson, 1957, 1964) holds that achievement behavior is the result of an emotional conflict between hope of success and fear of failure. Recently Weiner and his colleagues (Weiner, 1972, 1974, 1977; Weiner, Frieze, Kukla, Reed, Rest, & Rosenbaum, 1971) have suggested a cognitive reinterpretation of this traditional theory. It is proposed that need for achievement (nAch) is mediated by perceptions of causality that in turn influence affective reactions to success and failure, expectancy of future success, and subsequent performance. Thus achievement motivation, originally conceived solely in terms of differential emotional anticipations, becomes heavily imbued with cognitive elements. In this new schema nAch is assigned a subordinate status as one of many antecedents to achievement, with cognitive attributions becoming the major causal determinants of achievement behavior.

Following Heider (1958), Weiner proposes ability, effort, luck, and task difficulty as among the major perceived causes of achievement performance (Weiner, 1972, 1977, 1979). There is general consensus that high and low nAch individuals harbor differential explanations about the causes of their successes and failures (Weiner et al., 1971; Weiner, Heckhausen, Meyer, & Cook, 1972; Weiner & Kukla, 1970; Weiner & Potepan, 1970), and it is precisely these differences that are taken as the essence of individual differences in achievement motivation. Generally speaking, persons motivated to approach success (high nAch) attribute failure to lack of effort and success to their ability, whereas failure-avoiding (low nAch) persons tend to ascribe failure to lack of ability and success to external factors such as luck. The

The authors wish to thank Richard S. Lazarus, Joy Stapp, Alan Zonderman, the action editor, and anonymous reviewers for their helpful comments and suggestions on an earlier draft of this article.

Requests for reprints should be sent to Martin V. Covington, Department of Psychology, University of California, Berkeley, California 94720.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3709-1487$00.75

cognitive model specifies several sequential steps in the achievement process resembling an information-processing system.

1. First, following a given achievement outcome, success or failure is interpreted retrospectively in terms of the four major attributional elements and according to the predisposing individual differences in nAch.

2. Second, these differential perceptions of causality, in turn, mediate affective reactions to performance—shame in failure and pride in success—as well as the individual's expectations about the probability of future success. For example, in failure, self-ascriptions to low effort—characteristic of high nAch individuals—intensify shame and act to maintain an expectation for future success. By contrast, low ability ascriptions lead to less shame but lowered future expectations, a pattern typical of low nAch persons.

3. As a final step, achievement performance is determined by causal ascriptions acting indirectly through both pathways of affective reaction and expectations of future success. In effect, high nAch persons who have reacted to failure with chagrin yet have maintained an undiminished optimism for the future (due to their attributional biases) are likely to perform far better on the next occasion than are low nAch individuals. The entire process from Step 1 through Step 3 is repeated from one performance to the next.

This ingenious cognitive analysis has stimulated a large volume of research (for reviews see Weiner, 1972, 1974, 1977). Yet despite the impressive numbers, several shortcomings limit the theoretical utility of the cognitive viewpoint. First, with but rare exception (e.g., Weiner & Potepan, 1970), research has been carried out in the laboratory rather than in naturally occurring settings. Second, investigators have typically confined themselves to an examination of one link in the model at a time, to the exclusion of other antecedent and subsequent relationships. While a few studies have incorporated additional variables (e.g., Meyer, 1970 [fully described in Weiner et al., 1972]; Weiner & Sierad, 1975), none have yet evaluated the entire model as a dynamic system in a real-life setting. As a result, the central assumption of the primary causal role of attributions lacks sufficient empirical support to warrant its unqualified acceptance. Thus we ask: "Do achievement motive groups differ in affective reactions, expectancy of future success and subsequent performance *as a consequence* of attributions made for a previous test outcome?" At stake here is the basic assumption that attributions are in themselves a sufficient explanation for achievement behavior. According to the model, once nAch acts to trigger differential attributional biases in response to success or failure feedback (Step 1), it plays no further role as a major, direct determinant of the more distal psychological events that influence retest performance (i.e., expectancy and affect). Whether or not we are justified in deleting the concept of need achievement from theories of achievement behavior has yet to be empirically demonstrated.

The purpose of the present study is to trace the sequential cause and effect network of relationships assumed by the cognitive model in order (a) to explore both the separate and overlapping causal roles played by nAch and cognitive attributions in determining variations in achievement behaviors and (b) to determine whether or not the proposed attributional mediators act in accord with theoretical predictions. To this end we have gathered data on all the crucial elements specified by the model in an achievement context of test-taking in the classroom and in the chronological order that conforms to the theory. This analysis is confined to subjects who experienced subjective failure in the first of two test-taking opportunities.

The primary method for analyzing this system of nonmanipulated variables is path analysis (Anderson & Evans, 1974; Duncan, 1966, 1975; Heise, 1975; Kerlinger & Pedhazur, 1973; Werts & Linn, 1970). Path analysis allows for all determining factors as specified by a causal model to be incorporated into an overall predictive analysis, thereby permitting an estimation of the relative contribution (both direct and indirect) of each determinant to variations in the dependent variable(s). In this case, the dependent variables of interest are various facets of achievement behavior.

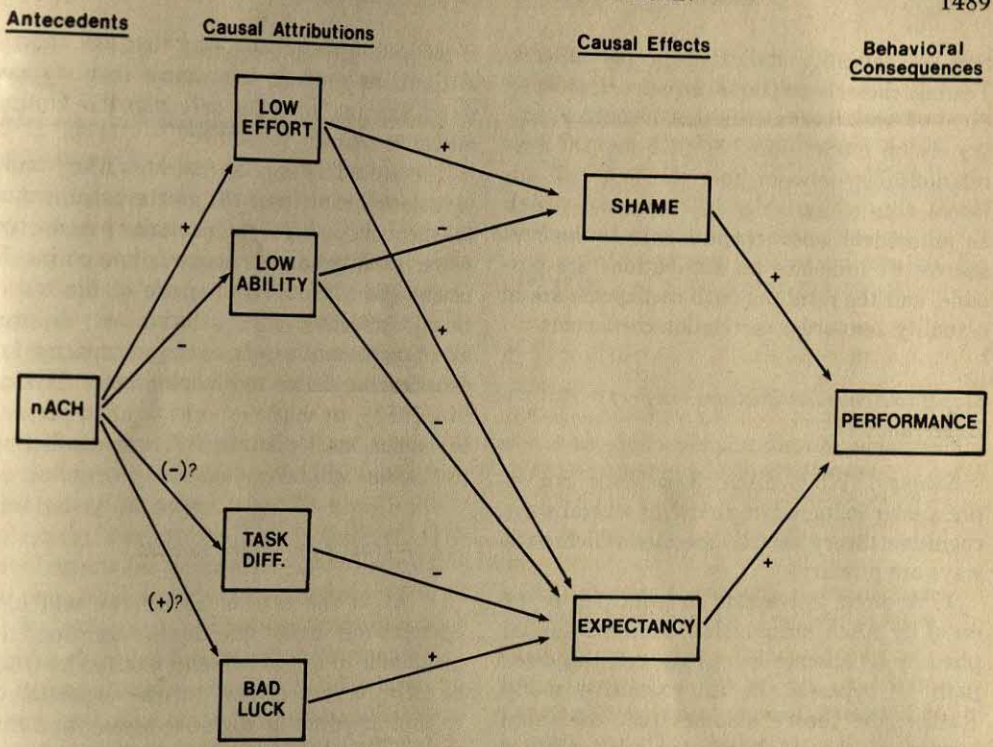


Figure 1. Causal pathways of influence postulated by the cognitive model of achievement motivation in the task reevaluation stage following failure.

For a proper interpretation of the analyses to follow, it is important to bear in mind that path analysis is not a procedure for demonstrating causality. Rather it is a method for tracing out the implications of a set of causal assumptions that the theoretician is willing to impose on a system of relationships. In essence, we attempt to answer the question "Does the cognitive model fit the data adequately?" by comparing the *observed* relationships among the variables with the *expected* relationships. Since the presumed relationships and directionality among the elements of the cognitive model have been so well specified, the application of recursive path analytic techniques does not allow for the testing of an alternate model. For instance, nAch is seen as an antecedent that predisposes individuals to particular attributional patterns. This causal ordering precludes the testing within the path analysis system of any alternate hypothesis that concerns the effect of attributions on nAch.

It will prove helpful to indicate in advance the kinds of hypotheses and specific causal assumptions we will examine, the order of the data presentation, and how path-analytic techniques will be employed. This presentation will be aided by reference to Figure 1, which schematizes the cognitive events described earlier. The unidirectional arrows and accompanying algebraic signs indicate the main pathways of causal influence to be tested.

Step 1: nAch→Attributions

Cognitive theory predicts a positive relationship between effort ascriptions and nAch in failure, that is, as nAch level increases, so do ascriptions to low effort. Similarly, we expect a negative relationship between nAch and ability ascriptions such that as nAch level increases, attributions to low ability decrease. The nAch pathways to luck and task ascriptions carry question marks, since the model makes no specific predictions regarding motive group predispositions. However, the stable (task) and unstable (luck) qualities of these two external elements suggest the direc-

tion of causality indicated in parentheses. Testing these predictions involves regressing each of the four attributional sources separately on nAch. Since the first step involves relationships between only the first two adjacent sets of variables in the model (nAch as antecedent and attributions), no *indirect* sources of influence on attributions are possible, and the resulting path coefficients are in actuality zero-order correlation coefficients.

Step 2: nAch→Attributions→Affect/Expectancy

Shame. While nAch dispositions can express their influence on shame in several ways, cognitive theory clearly specifies which pathways are primary.

1. A direct effect can, in principle, be exerted by nAch membership. However, as implied by its absence from Figure 1, this direct path is bypassed in the cognitive model. Rather, the theory assumes that the causal burden of the nAch→shame relationship is transmitted *indirectly* via attributions, with nAch serving as no more than an instigator of attributional biases. Path analysis allows for a comparison of the relative importance of these two nAch pathways.

2. The cognitive position also assumes that shame is influenced *directly* along an internal–external dimension of causal ascriptions, with internal factors being the more important determinants. While this dimensional distinction has been deemphasized in recent treatments and additional affective reactions identified (Weiner, 1979; Weiner, Russell, & Lerman, 1978, Note 1), it is still assumed that shame is a consequence of ability and effort attributions. Thus shame is said to be maximized when failure is perceived as occurring through lack of effort, especially among highly able individuals, and minimized when the individual, particularly if he is of low ability, has tried hard. This leads to the prediction of a positive relationship between effort ascriptions and shame, that is, as attributions to low effort increase, so does shame. Correspondingly, we expect a negative relationship between ability ascriptions and shame such that as low ability attributions increase shame is reduced. This is not to say that low ability attributions produce less shame than do external explanations, but only that low ability mitigates shame.

The preceding implies that high nAch individuals will experience the greatest shame due to their disposition to maintain a sense of power (ability) and to blame failure on insufficient effort. However, because of the traditional, affective interpretation of failure-avoiding persons as possessing a "capacity for experiencing shame and humilitation" (Atkinson, 1957), we might expect low nAch persons to be the most distressed. These conflicting predictions will be evaluated.

Expectancy for future success.

1. As in the case of shame, we will also address the basic question concerning the magnitude of causal influence exerted by nAch directly on expectancy (nAch→expectancy) as contrasted to the cognitive assumption that the only notable pathways of significance come indirectly from nAch mediated through attributions.

2. Expectancy is predicted to vary directly along the stable/unstable attributional axis. Specifically, expectancy is said to be higher when failure is attributed to unstable elements (effort, luck) and lower when ascribed to stable ones (ability, task). Accordingly, we ought to find significant, positive path coefficients between expectancy and ascriptions to low effort and bad luck; and conversely, negative path coefficients between expectancy and attributions to low ability and high task difficulty.

It follows from the foregoing that high nAch persons will maintain greater expectations for future success after failure than will low nAch persons, owing to the tendency of the former group to attribute failure to unstable elements, especially lack of effort. While several studies confirm that low nAch persons harbor lower expectations for success (Atkinson, Bastian, Earl, & Litwin, 1960; Kukla, 1972a, 1972b; Moulton, 1974) there is little evidence that this condition is influenced by attributional mediators. Path analysis allows us to evaluate the relative im-

portance of this presumed indirect causal pathway.

Step 3: nAch→Attributions→Affect/Expectancy→Retest

Cognitive theory holds that attributions influence performance largely through their determination of expectancy and affect.

1. To clarify the causal role of nAch, we will assess the degree to which nAch exerts a direct impact on performance in contrast to its indirect influence through cognitive mediators.

2. The assumption that affect and expectancy influence performance largely because of antecedent variations in attributions leads us to expect a relatively large indirect effect of attributions on performance as compared to the magnitude of the direct influence exerted by expectancy and shame. Correspondingly, the direct influence of attributions on performance should be negligible. This latter prediction takes on some importance in light of recent work suggesting direct attributional influence (e.g., Dweck, 1977; Latta, 1976).

As we move through the various steps of the cognitive model, the number of indirect paths of potential influence increases progressively. In principle, nAch can express its influence on performance indirectly through some 14 different indirect pathways. Although most of these paths contribute only negligibly to variations in performance, and few of the paths are of more than passing interest, we must nonetheless deal with the unwanted complexity that results. Consequently, at each of the three major steps in our analysis we will also employ a stepwise regression technique that allows for an incremental F test as a summary statistic. While not path-analytic in interpretation, it provides a valuable adjunct perspective indicating whether or not the addition of each successive predictor variable (or set of variables) increases significantly the total explained variance in the particular dependent variable under consideration.

Method

Subjects and Procedure

Subjects were drawn from among 439 students enrolled in an introductory psychology course that featured a modified mastery structure (Block, 1977, Note 2). At the end of each instructional unit two equivalent multiple-choice tests were administered two days apart. If students wished the chance to improve, they had the option of taking the second test after further study. Two hundred and six students (100 females; 106 males) who considered their first performance unsuccessful and chose to take the second test served as subjects. The present data were collected at the end of the second instructional unit.

Measures

Both tests were scored, and the results were known to students before they left the examination room. After receiving grade-equivalent feedback on the first test, students completed a brief questionnaire.

Causal ascriptions for failure. Students indicated the causal importance of each of Weiner's attributional sources using a format adapted from Feather (1969):

To what extent do you consider that your unsuccessful performance was due to each of the following factors?

The examination was difficult.
I did not study very much.
I had back luck (e.g., test stressed different materials).
I lack the skill and ability.

Each attribution was rated separately on a 7-point Likert-type scale from 1 (not a cause) to 7 (very much a cause).

Affective reactions. Students also estimated degree of shame as a result of their unsatisfactory performances. Again, ratings were made on a 7-point scale from 1 (not at all) to 7 (very much).

Grade expectancy. Students indicated the grade that they expected to get on retesting. Letter grades were converted to standard grade-point equivalents. While some researchers stress measurement of expectancy change (McMahan, 1973; Valle & Frieze, 1976), our use of subsequent expectancy level follows the practice of the majority of investigators (e.g., Weiner, Nierenberg, & Goldstein, 1976).

Retest performance. Number correct on the second test served as the performance measure.

Resultant achievement motivation. Following the procedures of Weiner and Potepan (1970), an index of resultant achievement motivation was derived for each subject by taking the difference between z-score transformations on the Mehrabian Achievement Risk Preference Scale (MARPS) (Mehrabian, 1968, 1969) and the debilitating anxiety subscale of the Achievement Anxiety Test (AAT) (Alpert & Haber, 1960). These two indices were administered during preenrollment for the course, thereby occuring in the antecedent position dictated by the model. Selection of the present failure sample did not result in a biased subset of the nAch distribution. The mean nAch of these subjects ($M = -.081$) does not differ from that of the entire class ($M = .068$), $t(560) = 1.21$, $p < .20$. Moreover, the resultant scores of failure

Table 1
Means, Standard Deviations, and Product-Moment Correlations for Variables in the Cognitive Model of Achievement Behavior (N = 206)

Variable	nAch	Effort	Ability	Task difficulty	Luck	Shame	Expectancy	Performance
nAch		.048	−.342*	−.066	−.237*	−.194*	.208*	.145*
Effort			−.168*	−.284*	−.076	−.147*	−.113*	−.067
			(−.123)[a]	(−.229*)	(.003)			
Ability				.141*	.332*	.210*	−.110	−.096
				(.069)	(.233*)			
Task difficulty					.237*	.007	−.047	−.025
					(.204*)			
Luck						.077	.001	.012
Shame							−.047	.091
							(−.020)	
Expectancy								.357*
M	−.081	3.74	3.01	5.28	3.80	4.02	3.33	21.11
SD	1.39	1.83	1.62	1.39	1.71	1.81	.52	3.18

[a] Coefficient of residual covariation.
* $p < .05$.

subjects classified as high ($M = 1.08$) and low ($M = -1.14$) on a median split do not differ from those in the class at large ($M_H = 1.21$, $p < .20$; $M_L = -1.07$, $p < .30$), indicating equivalent heterogeneity.

Results and Discussion

The means, standard deviations, and intercorrelations among the variables of the cognitive model are presented in Table 1. The cognitive model assumes no causal reciprocity between same-level variables (i.e., among attributions or between expectancy and shame). In path-analytic terms, any observed residual covariation (e.g., −.23 between task and effort ascriptions) need not be zero, but may be assumed to originate from a common dependency on nAch or on variables outside the model. Thus these same-level covariations do not influence the causal structure and are therefore not dealt with further.

Figure 2 presents the results of the total path analysis. The values associated with the unidirectional arrows are path (standardized regression) coefficients (P). Epsilon [$E = (1 - R^2)^{\frac{1}{2}}$] represents the path coefficient value of all sources of variation not specified by the model. Preliminary path analyses by sex produced virtually identical results, hence male and female data are combined in this present analysis. These null results are interesting in light of previous evidence of sex differences in achievement attributions and their consequences (e.g., Dweck, 1977; Feather, 1968, 1969; House & Perney, 1974; Nicholls, 1975). The present findings probably point to the relatively modest role played by sex when all aspects of the model are allowed to vary naturally.

Step 1: nAch→Attributions

Path coefficient values derived from regressing each of the four attribution elements singly on nAch are presented in the left-hand portion of Figure 2. The results indicate that variations in ability and luck ascriptions have a significant dependency on nAch ($P = -.341$ and $-.237$, $p < .05$, respectively), but that effort ($P = .048$) and task ascriptions ($P = -.066$) do not. Consistent with previous results (Kukla, 1972a; Weiner et al., 1971, 1972; Weiner & Potepan, 1970), failure experiences are found to elicit greater ascriptions to low ability among low nAch persons. However, the finding that low nAch persons ascribe their failures to bad luck was unexpected and may have been precipitated by the highly ego-involving nature of the present task. Those students who expect to fail may be acting defensively to disavow personal

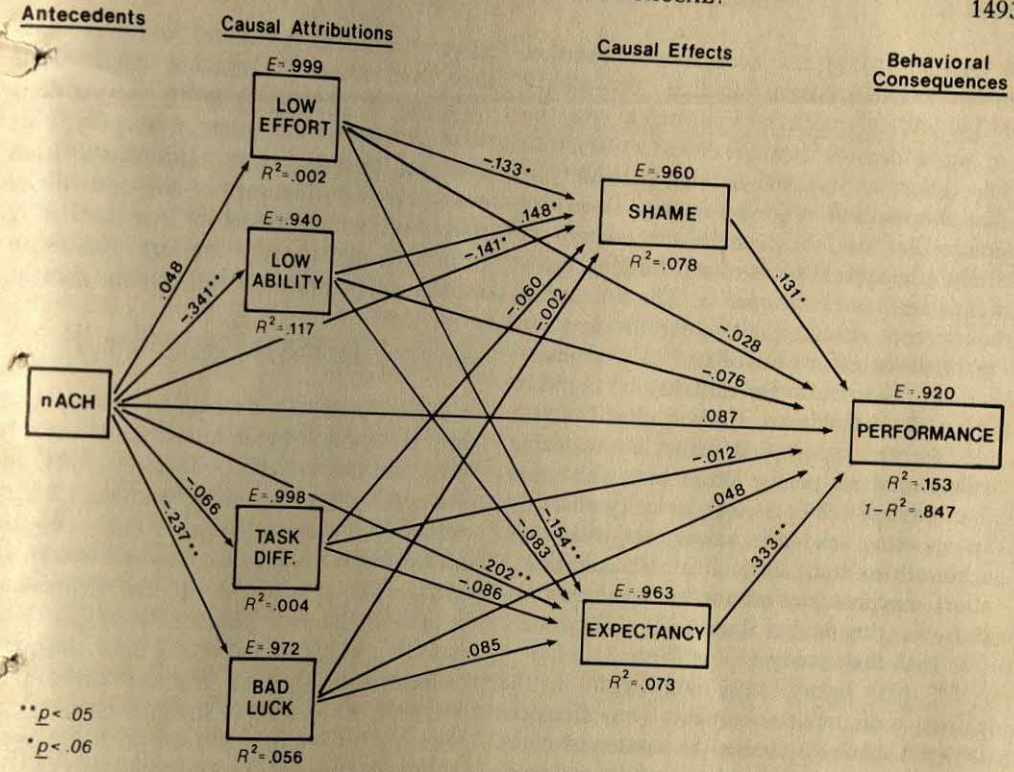


Figure 2. Path diagram of effects of nAch, causal attributions, expectancy, and shame on subsequent performance following failure.

responsibility (Covington & Omelich, Note 3). Whatever the explanation, inspection of Figure 2 indicates that luck as a causal attribution exerts no further influence on any of the distal components of the model. Similarly, task difficulty also drops out as a causal factor, having established neither a dependency on nAch nor a clear causal influence forward in the presumed chain of psychological events. Yet we cannot conclude that perceptions of task difficulty exert no influence on achievement behavior; we know they do. Variations in task complexity have been shown to influence expectations for future success (Atkinson & Feather, 1966) as well as reactions to a given performance (Parsons & Ruble, 1972) and to moderate performance (Marso, 1969; Woodson, Note 4). The fact that the task demands (test) were identical for all students probably led to a restricted range of task attributions, causing an underestimation of the true causal role of task variables.

The prediction that high nAch persons differently ascribe failure to insufficient effort was not upheld by the present data ($P = .048$, $p < .50$). Other attempts to substantiate this hypothesis have also met with disappointment (e.g., Weiner et al., 1971; Weiner & Kukla, 1970; Weiner & Potepan, 1970). Several explanations for this general failure to confirm the hypothesis appear tenable. First, according to the proposed distinction between stable and temporary effort (Rest, Nierenberg, Weiner, & Heckhausen, 1973; Weiner, 1974) cognitive predictions should be upheld if students are responding predominantly to low effort as a temporary state of affairs. But to the extent that some subjects, notably low nAch individuals, are reflecting a more trait-like disposition never to try (perhaps from a resignation born of a history of past failures), the predicted nAch→effort attribution linkage will be canceled out. Second, a self-serving, defensive interpretation is also viable. Inferences to low ability evoke shame (Covington

& Omelich, 1979), and because a combination of high effort/failure signifies low ability (Kun & Weiner, 1973), some students may be acting to deceive themselves and others with false effort ascriptions. A third possibility is that the real difference between nAch motive groups lies not so much in the *amount* of study (the typical focus of attributional questionnaires) as in its *quality*. The number of hours spent studying, either for the first test or for the retest, was unrelated to nAch membership. The reason for this may be found in the work of Goldman, Hudson, and Daharsh (1973), who report that many low-achieving college students persist longer in their studies than do high achievers, a superiority offset by an opposing tendency among yet other low achievers to study little if at all. Additional effort may not pay off for low achievers because nagging doubts about their ability interfere with their study (Wine, Note 5).

All three lines of speculation point to the existence of important qualitative differences between nAch groups in the matter of effort expenditure that have yet to be fully explored. Meantime, the predicted nAch→effort attribution relationship remains unconvincing. To the extent that nAch groups may differ in affective reactions to failure, future expectations, and retest performance itself, these variations do not appear to depend on their postdictive effort ascriptions. Naturally, this does not preclude effort attributions from exerting a direct influence on the more distal components of the model or from indirectly influencing performance through shame or expectancy, pathways we will examine shortly.

Step 2(a): nAch→Attributions→Affect

Table 2 presents the results of decomposing the relationship between nAch and shame into direct and indirect effects. Indirect effects indicate the degree to which a change in an independent variable (nAch) influences an intervening variable (cognition), which in turn results in variations in the dependent variable of interest (shame). They are estimated by summing the products of the path coefficients of the one or more paths that intervene. At present no significance tests are available for indirect effects except by evaluating the overall goodness of fit to the model.

Inspection of the first column in Table 2 indicates a significant zero-order correlation

Table 2
Decomposition of Zero-Order Correlations of Achievement Motivation With Shame, Expectancy, and Performance

Relationship	Shame	Expectancy	Performance	
Zero-order correlation	−.194*	.209*	.145*	
Direct effect (P)[a]	−.141	.202*	.087	
Total indirect effect	−.053	.007	.058	
Effort (E)	−.007	−.002	−.002	.014
Ability (A)	−.051	.026	.027	
Task difficulty (TD)	.004	.006	.001	
Luck (L)	.000	−.023	−.012	
Internal (A, E)	−.058	—	.025	
External (TD, L)	.004	—	−.011	
Stable (TD, A)	—	.032	.028	
Unstable (E, L)	—	−.024	−.014	
Internal attributions and shame	—	—	−.008	
External attributions and shame	—	—	.001	
Stable attributions and expectancy	—	—	.011	
Unstable attributions and expectancy	—	—	−.008	
Shame	—	—	−.018	
Expectancy	—	—	.066	

[a] Standardized regression coefficients.
* $p < .05$.

between nAch and shame of −.194 ($p < .05$) that decomposes into a direct effect from nAch of −.141 ($p = .054$), which falls just short of the conventional standard of significance, and into a total indirect influence on shame from nAch mediated through attributions of −.053. Reading down the remainder of column 1, we note that only ability emerges as a substantial transmitter of nAch influence (−.051), accounting for virtually all the total indirect influence on shame. The contribution of effort is negligible (−.007). Other attempts to substantiate attributions as mediators of affect among nAch groups have fared little better. While Meyer (1970) demonstrated that high and low nAch subjects differed in ability ascriptions following failure, no relationship was found between this differential ascriptive pattern and degree of affective reaction to failure. Similarly, Weiner and Potepan (1970) found nAch membership predictive of anxiety over an upcoming school test, but again affect was unrelated to accompanying attributions.

These data provide little support for the view that the principal influence of nAch on shame is mediated indirectly via cognitive mechanisms. Contrary to what would be expected if nAch acted only to initiate differential ascriptions, there is a tendency for nAch to exert a direct influence on shame in its own right, an effect certainly greater than that transmitted through attributions.

Now consider the direct influence of attributions on shame, irrespective of nAch. Figure 2 indicates two noteworthy direct pathways: one from effort ascriptions to shame ($P = -.133$, $p < .06$) and the other between ability ascriptions and shame ($P = .148$, $p < .06$). While this pattern of findings is consistent with Weiner's original contention that shame is determined primarily by variations in the internal elements of effort and ability, the direction of these influences is opposite to that predicted. First, ascriptions to low effort tend to decrease rather than increase shame; and second, ascriptions to low ability act to increase, not reduce, shame. These trends are consistent with recent findings (Covington & Omelich, 1979, in press; Nicholls, 1975, 1976; Sohn, 1977; Covington, Spratt, & Omelich, Note 6) indicating that students take pride in high self-ability attributions and experience shame at low ability, a dynamic intensified in repeated failure (Covington & Omelich, Note 7).

One consequence of this ability valuing is that high nAch subjects tend to experience less shame than do low nAch subjects, not only because of the maintenance of high ability ascriptions despite failure but also as a direct function of their motive group membership. This finding favors the traditional interpretation of failure-avoiding persons as "possessing a capacity for experiencing shame" (Atkinson, 1957) and lends weight to Nicholl's conclusion (1976) that the key difference between high and low resultant achievement motivation students lies in their differential self-concepts of ability.

Originally Weiner assumed that students would internalize teacher reinforcement patterns that favor achievement through effort and would view low ability as a mitigating circumstance in failure. However, in light of subsequent evidence (cited above), Weiner has recently acknowledged (Weiner et al., 1978) that ability and effort attributions can magnify affective reactions, depending on whether the individual is a teacher or student. The present data underscore the importance of maintaining a distinction between rewards and punishment from others and personal reactions to performance outcomes.

The left-hand portion of Table 3 displays the results of incremental F tests providing an overall summary of the path analysis for shame. As the first predictor variable entered, nAch accounts for 48.5% (.038) of the explained variation in shame through all paths of influence, direct and indirect, $F(1, 200) = 8.16$, $p < .01$. Next, the four attributional sources combined add .040 to R^2 and account for the remaining 51.5% of the explained variance ($R^2 = .078$), $F(4, 200) = 2.16$, $p < .08$. Only the internal elements of ability, $F(1, 200) = 3.81$, and effort, $F(1, 200) = 4.13$, contribute significantly to this total ($p < .05$), with the contribution of effort coming entirely from its direct influence on shame. However, it is reasonable to wonder if the introduction of nAch as the first variable in

Table 3
Incremental F Tests: Effects of Achievement Motivation and Attributions on Shame and Expectancy

Antecedent variable	*df*	Shame				Expectancy			
		R^2	Increment to R^2	F	% of R^2	R^2	Increment to R^2	F	% of R^2
nAch	1	.038	—	8.16*	.485	.043	—	9.38*	.597
Combined attributions	4	.078	.040	2.16	.515	.072	.029	1.58	.403
Effort	1	.057	.019	4.13*	.253	.058	.015	3.28	.209
Ability	1	.056	.018	3.81*	.227	.048	.005	1.20	.075
Task difficulty	1	.041	.003	.36	.036	.050	.007	1.46	.093
Luck	1	.038	.000	.11	.007	.045	.002	.42	.027
Stable	2	.058	.020	2.21	.263	.055	.012	1.31	.168
Unstable	2	.058	.020	2.12	.260	.060	.017	1.85	.235
Internal	2	.075	.037	3.97*	.479	.064	.021	2.23	.283
External	2	.041	.003	.36	.043	.052	.009	.94	.120
R^2 total		.078				.072			

Note. df residual = 200.
* $p < .05$.

our path analysis acted to attenuate the impact of attributions on shame, given the considerable variance shared by nAch and some of the attributional elements (see Table 1). Accordingly, an analysis with nAch deleted yields a significant effect of attributions on shame, $F(4, 201) = 3.22$, $p < .05$. While the internal attributions of ability ($p < .01$) and effort ($p < .07$) now account for virtually all the explained variance in shame (74% and 26%, respectively), there is a decrease in R^2 from .078 (with nAch included) to .060. However, this decrement fails to reach significance at the conventional level, $F(1, 200) = 3.68$, $p < .06$.

These results establish the importance of the direct influence of internal attributions on shame, even though their role as transmitters of motive dispositions is less than expected and in a direction opposite that predicted by the cognitive model. It could be argued that these data support the contention that nAch is "nothing more than one manifestation of differential attributions" (Weiner, Note 8), especially since deletion of nAch results in a nonsignificant decrement in explained variance. However, the present nAch data pattern, in conjunction with previously cited findings (Meyer, 1970; Weiner & Potepan, 1970), suggests that nAch does make a separate contribution to shame, independent of the predictive power it shares with ability attributions. Indeed, one further analysis deleting the four attributions from the prediction equation indicates a significant contribution from nAch, in the absence of any intervening influence of causal cognitions ($R^2 = .038$), $F(1, 204) = 7.98$, $p < .05$. Despite their commonality, we conclude that both motive dispositions and internal cognitive ascriptions exercise a direct and independent influence on shame.

Step 2(b): nAch→Attributions→Expectancy

The second column of Table 2 indicates a significant zero-order correlation between nAch and expectancy for future success of .209 ($p < .05$) that decomposes into a significant direct effect of .202 ($p < .05$) and a negligible indirect effect from nAch mediated through all four attributional sources of .007. We conclude that variations in expectancy have a substantial direct dependency on motive group membership and that contrary to cognitive predictions, attributions do not act as mediators for this covariation. The weight of previous evidence on failure does little to alter this conclusion. While Meyer (1970) reported that high and low nAch groups differed in their attributions to stable elements, these ascriptions led to expectancy shifts only among low nAch students. Moreover, Weiner

and Kukla (1970) found that although high nAch subjects reported greater ability levels, they did not as a consequence hold higher expectancies for future success. Thus it appears that whatever differential attributions high and low nAch persons make in response to failure, they play little part in determining expectancies for future success.

Now consider the evidence for a direct influence of attributions on future expectancy, irrespective of nAch membership. Whenever the direct attribution→expectancy linkage has been the sole focus of research, the evidence has overwhelmingly favored cognitive predictions (Fontaine, 1974; McMahan, 1973; Meyer, 1970; Rosenbaum, 1972; Valle, 1974; Weiner et al., 1976; the work of Rosenbaum, 1972, and Valle, 1974, is fully described in Weiner et al., 1976). These studies clearly demonstrate that expectations of future success remain higher when previous failures are attributed to unstable elements than to stable ones. In view of such agreement the present findings may appear puzzling. Figure 2 indicates that three of the four attributional elements (ability, luck, and task difficulty) exert no significant direct causal influence on variations in expectancy. This is also contrary to recently developed theoretical alternatives centering on learned helplessness (Abramson, Seligman, & Teasdale, 1978; Dweck, 1975, 1977; Covington & Omelich, Note 7), suggesting that self-ascriptions to low ability should produce expectancy decrements. However, it must be recalled that path analysis assesses the relative influence of each of many factors on a single dependent variable simultaneously. The role of attributions as direct causal mediators is probably overshadowed by the influential presence of nAch in the full model. When nAch is excluded from the path analysis and attributional sources are considered in isolation, the causal link between ability ascriptions and expectancy now approaches significance ($P = -.143$, $p < .06$), suggesting that this stable element acts to depress future expectancies of success. The insignificant role played by task ascriptions remains unchanged by these manipulations ($P = -.084$). This may reflect the possibility that, although it is a stable element, failure ascribed to a difficult task does not necessarily imply personal inadequacy and hence holds no particular implication for the future (Klein, Fencil-Morse, & Seligman, 1976; Tennen & Eller, 1977). Thus our data lend partial support to the hypothesized role of stable elements as mediators of future expectancy, while pointing up their relatively modest causal role when compared to other variables in the model.

We note in Figure 2 a significant direct, causal path from effort ascriptions to expectancy ($P = -.154$, $p < .05$), but in a direction opposite to that predicted. Rather than raising hopes for future success, low effort ascriptions for a previous failure appear to dash them. Although having studied little for a first test may in principle sustain optimism about future test outcomes, this is likely only if time permits sufficient additional study before the next test. When the time interval between tests is relatively short, as in the present study, students who did not prepare adequately the first time are probably less sanguine about a second chance. Weiner's notion of stable effort may also apply (Rest et al., 1973; Weiner, 1974, 1979). Consistent inaction in situations that clearly demand concentrated effort suggests an attitude of demoralization that carries with it negative expectations for future success. For some students a combination of low effort and pessimism for the future may simply reflect an underlying negative attitude regarding their sense of personal causality (deCharms, 1968, 1972). At the same time, these postdictive effort ascriptions may be partly defensive. The more effort expended in a failing cause, the more individuals are likely to underestimate actual time spent studying (Covington & Beery, 1976), a self-serving tactic that acts to confound cognitive predications.

The right-hand portion of Table 3 provides a summary of the findings for expectancy. An incremental F test indicates that nAch accounts for some 60% of the explained variance, $F(1, 200) = 9.38$, $p < .05$. Of all the attributional elements taken singly, only low effort contributes to any degree (although not significantly), accounting for half the remaining 40% of the explained variance in ex-

pectancy, $F(1, 200) = 3.28$, $p < .08$. Again to avoid underestimating the role of attributions as causal agents because of the significant covariation between nAch and attributions, the analyses in Table 3 were recalculated with nAch deleted. A significant decrease in R^2 is the result, from .072 (with nAch included) to .037, $F(1, 200) = 7.54$, $p < .01$. Thus nAch exerts a substantial influence on expectancy in the absence of any discernible impact of cognitions.

Step 3: nAch→Attributions→Affect/Expectancy→Retest

Inspection of column 3 in Table 2 reveals a significant zero-order correlation between nAch and retest performance of .145 ($p < .05$) that decomposes into a direct effect on performance of .087 and into a total indirect effect of .058. Five pathways make up the total indirect influence of nAch on performance. Three involve attributions as mediating links and should therefore make substantial contributions. The two remaining pathways involve only expectancy and affect, respectively, as the sole mediating links between nAch and performance and should in theory contribute relatively little to the total indirect effect. To simplify, we have squared and summed the individual contributions of each of these five pathways to permit a percentage interpretation.

1. *nAch→attributions→performance.* The value (.014) associated with the four bracketed entries in column 3 of Table 2 indicates the degree of nAch influence on performance mediated through variations in the four attributional sources combined. The absolute value of this effect (disregarding algebraic signs) represents roughly 15% of the total indirect effect of motive group on performance. Ability is the only attributional mediator of importance, accounting for most of the 15% contribution (13%). Previous research on this nAch→attribution→performance link has been inconclusive. Weiner and Sierad (1975) and Meyer (1970) report the relationship holding only for low nAch subjects, whereas Kukla (1972a) found it only among high nAch subjects. In light of the present data this inconsistency may be interpreted as due to the modest contribution of the nAch→attribution linkage to performance.

2. *nAch→attributions→shame→performance.* The internal dimensions of ability and effort have a negligible influence (−.008) on nAch group performance as mediated through shame. As expected, external elements play no part in this dynamic (.001). These combined sources account for only 1% of the total indirect effect.

3. *nAch→attributions→expectancy→performance.* Similarly, ascriptions to stable elements (.011) and to unstable elements (−.008) contribute little to the performance outcomes of nAch groups as mediated through expectancy. These combined sources account for 3% of the total indirect effect of nAch influence on performance.

4. *nAch→shame→performance.* Variations in shame produce a stronger indirect pathway of nAch influence on performance (−.018), accounting for 6% of the total indirect effect.

5. *nAch→expectancy→performance.* The most powerful source of indirect influences for nAch membership on performance is expressed through variations in expectancy (.066), accounting for 75% of the total indirect causal impact of motive group membership on performance.

In summary, the central proposition of cognitive theory, that performance outcomes among motive groups depend largely on differential attributions made for previous failures, finds little support. The main mediators of differential performance prove to be expectancy and, to a lesser degree, ability and shame, which together account for some 94% of the total indirect effect of motive group membership on performance.

Next, consider the direct effect of expectancy, shame, and the four elements of causal ascription on performance. Figure 2 indicates that the single most powerful source of direct influence is expectancy ($P = .333$, $p < .05$), followed by shame, which makes a modest, although nonsignificant, contribution ($P = .131$, $p < .06$). The direct path coefficient for each attribution is also nonsignificant (effort, $P = -.028$; ability, $P = -.076$; task difficulty, $P = -.012$; and luck, $P = .048$).

Table 4
Incremental F *Tests: Effects of Achievement Motivation, Attributions, Expectancy, and Shame on Performance*

Antecedent variable	*df*	R^2	Increment to R^2	*F*	% of R^2
nAch	1	.021	—	4.89*	.137
Effort	1	.027	.006	1.31	.037
Ability	1	.027	.006	1.32	.037
Task difficulty	1	.023	.002	.54	.015
Luck	1	.022	.001	.43	.012
Stable	2	.029	.008	.93	.052
Unstable	2	.038	.007	.87	.049
Internal	2	.032	.011	1.32	.074
External	2	.025	.004	.49	.027
Combined attributions	4	.036	.015	.90	.101
Shame	1	.037	.016	3.71	.104
Expectancy	1	.122	.101	23.61*	.660
R^2 total		.153			

Note. df residual = 198.
* $p < .05$.

These four path coefficients were recalculated excluding nAch. However, the new values differ little (−.027; −.098; −.012; and .036, respectively), and in no cases were changes sufficient to achieve statistical significance. Traditionally, theoretical interest has focused on the indirect mediating role of attributions, although the weight of evidence (Dweck, 1975, 1977; Kukla, 1972a; Weiner et al., 1972; Weiner & Potepan, 1970; Weiner & Sierad, 1975) suggests the existence of direct attributional links to performance, at least in the case of failure. However, the present findings cause us to wonder about the importance of such linkages relative to other direct and indirect pathways in the model.

As the last step, we assess the indirect influence of attributions mediated through expectancy and shame (attributions→shame/expectancy→performance), regardless of nAch. These data indicate that the total indirect influence from all attributional sources (disregarding algebraic signs) via expectancy is .136 and .045 via shame. While no statistical comparisons are possible between direct and indirect pathways, it is clear that expectancy ($P = .333$) and, to a lesser extent, shame ($P = .131$) exert a considerable direct influence on performance, independent of attributions, whereas, contrary to cognitive theory, variations in expectancy and shame that do depend on attributional mediators have a relatively modest influence on performance.

With the completion of the path analysis of performance, we come full circle, having investigated the entire presumed causal network among all elements of the cognitive model. Whether or not the magnitude of the various attributional influences, both direct and indirect, is sufficient to justify the cognitive theory of achievement motivation in its present form is a major question raised by this study. The results of incremental *F* tests presented in Table 4 help put this question in perspective. Need achievement accounts for some 14% of the explained variation in performance, $F(1, 198) = 4.89$, $p < .05$, through both its direct and indirect influences. The combined attributional sources account for an additional 10%, $F(4, 198) = .90$, $p < .50$. Expectancy, $F(1, 198) = 23.61$, $p < .05$, and shame, $F(1, 198) = 3.71$ $p < .06$, account for the remaining explained variance in performance, 66% and 10%, respectively, and in conjunction with nAch explain 90% of the total R^2. Once again, the data were reanalyzed excluding nAch. The removal of nAch does not appreciably decrease the total R^2 for performance, $F(1, 198) = 1.41$, $p < .25$. But neither is there a corresponding increase in the predictive power of attributions, $F(1, 199) = 1.01$,

$p < .50$. Cognitions, which now account for 14% of the total R^2, as compared to the original 10%, continue to exert only a negligible impact on performance through both direct and indirect pathways, $F(4, 199) = 1.20$, $p < .50$. It is expectancy that gains as a determinant of retest performance, from 66% to 77%, when nAch is removed. This is due to the considerable predictive power shared by nAch and expectancy in the full model.

Another approach for judging the importance of attributions is to assess the predictive power of the traditional theory of achievement motivation in the absence of mediating cognitions. In its simplest form Atkinson's model (1957, 1964) postulates that the performance of any individual depends on a multiplicative combination of achievement motivation (nAch), the incentive value (I_s) of the task (transliterated in Weiner's system as shame), and expectancy of future success (P_s). The final entry in Table 4 indicates that when these basic elements are combined along with attributions, a total of 15.3% of the variance in retest performance is explained. When the identical path analysis is repeated, but this time with all attributional sources removed, 14.8% of the variance remains accounted for, a decrease in explained variance that does not reach statistical significance, $F(4, 198) = 1.17$, $p < .50$. It appears that knowledge of causal ascriptions in failure tells us little more about the quality of subsequent performance than is already known from the elements of the traditional model.

General Discussion

The overall results of this study offer little support for the cognitive model of achievement behavior as presently formulated. First, achievement motive groups do not differ in test performance primarily as a consequence of differential attributions made for a previous failure. Second, although some attributions (notably effort and ability) do contribute to negative affect and future expectancy, these causal relationships are not always in the prescribed direction. Moreover, the effects of attributions on performance as mediated through expectancy and affect are disappointingly small. Third, nAch emerges as an important influence on achievement behaviors in its own right. In the case of shame, nAch contributes both independent of and in common with mediating cognitions, primarily ability. Moreover, nAch exerts a direct impact on expectancy in the absence of a significant contribution from attributions. And finally, nAch affects subsequent performance through a combination of direct influence and indirect influence mediated largely by expectancy, a network basically unaffected by cognitions. We agree with Weiner's recent reassessment that "the presumed causal biases ascribed to individual differences in . . . achievement needs remain somewhat tenuous" (Weiner et al., 1978).

What part, then, if any, do attributions play in the achievement dynamic? In our view, attributions, expressed postdictively, are reactions to rather than causes of performance. As postdictive reactions to failure, attributions are prone to reflect defensive, self-serving biases (Miller, 1976; Snyder, Stephan, & Rosenfield, 1976; Stephan, Rosenfield, & Stephan, 1976; Covington & Omelich, Note 3) and thereby obscure their causal role as information-processing determinants of subsequent achievement behavior. We do not deny the importance of rational considerations in achievement behavior, nor do we argue that self-perceptions of ability and effort ("power" and "try" in Heider's terms) are unimportant. We only suggest that such factors presently represented in the cognitive model as postdictive explanations of success and failure should be relocated as exogenous, antecedent determinants of achievement behavior where they are less likely to be confounded by self-serving tendencies.

The present data strengthen the case for differential self-perceptions of ability as the primary ingredient in achievement motivation (Kukla, 1972a, 1972b; Moulton, 1974; Nicholls, 1976). First, there is a substantial correlation between the Brookover Self-concept of Ability Scale (Brookover, LePere, Hanachek, Thomas, & Erickson, 1965) and the present index of resultant achievement motivation ($r = .44$, $p < .01$). Second, inspection of Figure 2 indicates that despite the

likely presence of self-serving bias, ability nonetheless emerges as the only attribution that links the entire model from nAch to retest performance (via shame). This finding strengthens the conceptual basis for recent work on learned helplessness that focuses on the ability→performance pathway (Abramson, Seligman, & Teasdale, 1978; Dweck, 1977; Covington & Omelich, Note 7).

The central importance of perceived ability to all achievement behavior, including affect and self-confidence, is underscored in a recently proposed self-worth theory of achievement motivation (Beery, 1975; Covington & Beery, 1976) holding that students attempt to maintain and promote self-ascriptions to high ability owing to society's tendency to equate worth with the ability to achieve competitively. In this self-serving process, attributions may not always reflect rational considerations. For example, Covington and Omelich (Note 3) found that males with low self-concept of ability devalued the ability of others compared to their own under identical circumstances of failure, whereas females with low self-concept denigrated their abilities beyond what was rationally indicated by the information at hand. The presence of such bias can only act to obscure, not clarify, the predicted relationships between postdictive attributions and shame, expectancy, and outcome. The study by Weiner and Sierad (1975) illustrates how the interjection of defensive attributions (excuses) can completely alter expected performance outcomes. A drug (placebo) alleged to interfere with hand–eye coordination was administered to college subjects of high and low nAch prior to a learning task. Armed with an explanation for potentially poor performance other than inability, low nAch subjects were expected to perform better than usual, while high nAch subjects were expected to perform poorer than usual owing to the presumed interference of the drug with ability. Both predictions were borne out. The authors correctly conclude that the experiment demonstrates the causal influence of cognitions on performance. Yet it illustrates something more. The presence of self-serving excuses, a factor overlooked by both traditional and cognitive theories, can act to confound the usual attributional linkages.

This reconsideration of the role of attributions underscores the need, in Weiner's words, ". . . to specify the conditions under which individuals avoid feedback and information as a self-protecting, coping strategy" (1977). Identifying the circumstances of failure that elicit defensiveness would appear a useful starting point, with special attention to Heider's dictum (1958) that in order for self-serving biases to be effective, they must fit the constraints of reason. In this connection, Covington and Omelich (Note 3) found that individuals aggrandize their own ability at the expense of others only when such self-protective judgments are plausible, either when their failure is accompanied by low effort or when excuses are available to explain why effort did not pay off.

A reinterpretation of the present data from a self-worth/ability perspective, corroborated by evidence from related sources, provides a glimpse of the complex interaction between affective, social, and cognitive factors in the achievement context. Consider first the dynamics of low nAch students as reflected in Figure 2. Owing to perennial self-doubts about their ability, failure elicits cognitions of incompetency that, in turn, cause shame (low nAch→low ability→shame), a dynamic confirmed independently by laboratory studies (Covington & Omelich, 1979; Sohn, 1977). Shame is further intensified directly through nAch dispositions (low nAch→shame), the linkage originally postulated by Atkinson (1957). This suggests that negative reactions to failure have long been conditioned by a history of failure, irrespective of any implication of low ability. A third source of shame depends on effort level (high effort→shame): effort intensifies shame because high effort/failure is evidence of low ability (Kun & Weiner, 1973) and inferences of inability evoke shame (Covington & Omelich, 1979). While this source of distress may be mitigated temporarily by not studying, as in the case of procrastination (Birney, Burdick, & Teevan, 1969), such tactics are maladaptive in the long run. Students who are perceived by

teachers as having studied little are rewarded less in success and punished more in failure than those students who do try (Rest et al., 1973; Weiner, 1972, 1974). Moreover, according to the present data configuration, low effort may also act to depress retest performance in two interlocking ways. First, low effort contributes to reduced shame (Covington & Omelich, 1979), which in turn tends to decrease performance, presumably to the extent that chagrin might otherwise mobilize effort (Weiner, 1974; Weiner et al., 1971). Second, low effort acts indirectly to suppress performance through lowered expectations of future success (low effort→low expectancy→ low performance). Unfortunately, the low nAch student who works hard is unlikely to fare much better even with the teacher approval that comes from trying. Studying typically occurs in the face of poor prospects for success (low nAch→low expectations), and the likelihood of failure despite high effort simply intensifies the threat to self-esteem as further evidence of low ability (Covington & Beery, 1976). Little wonder that the study habits of low nAch persons have been characterized as conflicted, tense, and ineffectual (Wine, Note 5).

Figure 2 explicates an entirely different dynamic in the case of high nAch individuals. These students tend to experience less distress in failure both because, as predicted by Atkinson (1957), they are success-oriented (high nAch→low shame), and because they are secure in their sense of competency (high nAch→high ability→low shame) (Moulton, 1974; Nicholls, 1976). Although the resulting lower level of shame tends to dampen subsequent performance (high nAch→low shame→ low performance), the relatively small magnitude of this negative suppression effect is more than offset by the tendency to maintain a high expectation of success that manifests itself in increased retest performance (high nAch→high expectancy→high performance). From this it can be seen why an occasional failure for the high nAch student gives little cause for self-doubting of ability or for lowered expectations, nor is it likely to precipitate a vicious cycle of self-defeat so common among low nAch pupils.

Conclusion

Progress toward a fuller understanding of achievement behavior will be made only as long as theories are tested in the naturally occurring contexts of achievement. We conclude from the evidence of the present field study that the achievement process is not solely, or even primarily, a manifestation of cognitive attributions. Rather it is most productively viewed as the operation of personality and socially conditioned dispositions emerging from the individual's efforts to protect and enhance a sense of competency. To be sure, these efforts are moderated by rational information-processing principles of the kind originally proposed by Heider and elaborated by Weiner. Indeed, it is this dynamic interaction of cognition and affect that presents the greatest challenge to the researcher and theoretician.

Reference Notes

1. Weiner, B., Russell, D., & Lerman, D. *The cognition-emotion process in achievement-related contexts.* Manuscript submitted for publication, 1978.
2. Block, J. H. *Mastery performance standards and student learning.* Unpublished manuscript, University of California, Santa Barbara, 1973.
3. Covington, M. V., & Omelich, C. L. *Sex differences in self-serving perceptions of ability.* Unpublished manuscript, Department of Psychology, University of California, Berkeley, 1978.
4. Woodson, C. E. *Motivational effects of two-stage testing.* Unpublished manuscript, Institute of Human Learning, University of California, Berkeley, 1975.
5. Wine, J. *Cognitive-attentional approaches to test anxiety modification.* In W. H. Holtzman (Chair), *Anxiety and instruction.* Symposium at the meeting of the American Psychological Association, Montreal, 1973.
6. Covington, M. V., Spratt, M. F., & Omelich, C. L. *Stable and unstable effort: A test of Weiner's hypothesis.* Unpublished manuscript, Department of Psychology, University of California, Berkeley, 1979.
7. Covington, M. V., & Omelich, C. L. *Mastery learning and learned helplessness.* Unpublished manuscript, University of California, Berkeley, 1979.
8. Weiner, B. Personal communication, May 1977.

References

Abramson, L. Y., Seligman, M. E. P., & Teasdale, J. D. Learned helplessness in humans: Critique and reformulation. *Journal of Abnormal Psychology,* 1978, *87,* 49–74.

Alpert, R., & Haber, R. N. Anxiety in academic achievement situations. *Journal of Abnormal and Social Psychology,* 1960, *61,* 207–215.

Anderson, J. G., & Evans, F. B. Causal models in educational research: Recursive models. *American Educational Research Journal,* 1974, *11,* 29–39.

Atkinson, J. W. Motivational determinants of risk-taking behavior. *Psychological Review,* 1957, *64,* 359–372.

Atkinson, J. W. *An introduction to motivation.* Princeton, N.J.: Van Nostrand, 1964.

Atkinson, J. W., Bastian, J. R., Earl, R. W., & Litwin, G. H. The achievement motive, goal setting and probability preferences. *Journal of Abnormal and Social Psychology,* 1960, *60,* 27–36.

Atkinson, J. W., & Feather, N. R. (Eds.). *A theory of achievement motivation.* New York: Wiley, 1966.

Beery, R. Fear in failure in the student experience. *Personnel and Guidance Journal,* 1975, *54,* 190–203.

Birney, R. C., Burdick, H., & Teevan, R. *Fear of failure.* New York: Van Nostrand Reinhold, 1969.

Block, J. W. Motivation, evaluation, and mastery learning. *UCLA Educator,* 1977, *12,* 31–37.

Brookover, W. B., LePere, J. M., Hanachek, D. E., Thomas, S., & Erickson, E. L. *Self-concept of ability and school achievement,* II. U.S. Office of Education, Cooperative Research Project #845. East Lansing: U.S. Office of Research and Publications, Michigan State University, 1965.

Covington, M. V., & Beery, R. *Self-worth and school learning.* New York: Holt, Rinehart & Winston, 1976.

Covington, M. V., & Omelich, C. L. Effort: The double-edged sword in school achievement. *Journal of Educational Psychology,* 1979, *71,* 169–182.

Covington, M. V., & Omelich, C. L. It's best to be able and virtuous too: Student and teacher evaluative responses to successful effort. *Journal of Educational Psychology,* in press.

DeCharms, R. *Personal causation: The internal affective determinants of behavior.* New York: Academic Press, 1968.

DeCharms, R. Personal causation training in the schools. *Journal of Applied Social Psychology,* 1972, *2,* 95–113.

Duncan, O. D. Path analysis: Sociological examples. *American Journal of Sociology,* 1966, *72,* 1–16.

Duncan, O. D. *Introduction to structural equation models.* New York: Academic Press, 1975.

Dweck, C. S. The role of expectations and attributions in the alleviation of learned helplessness. *Journal of Personality and Social Psychology,* 1975, *31,* 674–685.

Dweck, C. S. Learned helplessness and negative evaluation. *UCLA Educator,* 1977, *19,* 44–49.

Feather, N. T. Change in confidence following success or failure as a predictor of subsequent performance. *Journal of Personality and Social Psychology,* 1968, *9,* 38–46.

Feather, N. T. Attribution of responsibility and valence of success and failure in relation to initial confidence and task performance. *Journal of Personality and Social Psychology,* 1969, *13,* 129–144.

Fontaine, G. Social comparison and some determinants of expected personal control and expected performance in a novel task situation. *Journal of Personality and Social Psychology,* 1974, *29,* 487–496.

Goldman, R., Hudson, D., & Daharsh, B. Self-estimated task persistence as a nonlinear predictor of college success. *Journal of Educational Psychology,* 1973, *65,* 216–221.

Heider, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

Heise, D. R. *Causal analysis.* New York: Wiley, 1975.

House, W. C., & Perney, V. Valence of expected and unexpected outcomes as a function of locus of goal and type of expectancy. *Journal of Personality and Social Psychology,* 1974, *29,* 454–463.

Kerlinger, F. N., & Pedhazur, E. *Multiple regression in behavioral research.* New York: Holt, Rinehart Winston, 1973.

Klein, D. C., Fencil-Morse, E., & Seligman, M. E. P. Learned helplessness, depression, and the attribution of failure. *Journal of Personality and Social Psychology,* 1976, *33,* 508–516.

Kukla, A. Attributional determinants of achievement-related behavior. *Journal of Personality and Social Psychology,* 1972, *21,* 166–174. (a)

Kukla, A. Foundation of an attributional theory of performance. *Psychological Review,* 1972, *79,* 454–470. (b)

Kun, A., & Weiner, B. Necessary versus sufficient causal schemata for success and failure. *Journal of Research in Personality,* 1973, *7,* 197–207.

Latta, R. M. Differential tests of two cognitive theories of performance: Weiner versus Kukla. *Journal of Personality and Social Psychology,* 1976, *34,* 295–304.

Marso, R. N. The influence of test difficulty upon study efforts and achievement. *American Educational Research Journal,* 1969, *6,* 621–632.

McMahan, I. D. Relationships between causal attributions and expectancy of success. *Journal of Personality and Social Psychology,* 1973, *28,* 108–114.

Mehrabian, A. Male and female scales of the tendency to achieve. *Educational and Psychological Measurement,* 1968, *28,* 493–502.

Mehrabian, A. Measure of achieving tendency. *Educational and Psychological Measurement,* 1969, *29,* 445–451.

Meyer, W. U. *Selbstverantwortlichkeit und Leistungs-Motivation.* Unpublished doctoral dissertation, Ruhr Universität, Bochum, Germany, 1970.

Miller, D. T. Ego involvement and attributions for success and failure. *Journal of Personality and Social Psychology*, 1976, *34*, 901–906.

Moulton, R. W. Motivational implications of individual differences in competence. In J. W. Atkinson & J. O. Raynor (Eds.), *Motivation and achievement*. Washington, D.C.: V. H. Winston, 1974.

Nicholls, J. G. Causal attributions and other achievement-related cognitions: Effects of task outcome, attainment value, and sex. *Journal of Personality and Social Psychology*, 1975, *31*, 379–389.

Nicholls, J. G. Effort is virtuous, but it's better to have ability: Evaluative responses to perceptions of effort and ability. *Journal of Research in Personality*, 1976, *10*, 306–315.

Parsons, J. E., & Ruble, D. N. Attributional processes related to the development of achievement-related affect and expectancy. *Proceedings of 80th Annual Convention of the American Psychological Association*, 1972.

Rest, S., Nierenberg, R., Weiner, B., & Heckhausen, H. Further evidence concerning the effects of perceptions of effort and ability on achievement evaluation. *Journal of Personality and Social Psychology*, 1973, *28*, 187–191.

Rosenbaum, R. M. *A dimensional analysis of the perceived causes of success and failure*. Unpublished doctoral dissertation, University of California, Los Angeles, 1972.

Snyder, M. L., Stephan, W. G., & Rosenfield, D. Egotism and attribution. *Journal of Personality and Social Psychology*, 1976, *33*, 435–441.

Sohn, D. Affect-generating powers of effort and ability self attributions of academic success and failure. *Journal of Educational Psychology*, 1977, *69*, 500–505.

Stephan, W. G., Rosenfield, D., & Stephan, C. Egotism in males and females. *Journal of Personality and Social Psychology*, 1976, *34*, 1161–1167.

Tennen, H., & Eller, S. J. Attributional components of learned helplessness and facilitation. *Journal of Personality and Social Psychology*, 1977, *35*, 265–271.

Valle, V. A. *Attributions of stability as a mediator in the changing of expectations*. Unpublished doctoral dissertation, University of Pittsburgh, 1974.

Valle, V. A., & Frieze, I. H. Stability of causal attributions as a mediator in changing expectations for success. *Journal of Personality and Social Psychology*, 1976, *33*, 579–587.

Weiner, B. *Theories of motivation: From mechanism to cognition*. Chicago: Markham, 1972.

Weiner, B. (Ed.) *Achievement motivation and attribution theory*. Morristown, N.J.: General Learning Press, 1974.

Weiner, B. An attributional model for educational psychology. In L. Shulman (Ed.), *Review of research in education* (Vol. 4). Itasca, Ill.: Peacock, 1977.

Weiner, B. A theory of motivation for some classroom experiences. *Journal of Educational Psychology*, 1979, *71*, 3–25.

Weiner, B., Frieze, I., Kukla, A., Reed, L., Rest, S., & Rosenbaum, R. Perceiving the causes of success and failure. In E. E. Jones et al. (Eds.), *Attribution: Perceiving the causes of behavior*. Morristown, N.J.: General Learning Press, 1971.

Weiner, B., Heckhausen, H., Meyer, W., & Cook, R. Causal ascriptions and achievement behavior: A conceptual analysis of effect and reanalysis of locus of control. *Journal of Personality and Social Psychology*, 1972, *21*, 239–248.

Weiner, B., & Kukla, A. An attributional analysis of achievement motivation. *Journal of Personality and Social Psychology*, 1970, *15*, 1–20.

Weiner, B., Nierenberg, R., & Goldstein, M. Social learning (locus of control) versus attributional (causal stability) interpretations of expectancy of success. *Journal of Personality*, 1976, *44*, 52–68.

Weiner, B., & Potepan, P. A. Personality characteristics and affective reactions toward exams of superior and failing college students. *Journal of Educational Psychology*, 1970, *61*, 144–151.

Weiner, B., Russell, D., & Lerman, D. Affective consequences of causal ascriptions. In J. H. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research* (Vol. 2). Hillsdale, N.J.: Erlbaum, 1978.

Weiner, B., & Sierad, J. Misattribution for failure and enhancement of achievement strivings. *Journal of Personality and Social Psychology*, 1975, *31*, 415–421.

Werts, C. E., & Linn, R. L. Path analysis: Psychological examples. *Psychological Bulletin*, 1970, *74*, 193–213.

Received August 8, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 9, 1505-1518

Uncertainty-Reducing Properties of Achievement Tasks

Yaacov Trope
The Hebrew University of Jerusalem, Jerusalem, Israel

Two experiments were conducted to test the idea that choice among achievement tasks depends on the extent to which performance outcomes are expected to reduce uncertainty about one's ability. Two determinants of expected uncertainty reduction were investigated, the diagnosticity of the task at each zone on the ability scale and the person's uncertainty regarding the ability levels in each zone. As in previous research, the preference for a task increased with its overall diagnosticity, that is, its diagnosticity over the entire ability scale. Furthermore, among tasks of equal overall diagnosticity, subjects preferred tasks that were the most diagnostic at the zone where most of their uncertainty was concentrated, irrespective of the location of this zone on the ability scale. It is concluded that subjects selected tasks that would maximize expected reduction of uncertainty about their standing on the ability scale.

A theme running through a considerable amount of social psychological theory and research is the idea that people characteristically strive to gain self-knowledge. Students of social comparison and attribution processes have contended that a wide range of social phenomena, from emotional reactions and attitude change to affiliation and interpersonal communication, are affected by people's underlying need to assess their own attributes accurately (Festinger, 1954; Nisbett & Valins, 1972; Schachter, 1959; Suls & Miller, 1977). This article is concerned with the manifestations of this need in the domain of achievement behavior. More specifically, it focuses on the effects of (a) initial knowledge about one's own ability and (b) ability-relevant information contained in performance outcomes on choice among achievement tasks.

In the last few years, it has been argued that people choose among achievement tasks according to the amount of ability-relevant information they expect to gain from performance outcomes (Trope & Brickman, 1975; Weiner, Frieze, Kukla, Reed, Rest, & Rosenbaum, 1972). In their paper on the informational properties of achievement tasks, Trope and Brickman proposed that the well-documented preference for intermediate-difficulty tasks reflects a strategy of maximizing information about one's ability. This self-assessment interpretation was offered as an alternative to Atkinson's (1957) widely known theory of achievement motivation, which postulates that a choice of tasks where probability of success equals .50 maximizes positive achievement affect. Specifically, Trope and Brickman argued that the amount of ability-relevant information contained in performance outcomes depends on the *diagnosticity* of the task, that is, the extent to which performance outcomes are expected to vary as a function of ability level. The greater the differences in performance among people of different ability levels, the greater the diagnosticity of the task. Ordinarily, the differences between ability levels tend to be greater at intermediate difficulty tasks. However, Trope and Brickman showed that it is possible to construct tasks in which difficulty and diagnosticity vary independently. When

This research was supported by a grant from the Institute for Human Development at the Hebrew University of Jerusalem. The author would like to thank Gitza Freedman and Daphna Kobovy for research assistance.

Requests for reprints should be sent to Yaacov Trope, who is now at the Department of Psychology, University of Toronto, Toronto, Ontario, Canada M5S 1A1.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3709-1505$00.75

presented with such tasks, Trope and Brickman's subjects preferred high-diagnosticity to low-diagnosticity tasks, irrespective of whether the tasks were of intermediate difficulty. The effect of diagnosticity was replicated and extended in subsequent studies (Trope, 1975; Zuckerman, Brown, Fischler, Fox, Lathin, & Minasian, in press; Buckert, Meyer, & Schmalt, Note 1).

In addition to task characteristics (such as diagnosticity), the amount of prior knowledge about one's ability should also affect information gain. A person who has little prior uncertainty about his or her ability level cannot expect to gain much new information from performing tasks low or high in diagnosticity. As uncertainty about one's ability increases, the amount of information the individual can gain should also increase. The notion that uncertainty instigates information seeking is not new to psychology. Extensive research has demonstrated that in various domains, prior to choosing from among several alternatives, the amount of information seeking is directly proportional to the amount of uncertainty regarding the "correctness" of the alternatives (Lanzetta, 1963; Lanzetta & Driscoll, 1968). Lanzetta's research was guided by Berlyne's theory of curiosity (1960), which assumes that uncertainty has drive properties and postulates information seeking as the primary instrumental means for reducing the drive. Now, if achievement behavior is an instance of information seeking in general, we would expect that the more similar the subjective prior probabilities of possessing the various ability levels (i.e., the probabilities held by the person prior to task performance), the greater the uncertainty, and therefore the greater the tendency to perform a relevant task.

A recent study by Trope (Note 2) tested these propositions. Subjects first performed two tasks, each pertaining to a different ability. In one of the experimental conditions, the results were predetermined so as to induce very little uncertainty about their standing on one ability (low-uncertainty ability) and a great deal of uncertainty about their standing on the other ability (high-uncertainty ability). Subjects could then decide whether to continue working on tasks that were relevant to the low-uncertainty ability or on tasks that were relevant to the high-uncertainty ability. As expected, subjects preferred the latter. These tasks were equally attractive in the conditions in which prior results induced the same amount of uncertainty, low or high, about the two abilities.

In the preceding discussion of uncertainty and diagnosticity, the ability dimension was considered as a whole. However, uncertainty and diagnosticity may vary from one zone to another on the ability dimension. Thus, the prior probabilities of the ability levels within one zone may be *higher* and more nearly *equal* than the prior probabilities of the ability levels within another zone. To illustrate, suppose that two zones are considered, the lower half and the upper half of the ability dimension, and imagine a person who is almost certain that he or she is not in the lower zone of the ability dimension but is highly uncertain about his or her location within the upper zone. In probabilistic terms, compared to the ability levels within the lower zone, the ability levels within the upper zone have high and equal prior probabilities. The opposite case would obtain for an individual who is quite sure of not belonging to the upper zone but is unsure as to his or her standing within the lower zone of the ability dimension. That is, the prior probabilities within the lower zone are high and equally distributed among the ability levels. Clearly, these two persons have the same amount of overall uncertainty, but their uncertainty is distributed differently between the two zones: Most of one person's uncertainty is concentrated within the lower zone, whereas most of the other's uncertainty is concentrated within the upper zone.

The amount of information these hypothetical persons can gain from task performance depends on the accuracy with which the task can diagnose (or discriminate between) ability levels within each of the zones. Tasks of equal overall diagnosticity (i.e., equal total amount of variation in performance across all ability levels) may differ in the way their diagnosticity is distributed among the zones. The task may be (a) more diagnostic in the lower zone, producing greater differences in performance among ability levels within the lower zone than among ability levels within

the upper zone (e.g., a task in which performance is a monotone negatively accelerated function of ability); (b) it may be equally diagnostic in the two zones (e.g., a task in which performance is a linear function of ability); or (c) it may be more diagnostic in the upper zone (e.g., a task in which performance is a monotone positively accelerated function of ability).

One can maximize information gain or uncertainty reduction by choosing a task that is most diagnostic at the zone where one's uncertainty is concentrated. That is, information gain depends on the similarity between the distributions of uncertainty and diagnosticity across the various zones of the ability dimension. Hence, the self-assessment notion would predict that a person whose uncertainty is concentrated in the lower zone will be more interested in a task that is more diagnostic in the lower zone, less interested in a task that is equally diagnostic in the two zones, and least interested in a task that is more diagnostic in the upper zone. A person whose uncertainty is concentrated in the upper zone will display the reverse pattern of preferences.

The foregoing analysis was tested in two experiments. Experiment 1 manipulated two variables. The first variable was the subjects' distribution of uncertainty, that is, subjects' uncertainty was concentrated either in the lower zone or upper zone. The second variable was the distribution of task diagnosticity, that is, tasks were more diagnostic within the lower or upper zone or were equally diagnostic within the lower and upper zones. The tasks were equated on overall diagnosticity and difficulty. The second experiment manipulated these last two variables as well as the distribution of diagnosticity. In this experiment, the distribution of uncertainty was inferred from subjects' own beliefs about their ability rather than being experimentally manipulated.

Experiment 1

Method

Subjects

The subjects were 60 male candidates for an officers training school in the Israeli Defense Forces. They participated in the experiment during their stay in an officer selection unit located in an army base. All of the subjects were in regular army service, had completed high school education, and were 18–21 years old.

Procedure and Manipulations

Subjects participated individually in two experimental sessions. In the first, they performed a cognitive task and received predetermined feedback that was designed to manipulate the uncertainty distribution over the relevant ability dimension. In the second session, subjects were presented with three additional tasks varying in diagnosticity distribution over the ability dimension. Subjects then indicated their preferences among these tasks and answered several questions that served as manipulation checks.

At the beginning of the first session, subjects were told that they were taking part in a survey on conceptual ability carried out by the Department of Psychology at the Hebrew University. The experimenter (a civilian) emphasized that subjects' results were anonymous and would not be given to any army authority. Subjects were then informed that their first task would be to perform a test of conceptual ability called "common features test." The test consisted of 20 items in each of which subjects had to name a feature or concept that most precisely fitted a group of four words (e.g., fog, snow storm, electricity blackout, military coup).

After completing the test, subjects were handed a page containing information about conceptual ability and the scoring of the common features test. Subjects read:

> Conceptual ability is manifested in the ability to discover and generate conceptions and rules underlying sets of items of information. Extensive research has shown that young adults with high school education can be classified into three levels of conceptual ability: a relatively low level, a relatively high level, and an intermediate level of conceptual ability which lies halfway between the low and the high levels.

To minimize differences among subjects in prior beliefs about their ability, the statement further noted, "Conceptual ability is not measured adequately by standard intelligence tests. However, previous research has shown that approximately one third of the population of young adults with high school education has each of the three levels of conceptual ability." Subjects were told that the tester examined each answer and determined according to several well-established criteria whether it reflected a low level, an intermediate level, or a high level of conceptual ability.

Manipulation of uncertainty distribution. While the subject was reading this information, the experimenter pretended to score his test, ostensibly using an impressive table of norms and categories. The experimenter presented the results on a results sheet.

Table 1
Experiment 1: Reported Mean Percentages of Correct Items as a Function of Ability Level, Uncertainty Distribution, and Diagnosticity Distribution

Ability level	Uncertainty distribution: Descending — Diagnosticity distribution: Descending	Descending — Constant	Descending — Ascending	Ascending — Descending	Ascending — Constant	Ascending — Ascending
Low	29.8	37	44.2	15.8	23	30.2
Intermediate	65.8	57	48.2	51.8	43	34.2
High	69.8	77	84.2	55.8	63	70.2

In one condition, the results sheet indicated that out of the 20 answers, 9 (45%) reflected low ability, 9 (45%) intermediate ability, and 2 (10%) high ability. These results were designed to induce high probabilities of the low and the intermediate levels and a low probability of the high level. In the other condition, the results sheet indicated that 2 answers (10%) reflected low ability, 9 (45%) intermediate ability, and 9 (45%) high ability. These results were designed to induce a low probability of the low level and high probabilities of the intermediate and high levels. The former condition is referred to as *descending uncertainty* condition, meaning that most of the uncertainty was concentrated in the low half of the ability scale; the latter condition is referred to as *ascending uncertainty*, meaning that most of the uncertainty was concentrated in the upper half of the ability scale. Thirty subjects were randomly assigned to each condition.

After receiving the results, subjects were escorted to an adjacent room where another experimenter conducted the second session. This experimenter told subjects that he would like them to work on a test of conceptual ability called "classification test" and that they would receive their results. Subjects were told that each item in the test presented a list of objects or events that had to be assigned into four mutually exclusive classes. It was further indicated that there was only one correct classification and that the score on the test was the percentage of items correctly solved. Subjects were informed that the test consisted of three subtests, 30 items in each, with the kind of elements to be classified—words, figures, or letters—varying from subtest to subtest.

Manipulation of diagnosticity distribution. The subject was then shown a visual display of fictitious results that the low, intermediate, and high ability groups purportedly achieve at each of the subtests. Each subtest was represented by a diagram that displayed on a horizontal 0% to 100% scale the mean performance of each ability group and the interval within which the performance of 90% of the group fell. Table 1 presents the means for each subtest in the descending and ascending uncertainty conditions. All the intervals were 30% wide and symmetric around the means.

The following constraints were employed in deriving these means. In order to hold overall diagnosticity constant, the difference between the means of the lowest and highest ability groups was set at 40% in all of the subtests. Another set of constraints was designed to manipulate diagnosticity distribution. In subtests that were more diagnostic at the lower half of the ability scale, referred to as *descending diagnosticity* subtests, the difference between the low and intermediate levels was set nine times as great as the difference between the intermediate and high levels. In the subtests that were more diagnostic at the upper half of the ability scale—*ascending diagnosticity* subtests—the magnitude of these two differences was reversed. In subtests that were equally diagnostic at the lower and upper halves of the scale—*constant diagnosticity* subtests—these two differences were of the same magnitude. The third constraint was on expected performance at the subtests and was designed to hold perceived difficulty constant. Expected performance on a test was defined as the overall mean of the performances of the three ability levels, each weighted by the probability of having the corresponding ability level. Obviously, these probabilities should be different in the two uncertainty distribution conditions. In fact, pretesting indicated that the probabilities assessed by subjects were quite close to the proportions of answers in the common features test that purportedly reflected each ability level. Hence, the prior probabilities of the low, intermediate, and high ability levels were assumed to be .45, .45, and .10, respectively, in the descending uncertainty condition and .10, .45, and .45, respectively, in the ascending uncertainty condition. Using these probabilities, expected performance at the three subtests was set at 50% in the two uncertainty distribution conditions.

The order of presentation of the diagrams for the three subtests was varied across subjects according to a 3×3 Latin square design.

Measures. Subjects were informed that they were to choose a total of 28 items to work on and that they could decide how many items would come from each subtest. After indicating their selection of items, subjects were asked to fill out an evaluation questionnaire while the experimenter was ostensibly pre-

paring the test material. The first set of questions in this questionnaire pertained to the three subtests of the classification test. Subjects rated five questions on 7-point scales: (a) their willingness to work on each subtest, (b) their interest in finding out how they would perform at each subtest, (c) the accuracy with which each subtest could discriminate between the low and the intermediate levels of ability, (d) the accuracy with which each subtest could discriminate between the intermediate and the high levels of ability, and (e) the difficulty of each subtest. Questions a and b, together with the measure of item selection, served as measures of task preference. Questions c and d served as checks on the manipulation of diagnosticity distribution. Question e was designed to check the assumption that the subtests were equated on perceived difficulty. As a check on the manipulation of uncertainty distribution, subjects were asked to rate on three 11-point scales (ranging from 0 to 100) the chances of their having each of the three ability levels. After subjects had completed the questionnaire, the experiment was terminated. The experimenter thoroughly debriefed the subjects, checking for suspicion and explaining the actual purpose of the study and the reason for the deceptions.

Results and Discussion

Manipulation Checks

The manipulation of uncertainty distribution (i.e., the feedback from the common features test) had a strong effect on subjects' assessments of the probabilities of having each ability level. In the descending uncertainty condition, the means of the probabilities of the low, intermediate, and high levels were .36, .43, and .21, respectively, and in the ascending uncertainty condition, they were .10, .43, and .47. An analysis of variance (Uncertainty Distribution × Ability Level) yielded a highly significant interaction, $F(2, 116) = 65.07$, $p < .001$, indicating that in the descending uncertainty condition, the difference between the low and intermediate levels was small and insignificant and the difference between the intermediate and high levels large and significant ($p < .001$), whereas in the ascending uncertainty condition, the former difference was large and significant ($p < .001$), and the latter small and insignificant. Hence, as intended, in the descending uncertainty condition most of the uncertainty concerned the low and intermediate levels (i.e., the lower half of the ability scale), whereas in the ascending uncertainty condition most of the uncertainty concerned the intermediate and high levels (i.e., the upper half of the ability scale). It should be noted, however, that although the two conditions differed in the expected direction, the uncertainty distribution in the ascending uncertainty condition was more in line with the feedback from the common features test than it was in the descending uncertainty condition. In the latter condition, therefore, the uncertainty was more evenly distributed over the ability scale.

The first two rows in Table 2 indicate that the manipulation of diagnosticity distribution was successful. As expected, descending diagnosticity tests were perceived as the most accurate and the ascending diagnosticity tests as the least accurate in discriminating between low and intermediate ability. On accuracy in discriminating between intermediate and high ability, the ordering of the subtests was reversed. Analyses of variance (Uncertainty Distribution × Diagnosticity Distribution) supported these observations, yielding highly significant effects of diagnosticity distribution on ratings regarding accuracy of discrimination between low and intermediate ability, $F(2, 116) = 68.27$, $p < .001$, and between intermediate and high ability, $F(2, 116) = 52.08$, $p < .001$. No other source of variation was significant in these analyses. Finally, as intended, the perceived difficulty of the subtests (Row 3, Table 2) did not significantly vary as a function of the experimental variables. It seems, then, that the fact that expected performance was kept constant [1] made the subtests appear equally difficult.

Choice and Preference Among Subtests

The number of items selected from each subtest served as the main dependent variable. The pertinent means in Figure 1 indicate that the manipulation of uncertainty distribution produced two opposite combinations of items from the three subtests. As pre-

[1] As noted above, the subjective probabilities of the ability levels in the descending uncertainty condition somewhat deviated from the reported proportions of items that purportedly indicated the various ability levels. Even if these probabilities are used to calculate expected performance, however, the differences among the subtests remain extremely small.

Table 2
Experiment 1: Mean Ratings of Tasks as a Function of Uncertainty Distribution and Diagnosticity Distribution

Measure	Uncertainty distribution: Descending; Diagnosticity distribution: Descending	Descending; Constant	Descending; Ascending	Ascending; Diagnosticity distribution: Descending	Ascending; Constant	Ascending; Ascending
Accuracy in discriminating between low and intermediate ability	5.87	4.50	2.50	5.57	4.20	2.50
Accuracy in discriminating between intermediate and high ability	2.90	4.80	5.70	2.90	4.20	6.27
Difficulty	3.45	3.38	3.99	3.37	3.07	3.80
Willingness to work on subtest	5.23	4.50	3.53	3.63	3.93	5.47
Interest in finding out results	5.93	5.33	4.97	4.37	4.80	6.27

dicted, in the descending uncertainty condition the greatest number of items came from the descending diagnosticity subtest and the smallest from the ascending diagnosticity subtest, whereas in the ascending uncertainty condition the results were reversed. This dependency of choice among subtests on uncertainty distribution was reflected in a strong Uncertainty Distribution × Diagnosticity Distribution interaction effect, $F(2, 116) = 13.56$, $p < .001$, which was the only significant source of variation.

The same picture emerges from the ratings of desire to work on the subtests (Table 2, row 4) and the ratings of interest in finding out the results of performance at the subtests [2] (Table 2, row 5). Analyses of variance of these ratings yielded highly significant Uncertainty Distribution × Diagnosticity Distribution interaction effects ($p < .001$), with no other effect being significant. It is particularly interesting that the main effects of uncertainty distribution were insignificant ($F < 1$), meaning that subjects in the two uncertainty distribution conditions had the same overall interest in working on the subtests and in finding out how they would perform.

Finally, it seems that the slopes relating item selection and the preference ratings to diagnosticity distribution are steeper in the ascending uncertainty condition than in the descending uncertainty condition. The F ratios corresponding to these comparisons were insignificant, but it is worth noting that this difference in slopes may be due to the fact, indicated earlier, that in the descending uncertainty condition subjects' uncertainty was more evenly distributed between the two halves of the ability scale.

In conclusion, subjects' choice behavior suggests that they systematically employed a strategy of maximizing information about their ability. The descending uncertainty and the ascending uncertainty groups did not differ in this respect. First, the two groups expressed the same overall amount of interest in the ability-relevant tasks. Second, both groups were interested in working on a task to the extent that its diagnosticity distribution fitted their uncertainty distribution. Thus, subjecs were most interested in a task that was highly diagnostic at the ability zone where their uncertainty was concentrated, irrespective of the location of this zone on the ability dimension.

[2] The correlations between the three dependent variables (the two ratings and items selection) *within* conditions ranged from .42 to .79, with a mean of .56.

Experiment 2

This experiment manipulated three variables: overall diagnosticity, diagnosticity distribution, and difficulty. The first and last of these variables were held constant in Experiment 1. Distribution of uncertainty, which was experimentally manipulated in Experiment 1, was inferred in this experiment from subjects' assessments of their own ability level. It was assumed that the higher the perceived ability, the higher the zones on the ability dimension within which the uncertainty was concentrated.

Method

Subjects

The subjects were 54 male soldiers who belonged to the same population from which the subjects of Experiment 1 were drawn.

Procedure and Manipulations

Subjects participated in the experiments in groups ranging in size from three to five. The experimenter stated that the purpose of the study was to investigate opinions and feelings about ability tests. As in Experiment 1, subjects were assured that their answers were anonymous. The experimenter then handed the subjects a booklet titled *Survey of Opinions about Tests of Mental Concentration.* In the booklet subjects read that they would receive information about several tests of mental concentration in all of which the testee had to solve a series of items within a certain time limit. It was indicated that the tests varied in the type of item they included —numerical, verbal, or figural—and in emphasis on perception, memory, or reasoning. Subjects were then informed that extensive research had shown that young adults with high school education could be classified into seven levels of mental concentration, with approximately 14% of this population belonging to each of the levels.

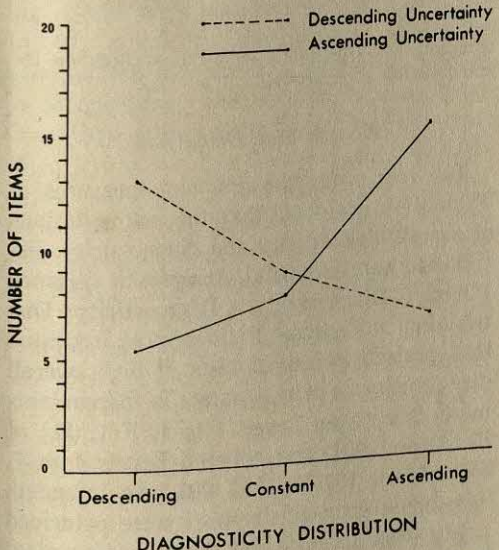


Figure 1. Experiment 1: Mean number of items selected from each test as a function of uncertainty distribution (descending uncertainty vs. ascending uncertainty) and diagnosticity distribution (descending vs. constant vs. ascending).

At this point, subjects were asked to assess their own level of mental concentration. Subjects were given a seven-interval scale with labels from Level 1 (lowest ability) to Level 7 (highest ability) and were asked to indicate their level of ability.

The booklet then indicated that previous research had obtained a large amount of data on the results that are typically achieved by people of different levels of mental concentration. An example of such data for one of the tests was then presented. Each ability level was represented by a bar whose height indicated the average percentage of correctly solved items. The bars, from left to right, indicated the lowest to highest ability levels, and were labeled beneath from 1 to 7, respectively. In this example, the average of Level 1 was 35%, which increased linearly with ability level up to 65% in Level 7. Subjects were told that in this test as well as in the other tests, the performance of 90% of each ability level lay between 4% below the average to 4% above the average.

Manipulation of test characteristics. After completing the example, each subject considered six tests varying in overall diagnosticity and distribution of diagnosticity. Each test was represented by a bar graph, as in the example. In tests of high overall diagnosticity the difference in performance between Level 1 and Level 7 (51%) was three times as great as the corresponding difference in tests of low overall diagnosticity (17%). The three tests within each level of overall diagnosticity varied in distribution of diagnosticity. Thus, the difference between Levels 1 and 7 (or overall diagnosticity) was divided such that the difference between Levels 1 and 4 (or diagnosticity at the lower half of the ability scale) was either greater, smaller, or equal to the difference between Levels 4 and 7 (or diagnosticity at the upper half of the ability scale). In the descending diagnosticity test the difference between Levels 1 and 4 was 7.5 times greater than the difference between Levels 4 and 7. In the ascending diagnosticity test the magnitude of these differences was reversed. In the constant diagnosticity test these differences were of the same magnitude. Within each half of the ability scale, performance was linearly related to ability level. Difficulty was manipulated by varying expected performance on the tests, that is, the overall mean of

Table 3
Experiment 2: Reported Mean Percentages of Correct Items as a Function of Ability Level, Expected Performance, Overall Diagnosticity, and Diagnosticity Distribution

% expected performance	Ability level	Overall diagnosticity: Low — Diagnosticity distribution: Descending	Low — Constant	Low — Ascending	High — Descending	High — Constant	High — Ascending
40	1	28.72	31.50	34.30	6.14	14.50	22.86
	2	33.72	34.33	34.97	21.14	23.00	24.86
	3	38.72	37.16	35.64	36.14	31.50	26.86
	4	43.72	40.00	36.30	51.14	40.00	28.86
	5	44.38	42.38	41.30	53.14	48.50	43.86
	6	45.05	45.66	46.30	55.14	57.00	58.86
	7	45.72	48.50	51.30	57.14	65.00	73.86
50	1	38.72	41.50	44.30	16.14	24.50	32.86
	2	43.72	44.33	44.97	31.14	33.00	34.86
	3	48.72	47.16	45.64	46.14	41.50	36.86
	4	53.72	50.00	46.30	61.14	50.00	38.86
	5	54.38	52.83	51.30	63.14	58.50	53.86
	6	55.05	55.66	56.30	65.14	67.00	68.86
	7	55.72	58.50	61.30	67.14	75.50	83.86
60	1	48.72	51.50	54.30	26.14	34.50	42.86
	2	53.72	54.33	54.97	41.14	43.00	44.86
	3	58.72	57.16	55.64	56.14	51.50	46.86
	4	63.72	60.00	56.30	71.14	60.00	48.86
	5	64.38	62.83	61.30	73.14	68.50	63.86
	6	65.05	65.66	66.30	75.14	77.00	78.86
	7	65.72	68.50	71.30	77.14	85.50	93.86

performance across the seven ability levels. There were three levels of expected performance: 40%, 50%, and 60%. The percentages presented to subjects in the bar graphs (see Table 3) were derived by using these constraints.

Expected performance was manipulated between subjects. Twenty subjects were randomly assigned to each of the three expected performance conditions. Each group considered six tests (varying in overall diagnosticity and diagnosticity distribution) with order of presentation varying across subjects according to a 6×6 Latin square design.

Measures. Subjects rated seven questions on 11-point scales: (a) their willingness to work on each test; (b) the attractiveness of each test; (c) their interest in finding out how they would perform at each test; (d) the accuracy with which each test could discriminate between Levels 1, 2, 3, and 4 of mental concentration; (e) the accuracy with which each test could discriminate between Levels 4, 5, 6, and 7 of mental concentration; (f) the accuracy with which each test could reveal their level of mental concentration; and (g) the difficulty of each test. Questions a to c served as measures of task preference. Questions d to f served as checks on the manipulations of overall diagnosticity and diagnosticity distribution, and Question g served as a check on the manipulation of difficulty.

Results and Discussion

The results for the first two measures in Table 4 clearly show that the manipulations of overall diagnosticity and diagnosticity distribution were successful. Analyses of variance (Overall Diagnosticity $\times$ Diagnosticity Distribution $\times$ Expected Performance) indicated that subjects perceived tests of high overall diagnosticity as more accurate in discriminating both between Levels 1 to 4, $F(1, 51) = 69.63$, $p < .001$, and between Levels 4 to 7, $F(1, 51) = 152.73$, $p < .001$. As intended, descending diagnosticity tests were perceived as the most accurate and the ascending diagnosticity tests as the least accurate in discriminating between Levels 1 to 4, $F(2, 102) = 155.95$, $p < .001$, whereas on accuracy in discriminating between Levels 4 to 7, the order-

Table 4
Experiment 2: Mean Ratings of Tests as a Function of Overall Diagnosticity, Diagnosticity Distribution, and Expected Performance

Measure	% expected performance	Overall diagnosticity: Low — Diagnosticity distribution: Descending	Low — Constant	Low — Ascending	High — Descending	High — Constant	High — Ascending
Accuracy in discriminating between Levels 1, 2, 3, and 4	40	5.67	4.44	2.56	7.72	7.33	3.94
	50	5.17	4.17	1.28	8.44	8.06	2.06
	60	5.33	4.61	2.06	8.28	7.78	2.50
	Combined	5.39	4.41	1.96	8.15	7.72	2.83
Accuracy in discriminating between Levels 4, 5, 6, and 7	40	2.89	4.83	5.67	4.28	7.44	8.28
	50	.94	4.44	5.89	1.78	8.06	8.89
	60	2.33	4.50	5.50	3.50	7.89	8.67
	Combined	2.06	4.59	5.69	3.19	7.80	8.61
Accuracy in determining one's own ability level	40	3.83	5.06	6.00	4.28	7.11	8.00
	50	1.78	3.50	4.78	2.83	7.72	8.28
	60	4.56	5.17	5.67	4.83	8.17	8.17
	Combined	3.39	4.57	5.48	3.98	7.67	8.15
Difficulty	40	4.67	4.11	5.50	4.78	4.28	4.50
	50	2.78	3.50	3.11	4.06	5.22	3.83
	60	2.89	2.78	3.11	2.39	3.33	3.72
	Combined	3.44	3.46	3.96	3.74	4.28	4.02
Index of test preference	40	5.52	5.76	6.59	5.87	7.17	7.91
	50	4.37	4.80	5.81	4.74	7.26	8.19
	60	4.94	5.94	6.81	6.09	7.87	8.67
	Combined	4.94	5.50	6.41	5.57	7.43	8.25

ing of the tests was reversed, $F(2, 102) = 112.07$, $p < .001$. Furthermore, since the *manipulated* differences among tests varying in diagnosticity distribution were greater when their overall diagnosticity was high than when it was low (see Table 3), the two variables were expected and found to have interactive effects on perceived accuracy in discriminating between Levels 1 to 4, $F(2, 102) = 18.17$, $p < .001$, and between Levels 4 to 7, $F(2, 102) = 13.06$, $p < .001$.

While expected performance did not have any effect on the ratings of perceived diagnosticity, it had a marginally significant effect on perceived difficulty (see fourth measure, Table 4), indicating that the higher the expected performance at the test, the easier the test, $F(2, 51) = 2.87$, $p < .07$. The only other significant effect on perceived difficulty was a Diagnosticity Distribution × Expected Performance interaction, $F(4, 102) = 2.94$, $p < .05$, which was unexpected and hard to interpret.

Prior to discussing the other measures, subjects' reports about their level of mental concentration should be considered. Quite surprisingly, the range of reported ability levels was extremely narrow. No subjects reported an ability below Level 4; only 2 subjects reported Level 4, 25 reported Level 5, 21 reported Level 6, and only 6 reported Level 7. These data suggest that the uncertainty of most of the subjects was confined to the high ability levels. It follows that subjects should rate the accuracy with which a test can determine their *own* ability according to the degree to which it can discriminate between the high levels, that is, the ascending diagnosticity tests should be rated highest and the descending diagnosticity tests lowest. The pertinent data in Table 4 (third measure) fully corroborated this predicted effect of diagnosticity distribution, $F(2, 102) = 49.26$, $p < .001$. Furthermore, as intended, ratings of the test's accuracy in determining one's own ability increased with overall diagnosticity, $F(1, 51)$

$= 76.70$, $p < .001$, and the effect of diagnosticity distribution on ratings of tests of high overall diagnosticity was stronger than its effect on ratings of tests of low overall diagnosticity, $F(2, 102) = 20.39$, $p < .001$.

Turning to subjects' preferences among the tests, we should note that the three measures —willingness to work on the test, attractiveness of the test, and interest in finding out the results—correlated highly across subjects for each of the six tests. The mean of the 18 correlations (3 intercorrelations for each of the six tests), using Fisher's r-to-z transformations, was .77. The three measures therefore were averaged (see Table 4, last measure). The self-assessment notion would predict that preferences among tests should be closely related to judgments about their accuracy in determining one's own ability. The means of the index of test preference substantiated this prediction. First, tests of high overall diagnosticity were preferred to tests of low overall diagnosticity, $F(1, 51) = 40.55$, $p < .001$. Second, there was a strong effect of diagnosticity distribution, $F(2, 102) = 26.99$, $p < .001$, indicating that tests discriminating mainly between high ability levels (ascending diagnosticity tests) were preferred to tests that discriminated equally between all levels (constant diagnosticity tests), which in turn were preferred to tests that discriminated mainly between low ability levels (descending diagnosticity tests). Third, this effect of diagnosticity distribution was more pronounced when overall diagnosticity was high than when it was low,[3] $F(2, 102) = 7.44$, $p < .001$. In none of these analyses did expected performance (or difficulty) have a significant effect on ratings. Finally, the close relationship between test preference and perceived accuracy in determining one's ability was reflected in the correlations between the two measures across subjects for each of the tests. All these correlations were highly significant, ranging from .415 to .674, with a mean of .55 (using Fisher's r-to-z transformations).

The effects of uncertainty distribution on task preference could not be fully manifested in this experiment because, as noted above, the ability levels that the subjects reported implied that their uncertainty was concentrated in the upper region of the ability dimension. However, this may be less true for subjects who reported the more moderate levels (i.e., Levels 4 and 5). It seems plausible that these subjects assigned higher probabilities to the low ability levels than did subjects who reported extremely high levels (i.e., Levels 6 and 7). In that case, the former group's uncertainty was more evenly distributed over the ability scale. Such uncertainty distribution should act to increase their interest in tests that can discriminate between low ability levels, namely, descending and constant diagnosticity tests. More precisely, the self-assessment notion would predict that the preference for ascending diagnosticity tests over constant and descending diagnosticity tests would be more pronounced among extremely high ability subjects than it would be among moderately high ability subjects.

Figure 2 presents means of the index of test preferences for subjects whose perceived ability was below the median (Levels 4 and 5) and for subjects whose perceived ability was above the median (Levels 6 and 7). It can be seen that the preferences of the extremely high ability group increased from the descending to the ascending diagnosticity test much more sharply than did the preferences of the moderately high ability group, $F(2, 104) = 14.62$, $p < .001$, for the Perceived Ability × Diagnosticity Distribution interaction. Moreover, this difference between the two perceived ability groups was greater with respect to tests of high overall diagnosticity than with respect to tests of low overall diagnosticity, $F(2, 104) = 3.03$, $p < .06$, for the Overall Diagnosticity × Perceived Ability × Diagnosticity Distribution interaction. Figure 2 also suggests an interaction between perceived ability and overall diagnosticity, $F(1, 52) = 8.23$, $p < .025$, which, unlike the other interactions involving perceived ability, was not derived from the self-assessment notion. This interaction suggests that among extremely high ability subjects the preference for high diagnosticity tests over low diagnosticity tests

[3] The same significant effects were obtained in separate analyses of each of the three measures of test preference.

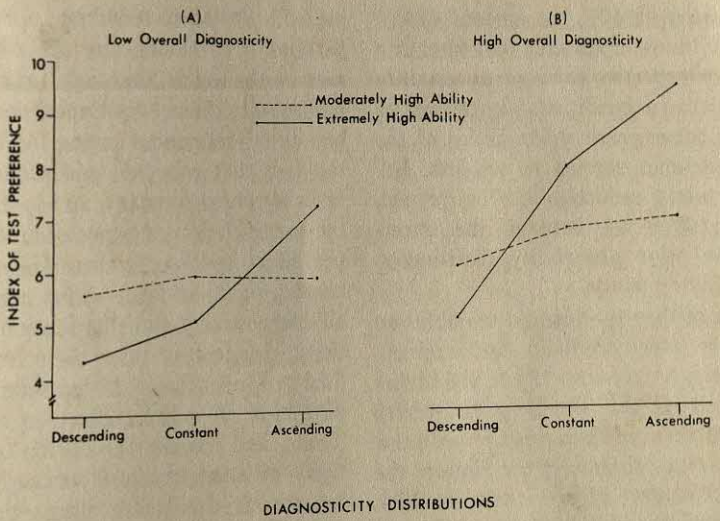


Figure 2. Experiment 2: Means of the index of test preference as a function of perceived ability (moderately high ability vs. extremely high ability), overall diagnosticity (low vs. high), and diagnosticity distribution (descending vs. constant vs. ascending).

was more pronounced than among moderately high ability subjects. It should be noted, however, that perceived ability did not have a significant main effect, that is, it did not affect overall interest in the tests.

General Discussion

The present experiments tested the idea that people choose among tasks according to the amount of ability-relevant information they expect to gain. Two general determinants of information gain were considered: the person's initial knowledge about his or her ability and the informational properties of the task. The former determinant was represented by distribution of uncertainty and the latter by overall diagnosticity and its distribution. On the whole, the effects of these variables on task preference supported the hypotheses and can be summarized as follows. First, as in previous research, tasks of high overall diagnosticity were preferred to those of low overall diagnosticity (Trope, 1975; Trope & Brickman, 1975; Zuckerman et al., in press; Buckert et al., Note 1). Second, with overall diagnosticity held constant, preferences were determined jointly by uncertainty distribution and diagnosticity distribution, that is, subjects preferred tasks that could most accurately diagnose within the zone where their uncertainty was concentrated. Third, the joint effect of uncertainty and diagnosticity distributions was more pronounced when overall diagnosticity was high than when it was low, as overall diagnosticity was manipulated as a multiplier of diagnosticity in each zone.

These results seem to reflect a preference for tasks that can add to previously acquired knowledge about one's own ability. Thus, when subjects could infer from previous performance that they were most likely to have, for example, one of the relatively high ability levels, they wanted to work on a task that could determine with precision which of these levels they actually had. These subjects were least interested in a task that made precise discriminations only between the low ability levels, which they were unlikely to have. Such a task would only reaffirm that they possessed one of the high ability levels.

Moreover, the observed task preferences conform to what a normative model of uncertainty reduction would prescribe. In such a model, each outcome revises prior probabilities of the ability levels into posterior probabilities (see Trope & Brickman, 1975). From each of these posterior probability distributions, it is possible to derive, as in information

theory (Attneave, 1959), the amount of uncertainty conditional upon each outcome. One can obtain an expected value of uncertainty by summing these conditional uncertainties across all outcomes, with probabilities of the respective outcomes serving as weights. Expected uncertainty reduction can be defined, then, as the difference between the uncertainty in the prior probability distribution and expected uncertainty.

The effects of the experimental variables on choice can be traced to their effects on expected uncertainty reduction. First, the higher the diagnosticity of the outcomes, the greater the amount of revision from prior to posterior probabilities, and therefore the lower the expected uncertainty. Hence, tasks of high overall diagnosticity reduce more uncertainty than tasks of low overall diagnosticity. Second, the distribution of diagnosticity over the zones of ability determines the distribution of diagnosticity over performance outcomes. Assuming that performance is a positive monotone function of ability, it follows that diagnosticity is related to ability and performance outcomes in a similar manner. Thus, if the task is more diagnostic of low ability levels, then relatively low outcomes are more diagnostic, whereas if the task is more diagnostic of high ability levels, then relatively high outcomes are more diagnostic. Third, the distribution of uncertainty affects the probabilities of the outcomes: The higher the zone of ability within which uncertainty is concentrated, the higher the probability of relatively high outcomes and the lower the probability of relatively low outcomes. Now, since the uncertainty reduced by each outcome is weighted by the probability of the outcome, expected uncertainty reduction is high to the extent that outcomes that reduce much uncertainty have high probability and outcomes that reduce little uncertainty have low probability. Clearly, this requirement is fulfilled by tasks that are highly diagnostic in the zone where uncertainty is concentrated. The fact that subjects preferred such tasks suggests that they were interested in maximizing expected uncertainty reduction.

In and of itself, expected performance or task difficulty does not necessarily affect expected uncertainty reduction, nor did subjects perceive it as having any effect on the accuracy with which the tasks could diagnose their ability. Hence, in Experiment 1 subjects had clear preferences among the tasks despite the fact that expected performance at these tasks was held constant, and in Experiment 2 the manipulation of expected uncertainty did not affect task preference. However, in the absence of direct information regarding overall diagnosticity and distribution of diagnosticity, people may infer them from task difficulty. Under these circumstances difficulty may have an indirect effect on choice. Thus, Trope and Brickman (1975) argued that tasks with intermediate success probabilities are perceived as more diagnostic and therefore more attractive than tasks with more extreme probabilities (see also Meyer, Folkes, & Weiner, 1976). Furthermore, it seems probable that easy tasks are perceived as more diagnostic of low ability levels (i.e., poor performance is more diagnostic than good performance) and difficult tasks as more diagnostic of high ability levels (i.e., good performance is more diagnostic than poor performance). Hence, to maximize information gain or uncertainty reduction, people with low perceived ability (whose uncertainty is concentrated in the low ability levels) should prefer relatively easy tasks, whereas people with high perceived ability (whose uncertainty is concentrated in the high ability levels) should prefer more difficult ones. The same logic applies to the case of two outcomes, success and failure, and two levels of ability, low and high. As difficulty increases, the perceived diagnostic value of success becomes greater than that of failure, and since the subjective probability of success increases with perceived ability, people with high perceived ability should prefer more difficult tasks than people with low perceived ability (see Zuckerman et al. for corroborating results). Poor performance shifts the uncertainty to lower ability levels and good performance shifts it to higher ability levels. Hence, in the former case, the person will attempt increasingly easier tasks, whereas in the latter case, he or she will attempt increasingly harder tasks. These well-documented

shifts in "level of aspiration" (Lewin, Dembo, Festinger, & Sears, 1944) are interpreted here as attempts to keep the amount of new ability-relevant information at a maximum.

The present findings are relevant to the view that self-enhancement and self-protective needs (Bradley, 1978; Jones, 1973) determine task choice. Kukla (1978) has recently suggested an attributional theory of choice based on the assumption that the tendency to work on a task increases with the level of ability that performance is expected to imply. A task will be attractive if it can demonstrate one's high ability, but it will be avoided if it may betray one's low ability. With regard to the present study, this approach would predict that subjects who suspect that they have one of the low ability levels will be less interested in working on the ability-relevant tasks than will subjects who believe that they have one of the high ability levels. However, the data did not show any trace of such a main effect of perceived ability (or uncertainty distribution). Furthermore, self-serving biases should produce a preference for a task that is more diagnostic of high ability levels over a task that is more diagnostic of low ability levels. More specifically, persons with high perceived ability should be attracted to the former task because it can demonstrate unequivocally that they have high ability, and persons with low perceived ability should *avoid* the latter task because it may disclose that they have low ability. Again, there was very little evidence for such asymmetry. Subjects chose the task that was diagnostic at the zone of ability within which their uncertainty was concentrated, even when performance of this task could give very unflattering results. Finally, it is worth noting that perceived ability did not produce any self-serving distortions of the informational properties of the tasks. Specifically, subjects with high perceived ability did not overestimate the accuracy with which performance could reveal high ability; nor did subjects with low perceived ability underestimate the accuracy with which performance could reveal poor ability. The judgments of both groups were appropriately related to the manipulated values of the informational characteristics of the tasks.

There was one result, a Perceived Ability × Overall Diagnosticity interaction, which *was* consistent with the self-enhancement view. This interaction indicated that the preference for the high overall diagnosticity tasks was more pronounced among subjects with extremely high perceived ability than among subjects with moderately high perceived ability. This result suggests that subjects' choice behavior was not entirely free from self-serving biases. It should be remembered, however, that in Experiment 2, where this interaction was obtained, perceived ability was varied by subject selection, which leaves open the possibility of a correlated variable. Achievement motive, which was found to be correlated with perceived ability (Kukla, 1978), may be such a variable; previous research by Trope (1975) has shown that high-achievement-motive people are more interested in ability-relevant information than low-achievement-motive subjects. In addition, while Zuckerman et al. obtained a similar effect of perceived ability, Buckert et al. (Note 1) did not. Clearly, more research on this problem is needed before definite conclusions can be reached.

Even if self-serving biases were entirely absent in the data, it would not have meant that our subjects were affectively indifferent to the ability-relevant feedback they received or expected to receive. Extrapolating from previous research on reactions to self-evaluation (see reviews by Bradley, 1978; Jones, 1973; Shrauger, 1975) it seems plausible that the subjects who learned that they had one of the low ability levels were distressed by the feedback and anticipated even greater distress should the next task diagnose a low level of ability. However, the present results cast doubt on the assumption that such affective reactions shape choice strategy. The desire to assess one's ability accurately seems a more powerful determinant of choice, at least under the circumstances investigated in this study. It remains for future research to determine the generalizability of this conclusion to different circumstances. One limiting condition is suggested by the fact that subjects in the present study were assured that their future

performance would remain anonymous. It is possible that when performance is expected to be public, the enhancement of self-esteem as well as social esteem may assume greater importance than accurate self-assessment. In such a situation, people may actually choose to work on tasks that are expected to diagnose their strengths but not their weaknesses.

Reference Notes

1. Buckert, U., Meyer, W. U., & Schmalt, H. D. *The effects of difficulty and diagnosticity on choice among tasks in relation to achievement motivation and perceived ability.* Unpublished manuscript, Universität Bielefeld, Bielefeld, Germany, 1978.
2. Trope, Y. *Uncertainty about one's own ability and achievement behavior.* Unpublished manuscript, Hebrew University of Jerusalem, 1978.

References

Atkinson, J. W. Motivational determinants of risk-taking behavior. *Psychological Review,* 1957, *64,* 359–372.

Attneave, F. *Applications of information theory to psychology.* New York: Holt, Rinehart & Winston, 1959.

Berlyne, D. E. *Conflict, arousal, and curiosity.* New York: McGraw-Hill, 1960.

Bradley, G. W. Self-serving biases in the attribution process: A reexamination of the fact or fiction question. *Journal of Personality and Social Psychology,* 1978, *36,* 56–71.

Festinger, L. A theory of social comparison processes. *Human Relations,* 1954, *2,* 117–140.

Jones, S. C. Self and interpersonal evaluations: Esteem theories versus consistency theories. *Psychological Bulletin,* 1973, *79,* 185–199.

Kukla, A. An attributional theory of choice. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 11). New York: Academic Press, 1978.

Lanzetta, J. T. Information acquisition in decision making. In O. J. Harvey (Ed.), *Motivation and social interaction.* New York: Ronald Press, 1963.

Lanzetta, J. T., & Driscoll, J. M. Effects of uncertainty and importance on information search in decision making. *Journal of Personality and Social Psychology,* 1968, *10,* 479–486.

Lewin, K., Dembo, T., Festinger, L., & Sears, P. Level of aspiration. In J. McV. Hund (Ed.), *Personality and the behavior disorders* (Vol. 1). New York: Ronald Press, 1944.

Meyer, W. U., Folkes, V., & Weiner, B. The perceived informational value and affective consequences of choice behavior and intermediate difficulty task selection. *Journal of Research in Personality,* 1976, *10,* 410–423.

Nisbett, R. E., & Valins, S. Perceiving the causes of one's own behavior. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior.* Morristown, N.J.: General Learning Press, 1972.

Schachter, S. The psychology of affiliation. Stanford, Calif.: Stanford University Press, 1959.

Schrauger, J. S. Responses to evaluation as a function of initial self-perceptions. *Psychological Bulletin,* 1975, *82,* 581–596.

Suls, J. M., & Miller, R. L. *Social comparison processes: Theoretical and empirical perspectives.* Washington, D.C.: Hemisphere, 1977.

Trope, Y. Seeking information about one's own ability as a determinant of choice among tasks. *Journal of Personality and Social Psychology,* 1975, *32,* 1004–1013.

Trope, Y., & Brickman, P. Difficulty and diagnosticity as determinants of choice among tasks. *Journal of Personality and Social Psychology,* 1975, *31,* 918–925.

Weiner, B., Frieze, I., Kukla, A., Reed, L., Rest, S., & Rosenbaum, R. M. Perceiving the causes of success and failure. In: E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior.* Morristown, N.J.: General Learning Press, 1972.

Zuckerman, M., Brown, R. H., Fischler, G. L., Fox, G. A., Lathin, D. R., & Minasian, A. J. Determinants of information seeking behavior. *Journal of Research in Personality,* in press.

Received October 15, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 9, 1519-1531

The Role of Facial Response in the Experience of Emotion

Roger Tourangeau and Phoebe C. Ellsworth
Yale University

Facial expression and emotional stimuli were varied orthogonally in a 3 × 4 factorial design in order to test whether facial expression is necessary or sufficient to influence emotional experience. Subjects watched a film eliciting fear, sadness, or no emotion, while holding their facial muscles in the position characteristic of fear or sadness, or in an effortful but nonemotional grimace; those in a fourth group received no facial instructions. The subjects believed that the study concerned subliminal perception and that the facial positions were necessary to prevent physiological recording artifacts. The films had powerful effects on reported emotions, the facial expressions none. Correlations between facial expression and reported emotion were zero. Sad and fearful subjects showed distinctive patterns of physiological arousal. Facial expression also tended to affect physiological responses in a manner consistent with an effort hypothesis.

Half a century ago, Cannon's decisive critique of the James–Lange theory ended scientific consideration of the hypothesis that peripheral responses provide the basis for qualitative distinctions among emotions. Nonspecific arousal theories have dominated the study of emotion ever since (Duffy, 1934, 1962; Lindsley, 1951). The James–Lange theory (James, 1890; Lange, 1885/1922) proposed that emotional stimuli elicit physiological responses specific to each emotion; the experience of an emotion, according to their view, is the perception of the corresponding physiological pattern.[1] By contrast, the nonspecific arousal theorists argue that physiological patterns do not correspond to specific emotions but only to the intensity of general emotional arousal and perhaps (Duffy, 1962) to a global, primitive approach–avoidance tendency. These theories have little to say about qualitative distinctions among the emotions. They tend to share, implicitly or explicitly, the assumption that such distinctions are the product of learning. The theory of Schachter and Singer (1962), for example, asserts that undifferentiated arousal is classified according to situational cues to determine the emotional experience. Qualitative distinctions derive from the classification; the classification presumably derives from social learning (see also Duffy, 1962).

The learning position of the arousal theorists is cast into some doubt by recent evidence for the widespread cross-cultural gen-

The research reported here was funded by National Science Foundation Grant BNS-75-22890 awarded to the second author. We are grateful for the assistance of Sidney Don and Steven Ganahl, who served as experimenters; Michael Brownstein, Barbara Fleming, Kathy Gould, and Jane Schweitzer, who assisted with scoring the data; Paul Ekman, Wallace V. Friesen, Alison Webb, Adelle Wenning, Nancy Wilner, and the Yale Audio-Visual Center, who helped us find, pretest, and edit the films; and Robert Novelly, who advised us on appropriate physiological measures.

Requests for reprints should be sent to Roger Tourangeau, Department of Psychology, Connecticut College, New London, Connecticut 06320.

[1] The theories proposed by James and by Lange differ in that Lange's theory was restricted to autonomic feedback (heart rate, stomach contractions, blushes, etc.), whereas that of James also included muscular feedback (such as changes in tonus, posture and, presumably, facial expression). Most subsequent writers, including the major critics, attributed the visceral version to both authors indiscriminately; in using the term "James–Lange theory" we will follow in this tradition, while recognizing that the muscular components of the James theory were not fully discredited by Cannon's research, and have much in common with the later facial feedback hypotheses.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3709-1519$00.75

erality of a small set of basic emotion categories, reliably used in labeling facial expressions of emotion (Ekman, Sorenson, & Friesen, 1969; Ekman & Friesen, 1971; Izard, 1971). If emotion categories are learned, cultural phenomena, why should all cultures studied so far share the same small set? In line with the evidence on the cross-cultural generality of the recognition of facial expressions, some theorists (Tomkins, 1962; Izard, 1971, 1977) have proposed that feedback from the facial muscles is important in the subjective experience of emotion. In the strongest version of these facial feedback theories, facial responses play the same critical role as more general visceral and muscular changes play in the James-Lange theory: the proprioception of the facial response *is* the experience of emotion. This shift in emphasis from visceral to facial feedback neutralizes most of Cannon's criticisms of the James–Lange theory. Cannon (1927), for example, argued that visceral responses were too slow and too undifferentiated to be the basis of the subjective experience; facial expressions are sufficiently immediate and sufficiently various; similarly, Cannon's demonstrations of "emotional" behavior in animals whose viscera were separated from their central nervous systems are irrelevant to the facial theories.

Although they differ on the causal priority assigned to each, these various positions predict that in general self-report, facial, and physiological measures of emotion should be positively correlated. In contrast, still another position predicts a negative correlation among these measures. This cathartic–hydraulic view was proposed first by James (1890); its chief exponent, however, is Freud (1946/1921). According to the hydraulic view, verbal, facial, and physiological responses are alternative channels for releasing the emotional energy evoked by a stimulus; if one channel is blocked, the response through the others should increase in intensity.

What is the evidence for the various views? The general arousal models receive a certain amount of indirect support from the numerous failures to find clear patterns corresponding to different emotions (Lacey, Kagan, Lacey, & Moss, 1963; Lindsley, 1951). In addition, Hohmann (1966) has found, contrary to Cannon's (1927) contention, that separation of the viscera from the central nervous system (due to spinal lesions) in humans is associated with reduced emotional responding. Finally, there is Schachter and Singer's (1962) demonstration that induced autonomic arousal can lead to various types of emotional (and unemotional) response (see also Schachter & Wheeler, 1962; Zillman & Bryant, 1974).[2] On the other hand, some of the better studies of emotional arousal *have* found evidence that different emotions are associated with different autonomic patterns (Ax, 1953; Funkenstein, 1955; Wolf & Wolff, 1947). Hohmann's findings on patients with spinal lesions invite numerous interpretations. Schachter and Singer's results do not always replicate (see Marshall & Zimbardo, 1979; Maslach, 1979), and even in their original experiment the differences between subjects given situational cues for euphoria and those given cues for anger were negligible. Finally, Cannon's arguments on the long latencies of visceral responses still pose difficulties for any theory that makes the sensation of autonomic arousal a necessary condition for emotional experience.

The strongest evidence for the facial feedback view comes from studies by Laird (1974) and Lanzetta, Cartwright-Smith, and Kleck (1976). Laird showed effects for manipulated facial expression on felt aggressiveness, and Lanzetta et al. showed similar effects on pain. The facial feedback theorists, however, cannot explain the results of Schachter and Singer's experiment (but see comments by Izard, 1977, chapter 2). In addition, the Laird study did not contain the control group necessary to determine whether the appropriate expression increases the response to a stimulus or the inappropriate expression inhibits the response (or both). Laird also used self-report measures in a within-

[2] Since the physiological measures in Schachter and Singer were extremely crude, we cannot be entirely confident that the physiological patterns were identical. It is quite possible that the situational cues modified the physiological response.

subjects design, leaving open the possibility that demand characteristics were responsible for the results. Lanzetta et al. (1976) also used a within-subjects design, but their inclusion of a galvanic skin response (GSR) measure makes an account in terms of demand characteristics less plausible. However, since pain is not typically included in theories of emotion, their findings may not generalize to the feeling states that are.

The best evidence for the hydraulic view comes from studies showing a negative correlation between facial expressiveness and measures of physiological arousal (Buck, Miller, & Caul, 1974; Buck, Savin, Miller, & Caul, 1972; Lanzetta & Kleck, 1970). This evidence tends to disconfirm the other theories. However, few of the studies cited in support for this view actually show the negative correlation expected to hold *within* a given individual. Instead, they find evidence that the most expressive people are not the most physiologically aroused; they do not find that the same individual is more expressive when he is less aroused. In addition, several studies (including Lanzetta et al., 1976) have found significant *positive* correlations.

Thus none of the evidence is decisive, and most of the questions about the roles and relative importance of the various components of emotional responding remain open. This study attempts to answer some of these questions by testing several predictions from the facial feedback hypothesis. What we are calling the "facial feedback hypothesis" obviously derives from the theories of Tomkins (1962), who was the first to claim that the emotions are primarily facial behaviors, and later Izard, who maintains that "awareness of facial activity or facial feedback is actually our awareness of the subjective experience of a specific emotion" (1977, p. 60). Nonetheless the general hypothesis we are testing here is not the same as that of either of these theorists. In the first place, both theories are comprehensive statements containing numerous propositions about the relationship of emotions to personality, motivation, communication, and each other. The facial feedback hypothesis involves just one of these propositions, albeit a central one.

In the second place, neither author is entirely static in defining the implications of his general statements about the importance of the face. Our experiment was designed to test (a) whether the appropriate facial expression is necessary for the subjective experience of the emotion, and (b) whether the voluntary assumption of an expression is sufficient to produce the experience. If facial expression is *necessary* for emotional experience, there should be no emotion unless the face responds. Even in the presence of emotional stimuli, without the appropriate emotional expression, no emotion should be felt. If the facial expression is *sufficient* for emotional experience, when the face responds, the emotion should follow. Even in the absence of emotional stimuli, an emotional facial expression should produce an emotional feeling. Weaker forms of the hypotheses predict that feedback from the face should have a significant main effect on the emotional experience, attenuating or intensifying it. These hypotheses are derived from some of the more strongly worded statements of Tomkins (1962) and Izard (1977), statements that are qualified in other parts of their work. Both theorists argue that adults may have learned to duplicate the effects of proprioceptive feedback by means of a reafferent loop from the subcortical centers directly to the cortex, rather than from the subcortical centers to the face to the cortex, so that actual movement of the face is not always *necessary*. Similarly, they have argued that voluntary movement of the facial muscles may not be *sufficient* to produce the corresponding emotion, because it does not create exactly the same proprioceptive cues as the involuntary movement created by a "real" emotional stimulus.

Although the theories allow for the possibility that neither the necessity nor the sufficiency hypothesis is true, we believe that these hypotheses are worth testing. In the first place, the practical, therapeutic implications of the theory are much greater if the facial muscles are actually involved in the experience. If the face were necessary for emotional experience, victims of accidents, disease, or surgery resulting in facial paralysis or sensory impairment would be expected to show corre-

sponding affective deficiencies, and additional therapeutic attention would be indicated. If the face were sufficient to influence the felt emotion, it would be useful to teach patients suffering from affective disorders how to control their facial expressions. In the second place, the qualifications render the theory much less testable. If the only influential facial expression is one that results from an involuntary natural response and if the facial muscles can be bypassed intracranially, the causal role of the face becomes inaccessible to any sort of definitive empirical test.

A final hypothesis tested by our research is one that *does* follow directly even from the weaker statements of Tomkins (1962) and Izard (1977), as well as from the stronger facial feedback hypothesis. It is simply the prediction that, in general, the relationship between the facial expression and the emotional experience should be monotonic and positive.

The three hypotheses of necessity, sufficiency, and monotonicity were tested in this study by a design in which facial expression was manipulated independently of emotional stimulation. The basic design included a replication across two emotions—fear and sadness; we observed facial, physiological, and self-report responses. We chose fear and sadness because we felt that it was important to demonstrate a distinction between negative emotions. The comparison of a single positive with a single negative emotion does not provide a very stringent test of the qualitative distinctions among emotions, and is particularly prone to demand characteristics and level-of-arousal artifacts.[3] Because it included measures of all three sets of responses, the study can also address additional questions about their interrelationships and thus can provide a basis for comparing all the theoretical positions.

Method

Overview

Believing that the experiment was a study of physiological responses to subliminal stimuli, subjects watched a sad, fear-arousing, or emotionally neutral film. Their heart rate, galvanic skin response, and respiration rate were recorded, and the placement of additional fake electrodes on their faces provided a rationale for asking them to hold their facial muscles in a constant position during the film. These positions corresponded to a fearful expression, a sad expression, or a grimace unrelated to any emotion. A final group of subjects watched one of the three films but received no facial instructions. During the film the subjects' faces were videotaped, and as soon as it was over the subjects rated their emotional experience. The design was thus a 3×4 factorial with three films (fear, sad, and neutral) and four facial expressions (fear, sad, neutral, no instructions).

Subjects

The subjects were 128 undergraduates, 85 of whom received both credit (in partial fulfillment of introductory course requirements) and $2.00. The remaining 43 subjects received $3.00. Five subjects were not included in the analyses: 2 left the experiment when the content of the film was described to them, 1 stopped the film in the middle, and 2 others were lost because of equipment problems.

Procedure

There were two experimenters in the study. The first was blind to the subject's facial instruction condition. This experimenter (E_1) told the subject the basic cover story: the experiment concerned physiological indices of subliminal perception—heart rate, skin conductance, respiration rate, "the orienting reflex," and "subvocal speech"; all of these responses would be recorded on a polygraph, and the two that involved small movements of the eyes and lips would also be recorded on videotape. The subliminal stimuli would be single frames spliced into a film. E_1 then explained that certain parts of the procedure might cause some discomfort: (a) to prevent the subject from concentrating too hard on finding the subliminal images, the film was intended to be distracting, and might be upsetting; (b) since normal muscle movements could distort some of the physiological measures, the subject might be required to hold certain muscles in a somewhat uncomfortable position during the film; (c) since observers would later score the videotapes for the orienting reflex and subvocal speech response, absolutely complete anonymity could not be guaranteed. After explaining how the subject would be "hooked up" to the polygraph, E_1 described the content of the film and obtained the subject's formal consent.

---

[3] We considered a comparison between fear and anger, in an attempt to include a replication of Ax (1955). We had to abandon this plan for the interesting reason that we could find no film that consistently elicited anger more than other emotions in all subjects.

E_1 then left and sent the second experimenter (E_2) into the room. E_2 had been out of earshot and so was unaware of the subject's film condition.

E_2 placed the electrodes on the subject (electrocardiograph [EKG] electrodes on the wrists, a respiration thermistor[4] on the nostril, GSR electrodes on the middle finger, bogus electromyograph [EMG] electrodes on the face), explaining how they worked and reinforcing the rationale of the cover story. After the electrodes had been placed, E_2 told the subject to keep his or her arms still, and gave the instructions for holding the face in the position that would facilitate recording the facial responses. When E_2 was satisfied with the subject's facial pose, he left the subject alone for a baseline period with instructions to relax.

The baseline period lasted until the subject's physiological responses had appeared stable for at least 1 minute. Then E_2 began videotaping the subject's face and repeated the facial instructions (in a shorter form). When the facial expression was approximately right, E_2 told the subject to hold the position during the film; he then turned the projector on and immediately left the room. As soon as the film was over (E_1 watched through a one-way mirror for the end of the film), E_2 returned and administered a questionnaire containing the self-report emotion items, along with filler items consistent with the cover story.

E_1 debriefed the subjects. Although several subjects expressed confusion about the complex cover story, no one guessed that facial expression was the variable of interest, nor that his or her expression had been an emotional one.

Two male undergraduates took the E_1 role; one male graduate student took the E_2 role.

Independent Variables

Film. We pretested eight films on a group of 50 undergraduates, and selected three because they elicited high agreement across subjects on a single dominant emotion, with low ratings for all other emotions. Subjects rated each of five emotions on two 9-point scales (0–8) for each film. The ratings are the sums of the two scale scores. The *fear* film reliably elicited fear ($M = 5.5$); lesser degrees of pain, disgust, and interest; and relatively low sadness ($M = 2.4$). It concerns two accidents in an industrial shop. The *sad* film traces the reactions of a small boy to his brief stay in an orphanage while his mother is in the hospital. The sad film reliably elicited sadness ($M = 3.6$) and interest ($M = 3.4$) from pretest subjects, and low fear ($M = 0.75$). The neutral film depicts a flower show in the botanical gardens of Golden Gate Park. It was seen as slightly pleasant ($M = 2.8$) and interesting ($M = 2.1$) and not at all sad ($M = 0.20$) or frightening ($M = 0.02$). Each film lasted for 2 minutes.

Facial instructions. The facial instructions were derived from the work of Tomkins (1962), Izard (1971), Ekman and Friesen (1975), and the instructions used by the second author in developing prototype photographs for the Facial Affect Scoring Technique (Ekman, Friesen, & Tomkins, 1971). In practice, the instructions varied somewhat depending on how easily the subject adopted the desired position. For the subjects in the *fear face* condition, the instructions typically ran:

> There are three sets of muscles around the eyes that can distort the measurement of the orienting reflex. We want you to contract all three of them. The first is the muscle between the eyebrows, the corrugator. Contract that muscle by pulling the two eyebrows together, toward each other in the middle. The second is the muscle in the forehead, the frontalis. Contract the frontalis by raising your eyebrows. The last muscle is the one which controls the eyelids. Contract that one by opening your eyes up wide. There is only one set of muscles you'll have to contract in your mouth. First, part your lips slightly; it will be easier. The muscles here [points below corners of mouth] are the triangularis muscles. Contract them by pulling the corners of your mouth down and back. If you're doing it right, you should feel your neck get tense.

For the *sad face* subjects, the instructions were similar, although the muscles differed: the corrugator was contracted (brows drawn together), frontalis contracted (*inner corner* of the brows raised), eyelids relaxed, mentalis contracted (lower lip pushed up and out slightly), and quadratus muscles contracted (corners of lips pulled down). The *nonemotional face* subjects were instructed to close one eye, purse their lips, and puff out their cheeks. *Unmanipulated face* subjects were told to ignore the facial electrodes and to act "naturally," as we were "interested in determining whether the orienting reflex and subvocal speech responses can be detected against a background of normal facial movements." For all subjects, electrodes were placed on the chin and below one eye (for subjects in the nonemotional face condition, this bogus electrode was placed below the eye they were supposed to close).

Dependent Variables

Self-reported emotion. Fear was measured on two 9-point scales labeled "scared" and "afraid"; the scales ranged from "not at all" (0) to "very strongly" (8). These two scales were highly related ($r = .87$). Sadness was also measured on two 9-point scales; these were labeled "sad" and "unhappy" ($r = .71$). As in the pretests, the overall emotion score is the sum of the two scales.

Facial expressions. Two trained raters blind to the subject's conditions scored videotapes on the emo-

---

[4] Respiration rate was recorded as a control for certain heart rate artifacts. These corrections proved unnecessary. The results for respiration rate were not analyzed.

tional content of the subject's facial expressions. For each subject, the raters judged how sad, unhappy, scared, and afraid the subject looked. (The scales for these judgments were identical to those used by the subjects in judging their own emotions; they were, thus, 9-point scales ranging from "not at all" to "very strongly.") The facial sadness measure is the sum of the two raters' judgments for "sad" and "unhappy"; similarly, the facial fear rating is the sum of the two ratings of "scared" and "afraid."

Physiological indices. Physiological responses were monitored by a Narco-Bio Desk Model Physiograph. EKG electrodes were attached to the subjects' wrists with a ground electrode on the ankle. A cardiotachometer averaged the beat-to-beat interval over five beats and recorded the heart rate in beats per minute. The skin resistance was monitored by passing a small direct current through two plate electrodes attached to the top and bottom of the subject's middle finger.

Artificiality. Our efforts to keep the experimenters blind to the subject's condition, to prevent the subjects from guessing the hypotheses, and to collect three different kinds of data, although successful, resulted in a situation that was complicated and unusual and may recall the futile efforts of early researchers to obtain photographs of "true emotions" in the laboratory (e.g., Landis, 1924). Although emotional experience may have been attenuated in our setting, there is no evidence that this was the case. First, the self-reported emotional reactions to the films were at least as high as those obtained in pretesting; second, several subjects spontaneously commented on their emotional arousal, and one stopped the projector because the film was too upsetting; third, there were no false alarms in the reporting of subliminal stimuli, as might be expected if the task demands distracted subjects from the arousing properties of the films; and finally, unlike the early research, our emotional stimuli produced significant differences on the dependent variables.

Results [5]

Facial Expression

The judgments of facial expression appear to be reliable across raters: for the two-item sadness index, the two raters' judgments correlated .76 ($p < .001$); for the fear index, the two raters correlated at .81. For the remaining analyses, ratings were summed over the two judges.

The facial instructions had the expected large effect on the ratings of facial expression. On the average, subjects given the sad facial instructions were rated as more than 15 points sadder (on a 32-point scale) than the subjects given any of the other three facial instructions, $F(3, 3) = 153.1$, $p < .01$. Subjects given the fear face instructions averaged 13 points higher on rated facial fear, $F(3, 3) = 327.1$, $p < .01$. The facial instructions thus seemed to have succeeded.

There is also some indication that the film affected ratings of facial expression. Subjects watching the fear film looked more fearful ($M = 8.2$) than those watching the sad (4.4) or neutral (5.4) films, $F(2, 2) = 116.7$, $p < .01$. There was no corresponding effect for rated facial sadness. This effect provides some evidence for the validity of the judges' ratings (since the film might have been expected to affect facial expression in the obtained direction) and of the success of the films in creating real emotions but also indicates that the facial instructions manipulation was not wholly successful. Apparently subjects could not consistently maintain their instructed expression when confronted with the strong stimuli of the film. Though reliable, the effect of the film is small, particularly when compared with the effect for facial instructions.

There were no interactions nor main effects for experimenter on rated facial expression.

---

[5] The cell sizes in this study are unequal. While the differences are small, they may in part reflect the experimental variables of interest; the film, for example, may have affected subject attrition slightly (two fear film subjects, one sad film subject, and no neutral film subjects refused to participate after learning about the film they would see). Under these circumstances, weighted means analysis seemed most advisable (Winer, 1962, p. 222). Partly because the inequalities are so small, unweighted analyses yield identical conclusions.

The film and face variables were treated as fixed factors and experimenter as a random factor. Thus main effects for the film and facial instructions were tested against their interaction with the experimenter factor. Error terms based on pooling these interaction terms with the within-group sum of squares (Winer, 1962, pp. 202–207) do not alter the conclusions except where reported. The pooling procedure is used whenever permissible for planned and a posteriori comparisons, because of the handicap of 1 and 1 degrees of freedom. Small variations in the degrees of freedom in these analyses reflect missing data and the specific terms that could be included in the pooled error term.

Self-Reported Emotion

As Tables 1 and 2 show, the film had substantial effects on the subjects' ratings of their own emotions: for fear, $F(2, 2) = 29.5$, $p < .05$; for sadness, $F(2, 2) = 42.1$, $p < .05$. A priori contrasts indicated that subjects who watched the fear film were more frightened than subjects who watched the other films, $F(1, 106) = 17.9$, $p < .01$; subjects who saw the sad film were sadder, $F(1, 103) = 77.1$, $p < .01$. Each contrast accounts for more than 90% of the variance among the means for the three film conditions. (See Footnote 5 for an explanation of the degrees of freedom).

An examination of the difference between fear (or sadness) and the mean of all other emotion ratings confirms this analysis. Compared to subjects who watched the sad and neutral films, fear film subjects felt *predominantly* fear. For subjects who watched the fear film, the self-rated fear was 2.65 points higher than the average self-rating of all the other emotions; for sad film subjects self-rated fear was 1.06 points *lower* than the average across the other emotions, and for neutral film subjects it was about the same—0.28 points lower, $F(2, 2) = 113.8$, $p < .01$. The overall analysis also shows an uninterpretable Face × E interaction. No other effects are significant. Compared to subjects who watched the fear and neutral films, sad film subjects felt *predominantly* sadness. Their self-rated sadness was 5.20 points higher than the average of all the other emotions; for fear film subjects self-rated sadness was about the same as the average of all other emotions (.06 points lower) and for neutral film subjects 1.42 points lower, $F(2, 2) = 20.5$, $p < .05$. No other effects were significant.

Table 1
Mean Self-Reported Fear

Film	Facial instructions: Fear	Sad	Nonemotional	Unmanipulated	*M*
Fear	6.0	7.5	7.8	4.6	6.5
Sad	3.3	3.2	3.9	3.9	3.6
Neutral	3.3	2.1	2.3	2.2	2.5

Note. Scale ranges from 0 to 16, with higher numbers indicating greater fear.

Table 2
Mean Self-Reported Sadness

Film	Facial instructions: Fear	Sad	Nonemotional	Unmanipulated	*M*
Fear	4.1	4.4	4.7	3.5	4.4
Sad	8.9	8.8	9.7	8.8	9.1
Neutral	0.9	2.6	1.9	1.4	1.6

Note. Scale ranges from 0 to 16, with higher numbers indicating greater sadness.

Both the necessity hypothesis (that facial responses are necessary for the felt emotion) and the sufficiency hypothesis (that facial responses are sufficient for felt emotion) of the facial feedback theory predict effects for facial instructions. There are, however, no significant main effects for facial expression instructions on either emotion, nor does facial instruction interact significantly with film.

One possible interpretation of the lack of effects for facial instructions is that the subjects were simply unable to maintain their facial expressions during the film. Part of the film effect, then, may reflect covariation of the facial expression with the film condition. Partialing out variation in felt emotion due to facial expression through analysis of covariance, however, fails to alter significantly the very substantial film effects. The effects of the film on self-reported emotion are thus not attributable to differences in facial expression across film conditions.

A weakened version of the sufficiency hypothesis might predict that, in the absence of strong situational cues, the facial expression may be sufficient to determine felt emotion. The neutral film subjects, by this argument, should show an effect for facial expression. This prediction receives very slight support: among those subjects who watched the neutral film, the sad face subjects are, on the average, slightly sadder than subjects in the other three facial instructions conditions; similarly, of the neutral film subjects, the fear face subjects report the most fear. Neither of these

Table 3
Means on Physiological Indices by Film and Facial Instructions Condition

Condition	Fall in HR	Rise in HR	Fall in GSR	No. GSRs
		Film		
Fear	7.4	21.3	71.0	14.4
Sad	7.7	16.4	67.1	11.0
Neutral	10.8	14.6	52.2	11.7
		Facial instructions		
Fear	8.4	17.0	43.8	12.6
Sad	8.8	17.5	60.2	10.2
Nonemotional	7.1	19.6	63.8	15.8
Unmanipulated	9.9	15.7	86.2	11.0

Note. HR = heart rate (change computed in beats per minute). GSR = galvanic skin response (change computed in thousands of ohms).

trends for the neutral film subjects, however, is statistically significant, either by a priori comparisons within the context of the analysis of variance or by Kruskal–Wallis one-way analyses of variance for ranked data.

Overall, the relationship between facial expression and reported emotion is slight. The correlation between the ratings of facial fear and self-reported fear is .01; between facial and self-reported sadness, it is .02. The strongly positive relationship between facial expression and self-reported emotion predicted by all the facial feedback theories does not seem to obtain.

Physiological Measures

Two coders scored the polygraph recordings of the physiological variables. Their agreement was substantial, correlations between them ranging from .96 to 1.00 (median r = .98). Indices were based on six variables: baseline heartrate, the average of the subject's heart rate 10 and 5 seconds prior to the end of the baseline period (each heart rate reading is itself an average based on five beat-to-beat intervals); baseline skin resistance, also an average of readings 10 and 5 seconds prior to the end of the baseline period; maximum heart rate during the film period; minimum heart rate during the film; number of skin responses (any fall exceeding 1,000 ohms prior to leveling off was scored as a response); and lowest skin resistance during the film. Both skin resistance measures are reported in thousands of ohms. The heart rate (HR) measures are in beats per minute. The relevant means for the physiological variables are given in Table 3.

Rise in heart rate. The largest rise in HR for each subject was calculated as the difference between maximum and baseline HR. The film had a significant effect on this index, $F(2, 2) = 36.4$, $p < .05$. A posteriori contrasts (Scheffé criterion, Winer, 1962, p. 88) indicate that subjects who watched the fear film showed the largest rises in heart rate; neutral and sad film subjects showed smaller rises and were similar on this index, $F(1, 105) = 11.00$, $p < .05$; the contrast accounts for 96% of the variation among the means of the film groups. None of the other main effects or interactions are statistically significant.

Fall in heart rate. By subtracting the lowest heart rate from the index of baseline heart rate, we can find the largest fall in heart rate. There is a nearly significant effect for the film on this variable, $F(2, 2) = 17.1$, $.10 > p > .05$. An a posteriori contrast is similarly marginal: neutral film subjects show the largest drop in heart rate; fear and sad subjects show similarly smaller drops, $F(1, 105) = 4.78$, $.10 > p > .05$; the contrast accounts for 99% of the variation. There is also an effect for the facial instructions variable, $F(3, 3) = 37.2$, $p < .05$. Subjects in the nonemotional face condition show the smallest drop in heart rate, subjects in the unmanipu-

lated face condition the largest. The interaction term used to test this effect for facial instruction is quite small ($F < 1$); with a pooled error term, the effect is no longer significant: $F(3, 105) < 1$, *ns*. None of the other effects is significant.

Fall in GSR. Subtracting the lowest skin resistance from the baseline index yields a measure of overall change in skin resistance. Facial instruction condition has an effect on this variable: Subjects in the unmanipulated facial condition showed the largest drop in skin resistance, the fear instructions subjects the smallest. As with the facial instructions effect on fall in heart rate, the statistical significance of this is probably overestimated by the use of the Face × E interaction term as the error term: $F(3, 3) = 44.4$, $p < .01$; a pooled error yields $F(3, 97) = 1.7$, *ns*. The other effects do not reach statistical significance.

Number of galvanic skin responses. Each time skin resistance fell by 1,000 ohms, it was scored as a response. The film variable had a significant effect of the number of GSRs: $F(2, 2) = 57.1$, $p < .01$. An a posteriori contrast is marginally significant: Subjects watching the fear film had more GSRs on the average than the sad and neutral film subjects, $F(1, 98) = 5.63$, $.10 > p > .05$; the contrast accounts for 95% of the variation.

Summary of physiological effects. Three of the four physiological variables showed significant film effects. For heart rate fall, sad and fear film subjects were similar, both being lower than neutral film subjects. For rise in heart rate and number of skin responses, the sad film subjects were closer to the neutral film subjects. There was no reliable effect for film on largest fall in skin resistance, although the pattern was similar to that for heart rate (sad and fear subjects were similar and both were different from neutral film subjects). This pattern of different physiological signs for different emotional stimuli also appears in Table 4. The pattern of correlations between physiological variables and self-reported emotion differs for sadness and fear.

Both fall in heart rate and fall in GSR show some evidence for facial instructions effects, although the statistical significance of the effects depends on the choice of the error term. For fall in HR, the nonemotional facial instructions subjects seem the most "aroused" (showing the smallest drop). This is in line with the hydraulic–cathartic view. It is the unmanipulated subjects, however, rather than the fear or sad face subjects, who show the least "arousal" (largest drop). Similarly, the fear face subjects show the smallest drop in skin resistance, again in line with the hydraulic–cathartic view; however, this time the unmanipulated rather than the nonemotional subjects show most arousal. Table 4 offers further evidence of the traditionally weak and somewhat inconsistent relationship between facial expression and physiological response.

Table 4
Correlations of Physiological Variables with Facial and Self-Report Measures

Item	Self-report		Facial expression	
	Fear	Sad	Fear	Sad
Fall in HR	−.05	−.10	.02	−.01
Rise in HR	.16*	−.05	.04	−.02
Fall in GSR	−.05	.03	.08	−.08
No. of GSRs	−.05	−.18*	−.02	−.14
Maximum HR	.19*	−.08	03	.02
Minimum HR	.16*	.00	.01	−.01
Lowest GSR	.06	.19*	−.07	−.06

Note. HR = heart rate (change measured in beats per minute). GSR = galvanic skin response (change measured in thousands of ohms).
* $p < .05$.

Discussion

The Facial Feedback Hypothesis

On the assumption that the facial instructions had their intended effect, the facial feedback hypothesis receives three major setbacks from the evidence of this study. First, adopting an emotional facial expression does not appear to be *sufficient* to produce the emotion. Even when there were no competing emotional stimuli from the film (i.e., for the neutral film subjects), manipulated facial expression did not produce significant differences in emotional responding. The trend for these subjects, insofar as there was a trend, was in the

direction predicted by the facial feedback hypothesis. This trend appears to replicate Laird's (1974) results: the differences he found were of about the same magnitude as those in our neutral film condition, and it is probably a safe assumption that his stimuli (still photographs) were less arousing than our fear and sadness films.

Second, adopting a nonemotional expression does not prevent emotional responding; thus, emotional expression does not seem *necessary* for emotional feelings. Similarly, correcting for changes in facial expression statistically by analysis of covariance does not remove the effect of the film. Although this technique may be biased in the direction of undercorrection, it is hard to see how any statistical method based on the correlation between facial expression and reported emotion could alter the film effect—the correlation is zero.

Finally, this lack of correlation constitutes especially damaging evidence against the theory. Examination of the scatter plots of facial and self-reported fear and of facial and self-reported sadness does not provide obvious support for any monotonic relationship between self-report and facial expression of emotion, let alone the linear relation measured by the correlation coefficient. Thus, even a threshold version of the facial feedback hypothesis seems untenable. Lanzetta, Cartwright-Smith, and Kleck's (1976) finding of an effect of facial expression on feelings of pain does not seem to extend to feelings of fear or sadness.

Even if the facial manipulation were *unsuccessful,* the facial feedback hypothesis would be difficult to maintain. The absence of any correlation between rated expression and reported emotion might buttress an argument based on failure of the facial manipulation. The convergence of the facial instructions with raters' judgments could be explained away: raters might have recognized the intended expression even though subjects' faces were poor reflections of the canonical fear or sad expression. Granting both of these arguments, we still must explain the absence of inhibiting effects for facial expression; even if the facial manipulation of fear, for example, were woefully inadequate to produce the canonical fear expression, it is difficult to believe that it didn't greatly interfere with the emergence of the sad expression, thus reducing felt sadness. Finally, the existence of the film effect, its independence of the facial expression, and its magnitude compared with that of any facial effect present difficulties for the facial feedback position. As to these relative magnitudes, our results are in complete agreement with those presented by Laird (1974).

The lack of any correlation between facial expression and reported emotion is damaging not only to the rather strong and unqualified version of the facial feedback hypothesis tested in this experiment but also to the more elaborated, qualified theories proposed by Tomkins (1962) and by Izard (1971, 1977). Even if there are reafferent loops and even if the proprioceptive feedback along voluntary and involuntary pathways is recognizably different, the theories ought to predict a generally positive correlation. The unmanipulated face condition is especially relevant here, since there were no instructions to introduce potentially confusing voluntary feedback. In these conditions, all facial expression was spontaneous, and the correlations between expression and reported emotion were still infinitesimal ($r = -.01$ for fear; $r = .07$ for sadness).[6]

Physiological Results

It is possible that self-reported emotion may reflect the subject's perception of the expected effect of the film. Such demand characteristics are also relevant to Laird's (1974) study and may account for the effects of the stimuli on self-reported emotion in both his study and ours. The effects of the film on the physiological variables cannot be so easily accounted for by demand characteristics.

---

[6] In general the subjects in the unmanipulated face condition showed little overt facial response. It is possible that covert facial expressions, unobserved by our raters, did correlate with self-report of emotion (cf. Schwartz, Fair, Salt, Mandel, & Klerman, 1976). Thus it is still possible that covert involuntary muscle activity has some causal influence, although the difficulties of separating this influence from the effect of the eliciting emotional stimuli are enormous.

Table 5
Mean Physiological Arousal for Nonemotional and Unmanipulated Face Subjects in Neutral Film Condition

Item	Nonemotional	Unmanipulated	t (18)	p
Number of GSRs	16.4	10.0	1.94	.07
Fall in GSR	68.6	58.5	1.53	*ns*
Rise in HR	18.0	13.4	1.53	*ns*
Fall in HR	6.8	12.6	−1.74	.10

Note. $n = 9$ in the nonemotional face condition; $n = 11$ in the unmanipulated face condition. GSR = galvanic skin response (change measured in thousands of ohms). HR = heart rate (change measured in beats per minute).

Subjects who watched the fear film showed generally greater "arousal" than subjects who watched the sad or neutral film. The pattern of rises and falls for subjects who watched the fear film is quite similar to the pattern reported by Ax (1953) for fearful subjects. For the heart rate variables, sad film subjects were intermediate between fear and neutral film subjects. However, sad film subjects showed even fewer GSRs than subjects who saw the neutral film. These data support the notion of different physiological patterns for different emotions. This position receives further support from correlations between physiological variables and self-reported emotion; again, sadness seems related to lower levels of arousal, fear to higher levels. These findings tend to render the demand-characteristics account relatively less plausible. Fear film subjects seem to show the physiological pattern for fear. These findings also tend to render suspect the theoretical utility of the nonspecific arousal concept. (Of course, the mere existence of physiologically distinct patterns does not guarantee that people use them, as the James–Lange theory asserts, as cues to their emotional state.)

Facial and Physiological Variables

The Freudian hydraulic model suggests that there are different channels for emotional expression; as one channel is used more, the others are used less in releasing emotional energy. We might expect, according to this theory, negative relations between measures of physiological arousal and facial expression. Manipulated facial expression did have some effect on physiological responding, although the statistical significance of the findings is dubious. For the two heart rate variables and the number of GSRs, nonemotional face subjects did, in line with the hydraulic model, show more physiological "arousal" than subjects who received the other facial instructions. The results for unmanipulated subjects, whose facial emotion was more than that of the nonemotional face subjects but less than that of subjects posed with a fearful or sad expression, create difficulties for the hydraulic view. In general, the unmanipulated face subjects showed the least arousal.[7]

On the whole, these results are in line with an effort or concentration hypothesis. The nonemotional facial position, which required subjects to close one eye and puff out their cheeks, was probably the most difficult position to maintain and required a great deal of concentration. The unmanipulated face, of course, required no special effort at all. That concentration on a task can produce physiological changes has been amply demonstrated (e.g., Lacey, 1950; Lacey, Kagan, Lacey, & Moss, 1963). This mechanism can also account for the individual differences reported by previous investigators (Lanzetta & Kleck, 1970; Buck, Savin, Miller, & Caul, 1972; Buck, Miller, & Caul, 1974): "natural" in-

[7] The exception is fall in skin resistance. Unmanipulated subjects had the highest skin resistance during the baseline period. They showed the largest drop during the film; despite this, their lowest skin resistance during the film still averaged higher than that of subjects in the other groups. Some of their large drop doubtless reflects regression artifacts.

hibitors concentrate on inhibiting their facial expressions; this concentration produces changes in GSR. The effects of facial instructions on physiological arousal are not entirely in line with the hydraulic model and are perhaps better explained by an effort or concentration mechanism.

This hypothesis is, however, both tentative and post hoc. Although some of the facial effects are significant, others are not. Comparing nonemotional and unmanipulated face subjects across all film conditions a posteriori, the only physiological measure that is significantly higher in the nonemotional facial conditions is the number of GSRs, $F(1, 98) = 9.97$, $p < .05$. It might be argued that the purest test of the effort hypotheses is in the neutral film condition, where effort is the major source of physiological arousal. Table 5 shows the means for the four physiological variables for the nonemotional and unmanipulated face subjects who watched the neutral film. Testing all four physiological variables together, these two groups did not differ significantly, Hotelling's $T^2(4, 15) = 9.9$, $p < .25$. Although all the means differ in the direction consistent with an effort hypothesis, they do not reach conventional levels of significance.

Summary

In an area where counterintuitive theories and puzzling results seem the rule, our results seem to support a common sense theory. Emotional stimuli such as our films affect subjective experience, facial expressions, and physiological processes associated with an emotion. The effect of the stimuli does not, as the facial feedback hypothesis predicts, depend on the facial response. Nor, as the nonspecific arousal theorists claim, are the physiological effects the same for all emotions (although people may not pay any attention to the differences). Finally, covering up an emotion facially may increase physiological responding, but this increase does not appear to result from the emotion's having to "come out somewhere else" (as the hydraulic view would have). Instead, it seems plausible that the concentration required to repress the outward expression of an emotion has an effect on physiological response. Besides supporting the common sense view, these results also support a general self-perception hypothesis. A variety of cues—facial, physiological, situational—may enter into the subjective experience of an emotion. Our results suggest that the situational cues receive the most weight.

References

Ax, A. The physiological differentiation between fear and anger in humans. *Psychosomatic Medicine*, 1953, *15*, 433–442.

Buck, R., Miller, R., & Caul, W. Sex, personality, and physiological variables in the communication of affect via facial expression. *Journal of Personality and Social Psychology*, 1974, *30*, 587–596.

Buck, R., Savin, V., Miller, R., & Caul, W. Communication of affect through facial expression in humans. *Journal of Personality and Social Psychology*, 1972, *22*, 367–371.

Cannon, W. The James–Lange theory of emotions: A critical examination and an alternative theory. *American Journal of Psychology*, 1927, *39*, 106–124.

Duffy, E. Emotion: An example of the need for reorientation in psychology. *Psychological Review*, 1934, *41*, 184–198.

Duffy, E. *Activation and behavior*. New York: Wiley, 1962.

Ekman, P., & Friesen, W. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, 1971, *17*, 124–129.

Ekman, P., & Friesen, W. *Unmasking the face*. Englewood Cliffs, N.J.: Prentice-Hall, 1975.

Ekman, P., Friesen, W., & Tomkins, S. Facial affect scoring technique: A first validity study. *Semiotica*, 1971, *1*, 37–53.

Ekman, P., Sorenson, E. R., & Friesen, W. Pan-cultural elements in facial displays of emotion. *Science*, 1969, *164*, 86–88.

Freud, S. Instincts and their vicissitudes. In *Collected Papers* (Vol. 4). London: Hogarth Press, 1946. (Originally published, 1921.)

Funkenstein, D. The physiology of fear and anger. *Scientific American*, 1955, *192*, 74–80.

Hohmann, G. Some effects of spinal cord lesions on experimental emotional feelings. *Psychophysiology*, 1966, *3*, 143–156.

Izard, C. *The face of emotion*. New York: Appleton-Century-Crofts, 1971.

Izard, C. *Human emotions*. New York: Plenum, 1977.

James, W. *The principles of psychology*. New York: Holt, 1890.

Lacey, J. Individual differences in somatic response patterns. *Journal of Comparative Physiological Psychology*, 1950, *43*, 338–350.

Lacey, J., Kagan, J., Lacey, B., & Moss, H. The visceral level: Situational determinants and be-

havioral correlates of autonomic response patterns. In P. Knapp (Ed.), *Expression of the emotions in man*, pp. 161–199. New York: International Universities Press, 1963.

Laird, J. Self-attribution of emotion: The effects of expressive behavior on the quality of emotional experience. *Journal of Personality and Social Psychology*, 1974, *24*, 475–486.

Landis, C. Studies of emotional reactions: II. General behavior and facial expression. *Journal of Comparative Psychology*, 1924, *4*, 447–509.

Lange, C. *The emotions*. In Istar Haupt (Trans.), (Originally published in Denmark, 1885.) K. Dunlap (Ed.), *The emotions*. Baltimore, Md.: Williams & Wilkins, 1922.

Lanzetta, J., & Kleck, R. Encoding and decoding of facial affect in humans. *Journal of Personality and Social Psychology*, 1970, *16*, 12–19.

Lanzetta, J., Cartwright-Smith, J., & Kleck, R. Effects of nonverbal dissimulation on emotional experience and autonomic arousal. *Journal of Personality and Social Psychology*, 1976, *33*, 354–370.

Lindsley, D. Emotion. In S. S. Stevens (Ed.), *Handbook of Experimental Psychology*. New York: Wiley, 1951.

Marshall, G., & Zimbardo, P. The affective consequences of "inadequately explained" physiological arousal. *Journal of Personality and Social Psychology*, 1979, *37*, 970–988.

Maslach, C. Negative emotional biasing of unexplained arousal. *Journal of Personality and Social Psychology*, 1979, *37*, 953–969.

Schachter, S., and Singer, J. E. Cognitive, social and physiological determinants of emotional state. *Psychological Review*, 1962, *69*, 379–399.

Schachter, S., & Wheeler, L. Epinephrine, chlorpromazine, and amusement. *Journal of Abnormal Psychology*, 1962, *25*, 121–128.

Schwartz, G. E., Fair, P. L., Salt, P. S., Mandel, M. R., & Klerman, J. L. Facial muscle patterning to affective imagery in depressed and non-depressed subjects. *Science*, 1976, *192*, 489–491.

Tomkins, S. S. *Affect, imagery, consciousness:* Vol. 1. *The positive affects*. New York: Springer, 1962.

Winer, B. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

Wolf, S., & Wolff, H. *Human gastric function*. New York: Oxford, 1947.

Zillman, D., & Bryant, J. The effect of residual excitation on the emotional response to provocation and delayed aggressive behavior. *Journal of Personality and Social Psychology*, 1974, *30*, 782–291.

Received September 1, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 9, 1532–1538

Deindividuation and Valence of Cues: Effects on Prosocial and Antisocial Behavior

Robert D. Johnson
Arkansas State University

Leslie L. Downing
Union College

Deindividuation has been shown to relate to increases in antisocial behavior. Typical manipulations, however, have confounded deindividuation with the presence of negatively valenced cues, such cues being inherent in the costumes or situations used to produce deindividuation. The present study manipulated deindividuation and valence of costume cues in a 2 × 2 factorial design. Zimbardo's theory of deindividuation suggests that deindividuation should disinhibit antisocial behavior, independent of cue valence, and should reduce any influence due to cues. Gergen, however, suggests that cues may have increasing influence, given deindividuation, and that deindividuation may increase prosocial behavior, given positive cues, and increase antisocial behavior, given negative cues. Results supported Gergen's position. Given options to increase or decrease shock level received by a stranger, no main effect was found for deindividuation. There was a main effect for costume cues, and an interaction of cues with deindividuation, with deindividuation facilitating a significant increase in prosocial responses in the presence of positive cues and a nonsignificant increase in antisocial responses in the presence of negative cues. Also cues interacted with trial blocks, prosocial behavior increasing with positive cues and antisocial behavior increasing with negative cues over trial blocks.

The construct deindividuation was first systematically investigated by Festinger, Pepitone, and Newcomb (1952). In their view, one consequence of an individual's involvement and identification with a group is a reduction of individual responsibility for individual behavior. They believed that deindividuation was a phenomenon on which groups would differ and that degree of deindividuation in a group could be indexed by the rate of failure of individuals to identify correctly the group members who had contributed different behaviors. It was expected that previously inhibited behaviors (i.e., saying negative things about one's parents) would be disinhibited as a result of deindividuation. As predicted, groups scoring highest on deindividuation did say more negative things about their parents. This was, however, a correlational study in which the direction of causality between deindividuation and disinhibition is unclear. Later studies (Singer, Brush, & Lublin, 1965; Zimbardo, 1970) using lab coats and hoods to obscure the identifiability of individuals, have shown that such a manipulation increases disinhibition of socially undesirable behaviors (i.e., speaking obscene words and administering electrical shock to another person).

Theoretical Issues

Exactly which behaviors will be disinhibited in a given situation is not yet clear. Nor is it clear how situational cues affect the type or amount of disinhibition that occurs. Two major positions on these issues have been advanced.

This article is based upon data submitted by the first author in partial fulfillment of requirements for the PhD degree at the University of Georgia, 1975. The second author was dissertation advisor. We thank Abe Tesser, Sid Rosen, Dick Endsley, and Dave Shaffer for their contributions to the project.

Portions of this article were presented at the 83rd Annual Convention of the American Psychological Association, 1975.

Requests for reprints should be sent to Leslie L. Downing, Department of Psychology, Union College, Schenectady, New York 12308.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3709-1532$00.75

Zimbardo's view. In Zimbardo's theory of deindividuation (1970), anonymity, along with other input variables, produces a state of the organism, *deindividuation*, that in turn produces a general disinhibition of previously inhibited behavior. Negative comments about parents, college women's use of obscene words, and subjecting others to pain are all behaviors inhibited by prior experience, possibly by expectation of punishment. Deindividuation disinhibits such behaviors by means of a general weakening of inhibitory mechanisms. Which behaviors will be disinhibited depends upon which ones have been inhibited. Inhibited behaviors, once initiated, will tend to increase in frequency and intensity because they are intrinsically reinforcing. According to this theory, because the real source of the behavior is its intrinsically self-rewarding nature, deindividuation should lead to decreasing influence of external cues.

These hypotheses were subjected to an experimental test using what is perhaps the best-known deindividuation paradigm (Zimbardo, 1970). Groups of subjects were individuated (with identifying name tags) or deindividuated (with lab coats and hoods obscuring their identity) and were then given a sanctioned opportunity to administer electrical shocks to another person. Zimbardo has interpreted the increased duration of shocks given by deindividuated subjects to be a result of the deindividuating experience disinhibiting aggressive behavior.

As predicted, individuated subjects decreased shock over trials for a "nice" target person and increased shock for an "obnoxious" target, whereas deindividuated subjects increased shock for both targets. Also, shock duration and negativity of ratings of the target person were significantly correlated for individuated subjects, $r = .67$, but not for deindividuated subjects, $r = .10$. Contrary to prediction, however, the deindividuated subjects' average shock duration was higher for the "obnoxious" than for the "nice" target, whereas during the first half of the trials, individuated subjects actually shocked the "nice" target more than the "obnoxious" one. Deindividuated subjects thus seemed at least initially more "appropriately" responsive to target characteristics than were individuated subjects. Thus evidence for Zimbardo's prediction of reduced influence of situational cues, given deindividuation, is at best equivocal.

Gergen's view. The prediction that antisocial behavior is most likely to be disinhibited by deindividuation was questioned by Gergen, Gergen, and Barton (1973). In their study, darkness- and anonymity-induced deindividuation led not to increased aggression but to increases in touching, caressing, and other affectionate behaviors. These workers advanced the notion that either prosocial or antisocial behavior could be enhanced by deindividuation, depending upon valence of situational cues. In this view, the darkness manipulation may have been more suggestive of intimacy than of aggression, hence anonymity-induced deindividuation increased the frequency of intimacy behaviors.

Experimental Confounds

A more important question is whether the disinhibition observed in prior research was in fact due to deindividuation and not to some unintentionally manipulated variable. One such variable relates to the cue value of the deindividuating costumes, which may be reminiscent of Ku Klux Klan outfits or perhaps of some Halloween ghouls, either of which might be considered cues eliciting aggression (Berkowitz, 1974). It is noteworthy that even without differential identifiability, Berkowitz and his colleagues have demonstrated increased duration of shocks to a target person due to the presence of cues previously associated with aggression (Berkowitz, 1974). A similar line of reasoning could serve as an alternative explanation for differential levels of the use of profanity in the Singer et al. (1965) study. Old clothes (deindividuated condition) may have provided cues for lowered restraints against obscene language, whereas dressy clothes (individuated condition) may have provided cues to a number of learned associations of verbal restraint and propriety. Finally, the dark chamber used by Gergen et al. (1973) to deindividuate subjects

may have provided cues (e.g., darkness, the awareness of the presence of others through the sense of smell, the sounds of breathing, and the sense of touch, even if at first accidental) suggestive of intimacy and thus facilitating intimate behavior. In this study, however, a control condition indicated that darkness cues did not disinhibit intimacy if subjects were not anonymous.

The major question addressed by the present study is whether the direction of behavior change induced by nonidentifiability is influenced by the valence of situational cues. Given an influence of cues, Zimbardo's (1970) theory would predict that nonidentifiability will disinhibit aggressive behavior and that any effect of cues should be less than the effect for identifiable subjects. On the other hand, the reasoning of Gergen et al. (1973) predicts that nonidentifiability will lead to increases in antisocial behavior if antisocial cues are present and to increases in prosocial behavior if prosocial cues are present. A third possibility is that anonymity per se has no effect, but that prior demonstrations of anonymity effects actually resulted from confounded differences in situational cues.

Method

Subjects

Sixty female subjects were recruited from introductory psychology, sociology, and child development courses to participate in a study described as concerned with changes in group evaluations of a stranger. Most subjects were given extra credit in their course as an incentive to participate.

Procedure

Subjects were randomly assigned, 15 to each of four conditions, in a 2×2 factorial design that manipulated individuation (identifiable) versus deindividuation (nonidentifiable) and prosocial versus antisocial cues. Each subject was informed by phone, at the time she was recruited, that the experiment was to be run in groups of four, and that it was therefore very important for her to appear on time. Upon arrival, subjects were informed that each student had been asked to report to a different room in order to preclude interaction with others prior to the experiment. It was also explained that Polaroid pictures of the other group members would be used to establish the essential feeling of being part of a group and that disguises would be worn in the pictures to obscure individual difference characteristics that might be influential. In fact, although as many as four subjects were run in each session, each subject was treated completely independently of the others, and all subjects at any one time were in different treatment conditions. No subject had any contact with any other subject during the experimental session.

It was explained to each subject that the paid nonstudent male volunteer whom they would be evaluating was taking part in a verbal learning experiment and that it was important for them to become involved with the stranger. Involvement was to be established by their participation in selecting the level of shock that this person received for failure to respond correctly in the learning task. Subjects were instructed that it was an exploratory study on the effect of arousal on learning and that the experimenters did not know what effect different shock levels would have. Following each error the learner would be given a shock the base level of which could be increased or decreased by the responses of the subjects. The actual shock level received following an error would be the base level adjusted by the average adjustment selected by the four group members. Each subject would select high, moderate, or slight increases (+3, +2, +1) or decreases (−3, −2, −1) in the shock to be administered. It was explained that following each trial the subject would see on her console the shock selections of the other three subjects plus her own, that these would be averaged to determine the level of shock increase or decrease to be administered on that trial, and that no record would be kept of individual responses but only of the group average for each trial. Thus all subjects' responses, regardless of condition, were to be nonidentifiable to the experimenter and to any others who might see the data. The "learning task" ended when the subject had made 15 errors. The number of correct responses made prior to the 15th error was the measure of successful learning that subjects were to try to facilitate by their shock selections.

Cue manipulation. The costume manipulation of cues was produced by having each subject wear either a robe resembling those of the Ku Klux Klan or a nurse's uniform. The ostensible purpose of the costumes was to obscure individual differences. The nature of the specific costume given to each subject was presented as an accident of convenience (i.e., "I'm not much of a seamstress; this thing came out looking kind of Ku Klux Klannish," or "I was fortunate the hospital *recovery room* let me borrow these nurses' gowns to use in the study.") A Polaroid picture was taken of each subject in her costume, and pictures of others in the group, in similar costumes, were attached to the subjects' consoles. Each subject was told that copies of her picture had been placed on the consoles of others in the group.

Deindividuation manipulation. In the individuation condition, consoles were labeled so that each subject could identify the shock level set by each person in the group and the person by whom it was set.

Also, large name tags were attached to the costumes of the individuals in the Polaroid pictures. In the deindividuation condition, pictures of others in costume were attached to subject consoles, but no name tags were worn, and subjects were provided with no means of identifying the person who made any given response.

To facilitate subjects' involvement with the target, they were allowed to see and hear the "learner" being interviewed. His behavior in the interview was designed to be obnoxious and distasteful to the subjects. A negative target was used in order to minimize the inhibitions against shocking another person, thus allowing for variability on this response dimension. In fact, subjects proved not to be too inhibited to use the shock increase option, yet the target was not so negative as to preclude use of the shock decrease option.

Following the interview, subjects filled out a preliminary evaluation of the confederate. Learning trials were then begun, errors being indicated by the lights on the subjects' consoles, signaling to them to select +3, +2, or +1 levels of increase or −1, −2, or −3 levels of decrease in intensity of the shock the confederate was to receive. After all subjects in the group had responded, feedback of the choices of others was displayed and shock was supposedly administered. Feedback was preprogrammed to average 0 across each of three blocks of trials. Following the learning task, subjects were asked to fill out a postexperimental questionnaire, after which they were probed for suspicion, debriefed, asked not to divulge the deceptions to others, and dismissed.

Results

Evaluation of the Stranger

Five bipolar scales for evaluating the stranger were administered immediately following the interview, prior to the learning task. The experimental conditions did not significantly differ from each other, the means for each condition being on the *insincere, dishonest, cold, phony,* and *unkind* side of the midpoint. It appears that the confederate was perceived to be obnoxious, as planned, in all conditions.

Manipulation checks

Perception of cues. Four bipolar scales on the postexperimental questionnaire were used to assess subjects' perception of costume cues. As intended, the Ku Klux Klan costumes were rated as significantly more tough, harmful, unkind, and cold than were the nurses' costumes, $p < .01$, for each scale. Composite totals of responses to these four scales indicated that the Ku Klux Klan costumes were perceived as significantly more negative than the nurses' costumes, $F(1, 56) = 24.95$, $p < .01$. This difference reflects the negative ratings of the Ku Klux Klan costumes, $M = 5.59$, versus the slightly positive ratings of the nurses' costumes, $M = 3.71$ on a 1 (extremely compassionate) to 7 (extremely aggressive) scale.

Table 1
Mean Shock Selection as a Function of Cue and Deindividuation Conditions

Cue	Condition	
	Individuated	Deindividuated
Prosocial	$-.35_a$	-1.47_b
Antisocial	$.76_c$	$.95_c$

Note. Possible range of scores was from −3 (the prosocial choice of maximally reducing shock level) to +3 (the antisocial choice of maximally increasing shock level). Means without a common subscript were significantly different from each other ($p < .01$).

Perceived sense of deindividuation. As expected, deindividuated subjects, compared to individuated subjects, indicated on the postexperimental questionnaire that it would be more difficult to identify shock selections of other individuals in their group, $F(1, 56) = 15.84$, $p < .01$, and that it would be more difficult to distinguish members of their group from nongroup members following the experiment, $F(1, 56) = 19.55$, $p < .01$. These measures suggest that in fact deindividuation was manipulated as intended.

Shock Selections

The primary dependent variable, shock selection, was analyzed by a $2 \times 2 \times 3$ analysis of variance, individuation versus deindividuation, and prosocial versus antisocial cues between subjects and trial blocks within subjects. This analysis revealed a significant main effect for cues, $F(1, 56) < 46.28$, $p = .01$, with shock decrease the mean response for prosocial cues and shock increase the mean response for antisocial cues. The main effect for individuation versus deindividuation was

not significant, $F(1, 56) = 2.92$, but there was a significant Cues × Deindividuation interaction, $F(1, 56) = 8.21$, $p < .01$. The interaction (see Table 1) is a result of an increasing effect of cues on behavior in deindividuated versus individuated conditions.

A Newman-Keuls test revealed that each condition was significantly different from each other condition, each at the $p < .01$ level, except the simple effect of deindividuation within antisocial cues. This comparison is essentially a conceptual replication of Zimbardo (1970). It was directionally consistent with that effect but only achieved $p < .15$ even by a one-tailed simple t test.

No significant main effect occurred for trial blocks, $F(2, 112) = 1.31$, or for the interaction of trial blocks with deindividuation, $F(2, 112) = 1.85$. There was, however, a significant Trial Blocks × Cues interaction, $F(2, 112) = 19.78$, $p < .01$, resulting from the tendency for shock levels to increase in the antisocial cues condition and to decrease in the prosocial cues condition, from the first to the second of three blocks.

Discussion

The experimental manipulations appear to have been effective. The prosocial costumes, though closer to neutral than had been intended, were rated as significantly less negative than the antisocial costumes. The manipulation of deindividuation was assessed by self-reports of ability to identify behaviors of specific group members (cf. Festinger et al., 1952) and perceived anonymity of group members (cf. Zimbardo, 1970). Both measures demonstrated that deindividuation was manipulated as intended.

Of primary interest in this research were the effects of cues and of deindividuation on shock selection. In the presence of Ku Klux Klan costume cues, subjects were likely to increase shock levels, whereas in the presence of prosocial cues, subjects were likely to decrease shock levels. This finding by itself suggests alternative interpretations of the Singer et al. (1965) study and the Zimbardo (1970) study presented in this paper's introduction. In those studies costume cues were completely confounded with the manipulations of deindividuation. It is entirely possible that the increased antisocial behavior observed in those studies (i.e., frequency of obscene words and duration of electrical shocks) was a function of the costume cue manipulation alone and depended less on anonymity than has been widely believed.

In contrast to the previous research, the present study allows us to look at the effects of deindividuation independent of costume cues. According to Zimbardo, (a) deindividuation should have resulted in more aggressive, antisocial behavior for both costume variations, and (b) any effect of cues should have been attenuated as a function of deindividuation. According to Gergen et al. (1973), however, cues should interact with deindividuation, deindividuation leading to more antisocial behavior in the presence of antisocial cues and to more prosocial behavior in the presence of prosocial cues. The interaction obtained in the present study (see Table 1) is consistent with Gergen's position.

Theoretical Implications

Antecedents of deindividuation. Conceptually, most theorists (cf. Zimbardo, 1970) have viewed deindividuation as a state of the organism that can be induced by a variety of input variables, including but not limited to anonymity. Experimentally, however, deindividuation has nearly always been manipulated by varying some aspect of identifiability. This is true of the research discussed in the introduction and is true of the present research. More recently (cf. Diener, in press), deindividuation has been manipulated by means of complexes of input variables including a sense of group unity, group cohesiveness, group responsibility, and even kinesthetic feedback from physical activity. These inputs may in fact induce a sense of anonymity, but very likely they have additional influences beyond those related to identifiability.

The state of deindividuation. Confirmation of the relationship between anonymity and disinhibition is not sufficient to support

the construct of deindividuation as a mediating state of the organism, and attempts to validate the construct have been largely unsupportive (Diener et al., 1975; Diener, in press). Anonymity-induced disinhibition may not require a reduction in the subjective sense of individuation but in many instances could reflect a simple reduction in perceived negative sanctions, hence a disinhibition of behavior previously suppressed by such sanctions. Possibly environmental cues, like demand characteristics, influence the subjects' interpretation of the types of prior sanction that have been lifted.

Individuation and objective self-awareness. Conceptually, a strong parallel seems to exist between Zimbardo's (1970) constructs of individuation versus deindividuation and the Duval and Wickland (1972) constructs of objective versus subjective self-awareness. For example, both individuation and objective self-awareness imply focusing of attention inward, toward the self. Recent research has shown that a typical manipulation of objective self-awareness does in fact lead to an increased sense of individuation (Ickes, Layden, & Barnes, 1978). Also, Diener (in press) has demonstrated a reduction in self-awareness resulting from a broad-based group involvement manipulation of deindividuation but has found this to be not equivalent to lack of self-awareness induced in a nonsocial way. This does not imply that anonymity will not yield a lack of self-awareness comparable to that of typical objective self-awareness manipulations, for Diener's manipulations of deindividuation included much more than anonymity.

Objective self-awareness and situational cues. The state of objective self-awareness has been shown to increase the influence of internalized standards for correct or appropriate behavior (Carver, 1975; Scheier, Fenigstein, & Buss, 1974). Conversely, one might expect objective self-awareness to reduce the influence of external, situational cues. If anonymity-induced deindividuation is comparable to reduced objective self-awareness, the anonymity should increase the relative influence of external cues. Basically, we are suggesting that reduced objective self-awareness, anonymity-induced deindividuation, and probably group involvement-induced deindividuation all lead to an outwardly focused attention with increased salience of and influence due to concurrent situational factors such as the cue manipulation in the present study. Also, individuation and objective self-awareness should focus attention inward, increasing the salience of and influence due to internalized standards for behavior (cf. Carver, 1975).

Under some circumstances, internalized standards may themselves be altered by concurrent stimulus information (cf. Carver, 1974), in which case the influence of such external stimuli may be increased by objective self-awareness. Consequently, statements as to whether anonymity will lead to increased or decreased influence of external cues may need to be qualified to take into account how or whether those cues affect a change in internalized standards for behavior.

References

Berkowitz, L. Some determinants of impulsive aggression: Role of mediated associations with reinforcements for aggression. *Psychological Review*, 1974, *81*, 165–176.

Carver, C. S. Faciliation of physical aggression through objective self-awareness. *Journal of Experimental Social Psychology*, 1974, *10*, 365–370.

Carver, C. S. Physical aggression as a function of objective self-awareness and attitudes toward punishment. *Journal of Experimental Social Psychology*, 1975, *11*, 510–519.

Diener, E. Deindividuation, self-awareness, and disinhibition. *Journal of Personality and Social Psychology*, in press.

Diener, E., Dineen, J., Westford, K., Beaman, A. L. & Fraser, S. C. Effects of altered responsibility, cognitive set, and modeling on physical aggression and deindividuation. *Journal of Personality and Social Psychology*, 1975, *31*, 328–337.

Duval, S., & Wicklund, R. A. *A theory of objective self-awareness*. New York: Academic Press, 1972.

Festinger, L., Pepitone, S., & Newcomb, T. Some consequences of deindividuation in a group. *Journal of Abnormal and Social Psychology*, 1952, *47*, 382–389.

Gergen, K. J., Gergen, M. M., & Barton, W. H. Deviance in the dark. *Psychology Today*, 1973, *7*, 129–130.

Ickes, W., Layden, A. L., & Barnes, D. Objective self-awareness and individuation: An empirical link. *Journal of Personality*, 1978, *46*, 146–161.

Scheier, M. F., Fenigstein, A., & Buss, A. H. Self-awareness and physical aggression. *Journal of Experimental Social Psychology*, 1974, *10*, 264–273.

Singer, J. E., Brush, C. A., & Lublin, S. C. Some aspects of deindividuation identification and conformity. *Journal of Experimental Social Psychology*, 1965, *1*, 356–378.

Zimbardo, P. G. The human choice: Individuation, reason and order versus deindividuation, impulse, and chaos. In W. J. Arnold, & D. Levine (Eds.), *Nebraska Symposium on Motivation* (Vol. 18). Lincoln: University of Nebraska Press, 1970.

Received September 1, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 9, 1539-1553

Social and Emotional Messages of Smiling: An Ethological Approach

Robert E. Kraut and Robert E. Johnston
Cornell University

Did smiling evolve as an expression of happiness, friendliness, or both? Naturalistic observation at a bowling alley ($N = 1,793$ balls) shows that bowlers often smile when socially engaged, looking at and talking to others, but not necessarily after scoring a spare or a strike. In a second study, bowlers ($N = 166$ balls) rarely smiled while facing the pins but often smiled when facing their friends. At a hockey game, fans ($N = 3,726$ faces) smiled both when they were socially involved and after events favorable to their team. Pedestrians ($N = 663$) were much more likely to smile when talking but only slightly more likely to smile in response to nice weather than to unpleasant weather. These four studies suggest a strong and robust association of smiling with a social motivation and an erratic association with emotional experience.

Everyday experience suggests that smiling is one of the most common nonverbal signals used in communication among humans. Despite this, and despite more than 100 years of research on facial expressions, we still know relatively little about the causation of smiling and its social functions. In this article we attempt to provide evidence about the causation of smiling in social settings and to raise some neglected questions about the analysis of facial expressions in general.

Research and thought on the facial expression of emotion has had a checkered history since the publication of Darwin's *The Expression of the Emotions in Man and Animals* in 1872, as has been documented by a number of excellent reviews (Ekman, 1973; Ekman, Friesen, & Ellsworth, 1972; Izard, 1971), but recently several of the perennial questions in this field have been settled. In a variety of studies researchers have shown that people can consensually and accurately recognize at least six emotional expressions from pictures of faces (e.g., Ekman, 1972; Ekman et al., 1972; Tomkins & McCarter, 1964) and that these abilities seem to be universal among humans (Eibl-Eibesfeldt, 1972 and 1973; Ekman, 1973; Izard, 1971; Vinacke, 1949; Vinacke & Fong, 1955). Ekman (1972) has proposed a general model for the factors influencing facial expression that we take as representative of this "emotional expression approach" to facial expressions, and we explicate this model in more detail below for the particular case of smiling. Tomkins' (1962) and Izard's (1971) positions emphasize the influence of facial expressions on emotional experience more than Ekman's approach does but are otherwise similar.

The Emotional Expression Approach: Smiling as the Expression of Happiness

According to the emotional expression view, a smile is the major component of a facial display associated with and caused by feelings

This research was supported in part by National Institutes of Health Grant MH 30041 to Robert Kraut and National Science Foundation Grant BNS 76-18681 to Robert Johnston. Peter Goldenthal, Virginia Engelhard, David Eaton, Celia Littauer, Cynthia Morgan, Sherrie Zweig, and Barbara Peterson provided valuable discussion in early stages of the research and helped collect data. Scott Hamilton, manager of Ide's Bowling Lanes in Ithaca, New York, was both generous and patient about our research efforts.

Requests for reprints should be sent to Robert E. Kraut, Departments of Psychology and Sociology, Uris Hall, Cornell University, Ithaca, New York 14853.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3709-1539$00.75

of happiness or joy. Anything that makes a person feel good or happy should produce smiling unless the individual wants to mask or inhibit this display. Laughing is considered to be the expression of either more intense happiness (Darwin, 1872/1965) or a particular type of happiness (Ekman & Friesen, 1975). Cultural and individual differences influence smiling both by determining the interpretation of events, which affects the cause of happiness, and by shaping display rules, which determine when it is socially appropriate to smile. But such differences do not alter people's innate and universal tendency to smile when they are happy. Thus, when a smile does occur, the message is usually happiness (Ekman & Friesen, 1975), although this may be a false message if the sender is masking another emotion with a smile or if the sender is simulating happiness for some other reason.

It is important to note that although workers in this tradition have emphasized the importance of facial expression in communication and social behavior, they have rarely studied such communication in natural social settings by studying the causes and consequences of smiling; rather, they have focused on the recognition and verbal labeling of emotions in facial expressions, generally in still photographs.

Ethological Studies: Smiling Indicates Friendliness

A different paradigm has of necessity been used by ethologists studying nonhuman primates. These workers have used naturalistic observation as a research tool and have drawn many of their hypotheses about humans from comparisons of humans with other primates; they have concentrated on the proximate causes of smiling, its consequences for the immediate interaction, and its evolutionary functions.

Many nonhuman primates have a submissive facial display, called a grimace, a grin, or a silent bared-teeth face. The display resembles the human smile, and in all species in which it occurs, it seems to have the function of deflecting hostile behavior of more dominant animals (Hooff, 1962). In a detailed study of chimpanzees, Hooff (1972, 1973) showed that in addition to averting attacks, variations of this display were used to maintain or increase affiliative behaviors between individuals. In some circumstances dominant animals used one variety of the bared-teeth face to reassure subordinates of their nonhostile, affiliative intent. Hooff hypothesized that the human smile is evolutionarily related to the chimpanzee's bared-teeth displays and serves the same functions of deflecting hostility and maintaining friendly contact. On the other hand, according to Hooff, laughter evolved independently and is related to the primate "play face."

If human smiling is a behavioral homologue of chimpanzee bared-teeth displays, one would expect smiling to occur most in face-to-face interaction, especially where friendly intent is problematic or where social bonds are being established or renewed. The smiler's motivation may be genuine friendliness or an intent to establish friendly relations. Researchers in this ethological tradition have not been concerned with the emotions or feelings experienced by those doing the smiling.

The two approaches outlined above are not necessarily incompatible, but they have talked past each other by using different methodologies on different species to ask different questions. The present article tries to compare predictions based on smiling as the expression of happiness with those based on smiling as an indication of friendliness, since the most straightforward extrapolations from each position do lead to different predictions about the causes of smiling in social settings. We asked about the motivational state of the smiler and the conditions under which smiling is produced, and we used naturalistic observation as our methodology. We chose public settings in which we could observe people's faces, in which the two theoretical orientations would predict that smiling would occur frequently, and in which we could distinguish the two theories. Thus, for a setting to be relevant to the emotional orientation, emotionally arousing and happiness-producing events had to be frequent. To be relevant to the social orientation, social interaction had to be frequent.

Study 1: Naturalistic Observations of Bowling

Watching bowlers is an excellent way to distinguish the social and emotional hypotheses about the causation of smiling. Observations of the progress of the game and the actions of friends and teammates provide evidence about external events that might cause smiles, and observations of the bowlers' behaviors as they are smiling provide evidence about their motivational states.

According to the emotional hypothesis, bowlers should smile whenever they feel happy, for example, immediately following a spare or strike. But according to the social hypothesis, smiling should occur during social interaction, and the score obtained should be irrelevant.

To some extent, the entire game of bowling is played in a social context. Bowlers generally play with several friends or teammates with whom they talk and drink between turns and who shout encouragement, taunts, and insults after the play (see White, 1955, for a lively description). Yet during the game, the extent to which bowlers are engaged in social interaction varies greatly from moment to moment. When bowlers are facing the pins, preparing to roll the ball and watching the outcome, social interaction is minimal. It increases when they turn to face friends and is greatest when they are engaged in face-to-face interactions in close proximity to them. The social hypothesis predicts that smiling should occur most during these bouts of more intense social involvement.

Method

The bowling alley in which we made observations had 36 lanes. The lanes were set up in pairs with a scoring table centered about 4.5 meters behind the foul line and a semicircular bench defining a pit for bowlers waiting their turn. Behind the pit and a guardrail was a gallery with small tables at which spectators and bowlers might sit, watch the game, and drink.

Although there were many variations, most bowlers followed a predictable sequence. They would arise from a seat in the pit, select a ball from the ball return located between the two lanes, approach and release the ball at the foul line, stand or back up while facing the pins to see the outcome of the roll, turn 180° to face the pit, and walk toward the pit. After turning from the pins, they might show a communicative display or engage in social behavior, and they would then go back to the ball return and start the second roll. At the end of the second roll, bowlers would return to their seats or leave the pit, occasionally talking to people on the way.

We made observations of bowlers who appeared to be at least 18 years old. Lanes to observe were arbitrarily chosen from among those with an open spectator table located about 7 meters behind the foul line. If no spectator tables were available, observers arbitrarily selected a lane and observed while leaning against a wall. An observer recorded behavior from all bowlers using a selected lane for a maximum of 20 rolls of the ball from all bowlers. At the end of 20 rolls, the observer moved to another arbitrarily chosen lane. Most bowlers were unaware that they were being observed by us. Those bowlers who asked were told that we were watching their technique and their score.

The observer recorded verbal and nonverbal data about bowlers from the time they turned to face the pit after bowling until they turned away to pick up the ball after the first roll or returned to their seat, left the pit, or turned away from the observer after the second roll. The mean length of an observation period for a sample of 120 rolls was 5.5 seconds. A pattern of body and facial behavior believed by the observer to have occurred simultaneously was recorded as a unit. If an element in the display changed or if the same pattern was held for what seemed a very long time to the observer, the observer recorded another unit.

Our observational study of bowlers was replicated three times, using slightly different lists of behaviors, different observers, and different recording techniques. Over all replications we observed 1,793 rolls of the ball, based on approximately 350 different bowlers. Observers were trained by watching and recording from videotapes of people bowling. In the first replication, observers were five students in a seminar on human ethology and the two authors. We were aware of the hypotheses of the study and recorded behaviors using a brief notational scheme with pencil and paper. In the second replication, two observers naive to the hypotheses of the study watched 550 rolls of the ball and recorded behaviors by speaking code names of the behaviors into portable tape recorders, pausing between behaviors in order to indicate unit boundaries. Using this technique, observers didn't have to remember behavioral sequences as long and didn't have to look away from the research subjects to record data; they were therefore able to record more behaviors and units per observational period.

In the third replication we videotaped 155 rolls of the ball and made detailed analyses of these tapes. Recordings were taken with bowlers' awareness and permission. We set up the lights and camera at the spectator table and pretended to film for 5 minutes prior to the actual filming to acclimate subjects to

the equipment. Two observers recorded data from these tapes. As soon as both eyes of the subject could be clearly seen, the observers stopped the tape, recorded behaviors, hand-turned the tape one quarter-turn, recorded behaviors, and repeated this sequence until the subject moved out of focus or turned so that only one eye could be seen. Using this method, data were recorded at approximately two-thirds-of-a-second intervals. After this stop motion analysis, the tapes were played at least once at normal speed to validate the behaviors recorded.

We did no reliability checks on the live recording. On the analysis of the videotaped sequences the two observers agreed on 97% of the behaviors that either recorded.

We recorded the following behaviors because prior research suggested that they might be communicative and because pretesting showed that they had occurred with sufficient frequency to allow meaningful statistical analyses:

Neutral face: blank expression; mouth relaxed; head straight forward; absence of all other coded behaviors except *talked to, groom,* and *good score.*

Closed smile: corners of the mouth turned up; lips together; teeth together.

Open smile: corners of mouth turned up; lips are parted to show teeth. In Replication 1 only, smiling was coded without distinguishing between closed- and open-mouth smiles.

Laugh: mouth open, corners of mouth sometimes turned up; laughter accompanying open mouth.

Tight lips: lips compressed tightly; mouth in straight line; teeth probably clenched.

Look: gaze fixed on another person.

Look down: gaze directed at floor.

Look away: head turned off body axis, gaze not directed at others in group.

Talk: vocalization by subject to another person.

Talked to: vocalization directed to the subject.

Groom: subject preening, scratching, or rubbing any part of his body or face.

Face cover: one or both hands covering facial features for several seconds. This was not recorded in Replication 2.

Head shake: a continuous horizontal movement of the head, usually repeated several times.

Ham: a nonspontaneous, exaggerated facial or body expression apparently intended to communicate; for example, "funny" faces, sticking out tongue, wrinkling nose, jumping up and down, dancing a "jig."

Positive exclamations: spontaneous and often exaggerated actions or words that the observer believed corresponded to a pleasant experience or one during which the individual felt pleasurably excited. For example, a leap and squeal of joy after a strike.

Negative exclamations: spontaneous and often exaggerated actions or words that the observer believed corresponded to an unpleasant experience. For example, swinging the fist across one's body, stamping one's foot, or swearing after missing the pins. Exclamations were more spontaneous, less exaggerated, and shorter, and appeared to be less intentially communicative than hamming, with which they might be confused.

Good Score: a strike or spare.

Sequential Analysis

In order to understand the causes of smiling, it is necessary to examine the events with which it is temporally associated. If smiling often followed an external event in the smiler's environment, it is possible and even likely that the event caused the smile. If smilers often performed other behaviors while smiling, it is likely that the motivation underlying the other behaviors is also underlying the smiling.

Whether we consider two events or behaviors temporally associated depends on the time unit we select; the meaning of *simultaneous* depends on temporal resolution. In Replication 1 we have considered two behaviors as temporally associated or co-occurring if the original observer thought that they had occurred simultaneously, that is, placed them in the same behavioral unit, or if the observer placed them in adjacent behavioral units. Replication 2, using tape recorded data collection, and Replication 3, using analysis of videotapes, had finer temporal resolution. Therefore, to make all data analyses comparable, we considered two behaviors as co-occurring if they occurred within 3 behavioral units of each other in Replication 2 and within 8 behavioral units of each other in Replication 3. This means that we considered two behaviors as temporally associated if they occurred within about 4 seconds of each other.

Because the frequency of co-occurrence is biased by the frequency of each behavior, Yules Q or gamma is the appropriate measure of association to use. We use this statistic rather than the more familiar phi correlation because we considered a positive or negative association between two behaviors as perfect if the more frequent behavior always or never occurred with the less frequent behavior (Blalock, 1972, pp. 298–299).[1]

---

[1] The formula for gamma in this research is as follows. Let A = the frequency of co-occurrence. T = total behavioral units; this is the total number of opportunities for co-occurrence. Ant = the frequency

Results

The data were stable and the co-occurrence of smiling with other behaviors is similar across the three replications. The mean Pearson correlation between columns in Table 1 is .80 ($p < .001$).

The social hypothesis leads us to expect that bowlers will smile when they are socially motivated independently of emotional experience. Table 1 provides evidence of this. Bowlers were more likely to smile when they were engaged in social contact, for example, when looking at their friends ($M\ \gamma = .71$), and less likely to smile when they were temporarily avoiding social contact for some reason and were looking at the ground or away from their friends ($M\ \gamma = -.56$). In addition, bowlers showed a tendency to smile more when they were talking to friends or being talked to by them ($M\ \gamma = .20$).

Smiling, especially open-mouth smiling, also occurred with playful behaviors; laughter ($M\ \gamma = .23$ over all smiles; $M\ \gamma = .80$ for open-mouth smiles), hamming ($M\ \gamma = .33$ for all smiles; $M\ \gamma = .47$ for open-mouth smiles), and face-covering ($M\ \gamma = .34$ for all smiles; $M\ \gamma = .93$ for open-mouth smiling). At the bowling alleys, open-mouth smiling occurred when bowlers were being socially playful. Since nonsocial playfulness rarely occurred in this setting, we cannot tell if the social or the playful motivation was a more important determinant of open-mouth smiling.

The happiness hypothesis leads us to expect that bowlers would smile more after playing well and bowling a good score, but this was not the case. The association between smiling and scoring a spare or strike was weak ($M\ \gamma = .13$). Of the 1,793 rolls for which we collected data, 448 were spares or strikes. Bowlers smiled 30% of the time after good scores and 23% otherwise. Other support for the happiness hypothesis is mixed. In general, smiling showed weak associations with the behaviors that we had identified as subtle indicators of negative affect and stronger associations with larger scale emotional displays. Thus, smiling had little association with the subtle cues of grooming and head shaking. Smiling had a substantial negative association with the tight-lip display, which our observations had led us to believe was an indicator of anger, frustration, and perhaps determination. It is a component in the traditional anger display (Ekman & Friesen, 1975). However, the negative association could have been partly caused by the physical difficulty a bowler would have smiling and lip pressing simultaneously (and the difficulty coders would have distinguishing two behaviors in the mouth region) as well as by the emotional incompatibility of smiling and lip pressing.

On the other hand, the spontaneous emotional displays we have termed positive and negative exclamations showed substantial associations with smiling. When bowlers were communicating surprise and glee, generally after getting a good score, they smiled ($M\ \gamma = .49$); when they swore and showed other signs of disappointment and anger, generally at not getting a good score, they failed to smile ($M\ \gamma = -.49$). Since smiling showed no association with score, which is the likely cause of bowlers' emotional experience, these data suggest that when bowlers were attempting to communicate their happiness through positive exclamations, they used smiles as part of the communication; when they were experiencing positive emotions but were not attempting to communicate them, however, smiling did not covary with other subtle behavioral indicators of their emotional state.

Several problems limit confidence in the conclusions one can draw from these co-occurrence data. As we have suggested above, the degree of temporal co-occurrence between behaviors may reflect as much their physical

---

of the antecedent behavior; this is the maximum number of times it could have co-occurred with a subsequent behavior. Sub = number of behavioral groupings (i.e., two behavioral units in Replication 1, three units in Replication 2, and eight units in Replication 3) in which a subsequent behavior occurred; this is the maximum number of times it could have co-occurred with an antecedent behavior. Then $B = Ant - A$; $C = Sub - A$; $D = T - A - B - C$; $\gamma = [(A \times D) - (B \times C)]/[(A \times D) + (B \times C)]$. Since the gamma based on Behavior X preceding Behavior Y is not necessarily the same as Y preceding X, we used the mean of these two measures in all our analyses.

Table 1
Temporal Associations (Gamma) of Smiling and Other Behaviors

Behavior	Replication 1	2		3		*M* all smiles[a]	*t*
	All smiles	Closed smile	Open smile	Closed smile	Open smile		
Neutral face	−77	−52	−88	−70	−93	−76	10.5
Closed smile	—	—	−19	—	05	−07	—
Open smile	—	−19	—	05	—	−07	—
Laugh	−48	07	91	−06	70	23	.90
Ham	20	08	44	43	50	33	4.09
Look at	70	63	88	53	83	71	11.14
Look down	−59	−48	−82	−40	−66	−59	8.10
Look away		−56	−51	−38	−56	−52	13.95
Talked to	42	05	44	37	−26	20	1.50
Talk	37	08	41	07	−02	18	2.09
Groom	−47	01	−26	−14	04	−16	1.75
Tight lip	−79	−52	−68	−33	−63	−59	7.55
Face cover	28	—	—	−18	93	34	1.07
Head shake	−18	17	−07	12	15	04	.55
Positive exclamation	24	23	71	72	56	49	4.53
Negative exclamation	−65	−59	−40	−16	−64	−49	5.22
Good score	07	37	29	01	−09	13	1.50

Note. Decimal points have been omitted.
[a] In this table significance tests have not been performed on the individual gammas, since the units on which they are based were not independent of each other. They are based on 5,527 overlapping 4-second periods in which two behaviors could co-occur, spread over 1,793 rolls of the ball and over approximately 350 different bowlers. The *t* tests performed on the mean gammas are based on five observations and test whether smiling is associated with the other behaviors reliably over the different replications and types of smiles.

compatibility as their similarity in underlying motivation or external causation. For example, since smiling, laughing, talking, and compressing one's lips are all behaviors done by the mouth, their mutual co-occurrence is limited. Similarly, the neutral display, by definition, cannot co-occur with other behaviors. A second problem could be a result of observer errors. Faced with a rich and complex event, observers' errors tended to be omissions; our training procedures showed that observers were more likely to ignore behaviors that did occur than to record behaviors that did not occur. What might be termed a climax error may be characteristic of event sequences in which observers underreport low intensity forms of a behavior that gradually change to a climax form of the same or another behavior. For example, it is possible that smiling might be underreported in the sequences leading to laughter, hamming, and positive exclamations, while tight lips and grooming might be underreported in the sequences leading to negative exclamations.

A third problem is more conceptual. Our research strategy has been to infer the messages of smiling by examining other behaviors with which it occurs, that is, marker behaviors. Our beliefs about the significance of these other behaviors were based on the prior literature on nonverbal communication, on informal observation in our research setting, and on intuition. More systematic analysis of the patterns of co-occurrences among marker behaviors could provide direct evidence on the significance of marker behaviors and might provide deeper insight into the messages of smiling.

Principal Components Analysis

One partial solution to these problems is to analyze the similarities in co-occurrence that pairs of behaviors had with other behaviors

rather than analyzing only the temporal association between pairs. For example, if lip tightening and closed-mouth smiling were equivalent but alternative behaviors, an analysis of temporal association would show that they never co-occurred, but an analysis of their similarities in co-occurrence would show that they occurred in exactly the same contexts. Following Hooff's logic (1973), we started with the assumption that to the extent that a pair of behaviors occurred in the same contexts, that is, to the extent that they had similar co-occurrences with other behaviors, they also shared underlying motivations or external causes. The Pearson correlation between rows in a matrix of gamma scores (our measure of co-occurrence) is one measure of the extent to which the pair of behaviors represented by the rows had similar co-occurrences with each other behavior, including co-occurrences with themselves. Because the three replications of our basic study resulted in three estimates of the co-occurrences between each pair of behaviors, we took the median of these three as our best estimate of the co-occurrence. The Pearson correlation between rows was conducted on this matrix.[2]

One can factor analyze this correlation matrix. Table 2 is a varimax rotation of the principle components analysis of this matrix. The five factors that emerged prior to an eigenvalue falling below 1.0 represent 86% of the variance in the original matrix. An examination of the factor structure provides evidence both about the pattern of co-occurrences and the underlying motivations represented by the marker behaviors and about the association of closed- and open-mouth smiling with these underlying motivations.

On what factors did smiling load highly? Factor 3 seems to represent a social motivation. Talking, being talked to, looking at another, and not looking down or away all load highly on this factor. Closed-mouth smiling has its highest loading on this factor, and in addition, open-mouth smiling and laughing both loaded highly here. This result supports the previous analysis and clearly suggests the social motivation underlying both closed- and open-mouth smiling and laughing.

Table 2
Component Analysis of Similarities in Co-occurrence

Behavior	Factor 1	2	3	4	5
Tight lip	*−87*	02	−34	08	−03
Head shake	*−85*	−24	13	23	−01
Laugh	*84*	−00	31	18	−30
Face cover	*75*	27	05	08	42
Open smile	*60*	32	44	−24	43
Good score	18	*92*	02	−23	−19
Negative exclamation	−23	*−91*	−09	−19	−11
Positive exclamation	−08	*70*	30	*−54*	06
Talk	13	−10	*92*	−04	−11
Talked to	12	11	*87*	−19	08
Closed smile	10	44	*64*	−16	29
Look at	31	41	*63*	−20	47
Look down	*−57*	−14	*−61*	30	−05
Ham	08	08	27	*−89*	15
Groom	−34	−05	−08	*75*	−14
Neutral face	−25	24	−07	17	*−81*
Look away	−09	−22	−47	16	*−67*

Note. Decimal points have been omitted. Loadings greater than or equal to 60 are shown in italics.

Smiling shows a complex pattern of relationships with the emotional displays. Factor 2 represents the explosive emotional displays associated with extreme scores: good scores and positive exclamation loaded highly, and

[2] In Replication 1, closed-mouth smiles were not distinguished from open-mouth smiles and looking down was not distinguished from looking away. In computing the matrix of median gammas, the undifferentiated categories of Replication 1 were generally averaged with each differentiated category of Replications 2 and 3. For example, the association between smiling and grooming in Replication 1 contributed in the final matrix to both the association of closed-mouth smiling and grooming and the association of open-mouth smiling and grooming. However, in a few cases closed and open smiles and looking down and away had substantially different associations with another behavior in both Replications 2 and 3 (mean difference in gamma > .30). In these cases, Replication 1 was ignored in computing the final measure of association. Specifically, Replication 1 was ignored in computing the median relationship of closed- and open-mouth smiling to laughter and to face cover.

negative exclamations loaded negatively. The association of closed-mouth smiles with this set of behaviors may reflect the role of closed-mouth smiling in the communication of emotion. Factor 1 seems to represent the contrast between affect and playfulness. The tight lip, head-shake, and looking-down displays cluster at one pole; and laughter, face covering, and open-mouth smiling cluster at the other. This factor may reflect the incompatibility of genuine playfulness with feeling bad.

Factors 4 and 5 are especially difficult to interpret. Factor 4 primarily represents the contrast between hamming and positive exclamations, on the one hand, and grooming, on the other. Factor 5 may indicate expressiveness. The neutral and looking-away displays are at one extreme, and open-mouth smiling and face covering are at the other.

In summary, both the temporal co-occurrence of smiling with other behaviors and the factor analysis of similarities of co-occurrence suggest that bowlers smiled when they were being social, when they were being playful, or when they were otherwise communicating an emotional statement to an audience. Both closed-mouth and open-mouth smiling shared a nonemotional, social motivation. In addition, open-mouth smiling was similar to laughter in adding a playful motivation; to the extent that one cannot be playful and at the same time distressed, angry, or disappointed, open-mouth smiling was incompatible with negative affect.

Study 2: Bowlers Facing the Pins or Their Companions

One might dismiss our failure to find a strong association between smiling and emotional experience in Study 1 with the claim that bowlers were masking their emotional experiences; they hid the joy or disappointment they felt in order to appear modest or sportsmanlike. To meet this objection we made observations of people who were bowling alone and were therefore under no pressure to mask their emotions. Lone bowlers rarely showed any facial displays or other gestures but instead maintained a generally neutral face. The most common expressions seen were relatively antisocial or negative—looking down, tight lips, and negative exclamations; they rarely smiled. However, one may object to these data, since people who bowl alone may be the type of person who is unexpressive in all circumstances.

Table 3
Percentage of Bowlers Smiling According to Bowling Score and Social Focus

Social focus	Score		
	Good	Not good	Total
Yes	42	28	31
No	4	3	3
Total	46[a]	31[b]	—

Note. $N = 116$.
[a] $n = 26$.
[b] $n = 90$.

Therefore, to further examine the social and emotional messages of smiling, we looked at bowlers when they were facing the bowling pins and reacting to their rolls, and at the same bowlers when they turned to face their friends. The social hypothesis would lead us to expect very little smiling as bowlers faced the pins, since this is a relatively nonsocial setting. The emotional hypothesis would lead us to expect that bowlers would smile after rolling a good score and would engage in behaviors revealing negative affect after bad scores, regardless of their social orientations. When bowlers roll a good score and remain facing the pins, they should feel happy and should not need to mask or hide the expression of this emotion, since they believe that they are not being observed.

Method

An observer knelt on a platform among the pin setting equipment at the end of the bowling alley behind the bowling pins and watched bowlers through binoculars as they finished their roll. The observer was 19.2 meters from the bowlers and observed through a narrow slit in the facade of the pin setting equipment. The observer was invisible to bowlers. As bowlers finished their roll and stepped back while watching its outcome, the observer recorded characteristics of the bowler and the behaviors listed above. Data were recorded from the moment the bowler stepped into view until he turned toward

the pit. Simultaneously, a second observer positioned as in Study 1 recorded behaviors in the standard way after the bowler turned to face friends in the pit. One hundred and sixteen rolls from 34 bowlers were observed from both positions (behind the pins and facing the pit).

Results

As shown in Table 3, bowlers were generally unexpressive while facing the pins, in comparison with their behavior when they faced their friends in the pit. Smiles were particularly rare. In 116 observations, bowlers smiled 36 times when facing friends but only 4 times while facing the pins, $t(115) = 6.25$, $p < .001$. Smiling was unrelated to how well they bowled. Only one of the pin-facing smiles was after a good score, although bowlers in this sample rolled 26 strikes or spares. These data clearly support the social hypothesis. People rarely smiled in nonsocial settings, regardless of emotional experience.

Study 3: Smiling by Fans at a Hockey Game

Hockey at Cornell University is probably the most important school sport. Students, faculty, and townspeople line up overnight to get season tickets. In 1977, Cornell's team had finished its regular season first in the Ivy League and was in the playoffs for the Eastern College Athletic Conference Championships. On March 8, 1977 the team faced Rensselaer Polytechnic Institute in the quarter-final at Cornell. As might be expected, the game was sold out, mainly to enthusiastic Cornell fans. Given this situation, the emotional hypothesis about the causation of smiling would predict that Cornell fans would smile more when the hockey match was going well for Cornell than when it was going well for the opposing team. On the other hand, the social hypothesis would predict that fans would smile more when socially interacting with other fans than when watching the game.

Method

A photographer sat in the stands during the hockey game and, using a telephoto lens, took pictures of the spectators across the ice from him (about 30 meters away) immediately following events favorable, neutral, or unfavorable to Cornell's chances of winning the hockey game. About 220 people were photographed at each exposure ($M = 223$; $SD = 21.8$). The photographer used a prefocused camera on a tripod and took pictures while watching the hockey game in order to be uninfluenced by the behavior of his subjects as he was photographing them. Photographs were taken after (a) goals for Cornell or the opposing team, (b) penalties called on Cornell or the opposing team, (c) face-offs, before the puck came into play, and (d) time-killing passes of the puck at center ice. Cornell goals and opposing-team penalties were considered events favorable to Cornell, the opponent's goals and Cornell penalties were considered unfavorable, and face-offs and center-ice passes were considered neutral. The section of the stands that was repeatedly photographed mainly contained season ticket holders and was filled with Cornell fans. There were few opposing fans in the entire arena, and they were segregated in another area.

Each transparency was coded by a person naive to both the hypothesis of the study and the events preceding the photograph. Five randomly selected transparencies were then recoded by a second person to check reliabilities. The transparencies were first scanned for social units, which were defined as a group of two or more spectators, at least one of whom was turned towards the other or others in the group. The two independent coders agreed on 73% of the social units that either identified. Each member of a social unit was defined to be socially involved with others in that unit, and all other spectators were defined to be socially uninvolved. The transparencies were next coded for smiling. The two coders agreed on 73% of the smiles that either identified, 69% for smiles within social units, and 74% for other smiles.

Table 4
Percentage of Spectators Smiling According to Valence of Hockey and Social Involvement

Social involvement	Valence of play for home fans: Good %	Good *n*	Not good %	Not good *n*	Total %	Total *n*
Yes	27	59	22	165	23	224
No	12	1,258	2	2,244	6	3,502
Total	13	1,317	3	2,408	7	3,726

Results

Because the results involving neutral and bad events didn't differ substantially from each other, they have been combined in the analyses that follow. Table 4 shows the per-

centage of fans smiling who were or were not part of a social unit following events that were good or not good for the Cornell hockey team.

We coded a total of 3,726 faces in 16 photographs. Over all photographs, the probability of any one of these faces smiling at the moment that the picture was taken was .069. Some data supported the emotional hypothesis. Spectators were more likely to smile following events favorable to the home team (13%) than following neutral or bad events (4%; $\gamma = .60$; $N = 3{,}726$; $p < .001$). Some data supported the social hypotheses. Regardless of the valence of the events that preceded the photograph, spectators were more likely to smile if they were members of social units (23%) than if they were not (6%; $\gamma = .66$; $N = 3{,}726$; $p < .001$). The effects of being socially involved on smiling were stronger following bad and neutral events ($\gamma = .85$) than following good ones ($\gamma = .45$, $p < .01$), perhaps because the base rate of smiling was lower following bad events.

Study 4: Smiling, Social Interactions, and the Weather

People feel happier on days when the weather is nice. Given this assumption, the emotional orientation would lead us to expect that people walking outdoors would smile more in pleasant than unpleasant weather. As they walk, they can either be socially involved with someone or not. The social orientation would lead us to expect that people would smile more if they are socially involved but that the weather alone should not influence the frequency of smiling.

Method

A single observer made observations of pedestrians twice at each of four public walkways in Ithaca, New York in September and October 1977. Each site was observed once during pleasant weather, when the temperature was between 50 and 70° F. and the sky was sunny or partly sunny, and once during unpleasant rainy weather. Pairs of observations were made at approximately the same time of day and were made within 3 weeks of each other. A total of 663 subjects were observed in the eight observation periods.

Table 5
Percentage of Pedestrians Smiling According to the Weather and Social Interaction

Social interaction	Weather: Good %	Good *n*	Bad %	Bad *n*	Total %	Total *n*
Yes	61	61	57	60	59	121
No	12	264	5	288	8	552
Total	21	325	14	348	—	—

At the site, the observer selected two reference marks about 10 meters apart on the sidewalk to indicate the limits of observation. The next pedestrian approaching the starting mark was selected as a subject. If pedestrians were walking in pairs or larger groups, the person on the extreme left or the person on the extreme right was alternately selected as the subject for that group, before the group passed he starting mark. Subjects were observed for the 8 to 12 seconds it took them to walk between the two reference marks. If they talked to anyone, greeted anyone, or were talked to by anyone at any time during the observation period, they were classified as socially interacting during that period. If they smiled at any time during the observation period, they were classified as smiling during the observation period. Reliability of the smile classification, based on a second observer's judgments of 100 pedestrians, was high ($\phi = .86$).

Results

Table 5 shows the cross classification of social interaction and smiling among pedestrians during days with good and bad weather. These results show that pedestrians are no more likely to engage in social interaction on nice days than on bad ones (19% vs. 17%, $\phi = .02$). They were slightly more likely to smile on nice days than on bad ones (21% vs. 14%, $\phi = .10$, $p < .01$) and were very much more likely to smile if they were conversing with or greeting someone than if they were not (59% vs. 8%, $\phi = .54$, $p < .001$). The effect of social interaction on smiling was slightly stronger on bad weather days than on good days ($\phi = .58$ vs. .47, $z = -1.41$, $p < .10$), but this interaction was minor compared to the main effect of social interaction and may not replicate.

We conducted an extensive pretest in which 1,489 subjects were observed at six sites on

the Cornell campus in the summer of 1977. Good weather was defined as between 65° F. and 75° F. with low humidity, whereas bad weather was over 85° F. with high humidity. Subjects were observed as long as they remained in sight, about 10–20 seconds, and observation times were not standardized across sites. In this pretest, subjects were again slightly more likely to smile on nice days than on unpleasant ones ($\phi = .13$, $p < .01$) and were much more likely to smile while socially engaged than when not ($\phi = .56$, $p < .001$). Here, however, social interaction had similar effects on good weather and bad weather days ($\phi = .58$ vs. $.54$, $z = .89$, $p > .10$).

In both comparisons, social interaction was thus a much more powerful predictor of smiling than was the weather. In the main study, variations in social interaction accounted for 29% of the variance in smiling, whereas variations in the weather and the positive and negative emotions they may have produced accounted for about 1% of the variance in smiling.

Discussion

The Social Hypothesis

Both the present and earlier research provide strong evidence that social involvement is a major cause of smiling, independent of the smiler's emotional state. In each of the four studies described above, smiling was strongly associated with social interaction: talking to and looking at others in Study 1, facing fellow bowlers in Study 2, orienting toward other fans in Study 3, and talking to another person in Study 4. Other researchers have also found that smiling often occurs in a social context. Almost from its first appearance, smiling is socially produced and has social consequences. Human infants, from the age of 1 to 5 months, smile most in response to the human voice and the human face or abstractions of it (Sroufe and Waters, 1976). These smiles seem to be a major determinant of the bond between an infant and its caretaker (Fraiberg, 1977; Spitz & Wolf, 1946). Among nursery school children, smiles are likely to occur in the context of other social behaviors such as pointing, giving, receiving, and talking (Blurton-Jones, 1972). When nursery school children approach a stranger, they often smile, and they are more likely to approach when the stranger smiles and talks to them (Connolly & Smith, 1972). In addition, the smile appears to be a universal component of greetings (Kendon & Ferber, 1973; Eibl-Eibesfeldt, 1972). Even when people smile in response to humorous or other nonsocial stimuli, they smile more in the presence of other people than when they are alone (Mackey, 1976; Leventhal & Mace, 1970).

It could be argued that positing a separate social cause for smiling is a theoretical extravagance. According to this view, the presence of others is just one of the many events that make people happy, and apparent social smiling is mediated by the pleasant emotions the smiler feels in the presence of others. This is undoubtedly true on occasion.

However, there are several reasons to believe that many smiles have purely social causes independent of happiness. The first is parsimony. We need not assume that happiness is a cause when we have evidence of a strong relationship between social stimuli and smiling, but no evidence of a mediating emotion. A theory of friendliness displays, based on independent comparative data, can account for the empirical relationship. The component analysis of Study 1 showed a social, nonemotional motivation behind some smiling (Factor 3 in Table 2): both closed- and open-mouth smiling occurred in the same contexts as each of the social behaviors we recorded at a bowling alley. These in turn were independent of both gross and subtle emotional displays of happiness, disappointment, and anger. This evidence of purely social smiling exists even though smiling has an emotional component under some circumstances (Factors 1 and 2 in Table 2).

The existence of smiles in uncomfortable social settings is further evidence against the hypothesis that social smiles are mediated by happiness. Repeated viewings of some of our bowling videotapes convinced us that some smiling was done to apologize for an especially clumsy performance or for poor bowling, such as dropping the ball before bowling

or bowling a gutter ball immediately after bowling a strike. To examine this possibility more systematically, we have started looking at facial expressions in uncomfortable social settings. If smiling evolved from primate appeasement displays, we would expect that in humans some smiling should occur when a person is trying to placate or appease another, for example when he or she has made a mistake or has violated a social norm and is apologizing for it. An exploratory field experiment suggests that subjects smile more in an appeasement than in a control condition. In the appeasement condition, customers in a store were made to think that they had made a mistake when they interrupted a clerk busy with paperwork who told them "I'm not working here. She (another clerk) will help." In the control condition, customers were made to think that no mistake had been made ("Fine, I'll get it for you, and she will ring it up"). Customers apologized more in the appeasement (11%) than the control (0%) condition. More to the point, customers smiled more in the appeasement condition (28%), when they were presumably trying to rectify a mistake, than in the control condition (5%, $\gamma = .77$; $n = 99$; $p < .001$).[3]

Ekman (1972) also notes that people often smile when experiencing an unpleasant emotion in the presence of others, although he interprets the smile as a mask for a socially inappropriate facial expression that the emotion would cause rather than as an appeasement display. A question for future research is to determine whether an appeasement or a masking hypothesis can better account for the occurrence of smiling in tense or uncomfortable social situations.

The detailed analysis of patterns of behavior in which smiling is embedded provides additional evidence that social smiling need not be mediated by happiness. For example, Kendon and Ferber (1973) have carefully described the behavior in a greeting. They divided the greeting into three stages, a distance salutation in which the participants establish their readiness and willingness for interaction, an approach phase in which one walks toward the other (or both do), and a close salutation that is a prelude to conversation or other social interaction. During the distance salutation, the greeters orient their bodies toward each other, look at each other, and show a greeting gesture such as a head toss, an eyebrow flash, a wave, or a verbal greeting. During the approach phase they decrease the amount they look at each other, until immediately prior to the close salutation, when all participants looked at and then talked to each other. Clearly the greeters were regulating their social contact during the greeting, first agreeing to it, then postponing it, and finally engaging in it when their physical separation made it convenient.

Significantly for the present argument, smiling almost invariably occurred in both the distant and close salutation, when greeters were showing a willingness to greet each other and establish or reaffirm their relationship. Smiling occurred much less frequently and intensely during the approach phase, when the greeters were showing a temporary withdrawal from social contact by looking away from each other. Smiling thus varied with the intensity of social contact. Since emotions are often regarded as diffuse with a gradual decay, it is difficult to account for the rapid and asymmetrical shifts in smiling during a 10–20 second greeting by referring to shifts in happiness.

Through this literature review, we have tried to establish that social involvement is a major cause of smiling and that happiness does not seem to be a necessary mediator. Why do people smile in the presence of others? Drawing on comparisons of humans with other primates, Hooff (1972) argues that most human smiling is affinitive, used in the expression of sympathy, reassurance, or appeasement, that is, that the smiler's motivation is to insure the establishment and maintenance of friendly interaction. The message might be paraphrased "I am friendly" or "I would like us to be friendly for a while." This may occur when friendliness is highly probable, as when two old friends greet each other after an absence, or when friendliness is problematic, as when a client interrupts a conversation between two professionals to ask one of them a question. The smile is an evolu-

[3] Peter Goldenthal, Lou Zambello, and Donna Brown planned and conducted this study.

tionarily designed signal to smooth interaction among members of a species who must cooperate in group living.

The Emotional Hypothesis

In our studies smiling had a weaker and more erratic association with happiness than it had with social interaction. Although people smiled when shouting, jumping, and gesturing after bowling a spare or a strike or when cheering their hockey team's goals, they did not smile much when they had gotten a good bowling score and were alone or had not yet turned to face their friends, or when they were walking down the street alone on a nice day. Given the very strong link between smiling and the experience of pleasant emotions posited by both the popular culture and the 100-year-old Darwinian tradition, our failure to document this link convincingly demands explanation. We will discuss three attempts to reconcile our results with the happiness hypothesis.

First, we may have found more smiling when people were social than when they were happy because our independent variables were not equated for strength. If bowling, hockey games, and walks on fine days do not produce strong pleasant emotions, they could not be expected to produce much smiling either. We believe that in general, smiling as a display of happiness is relatively infrequent in daily life, partly because the strong emotions that may be needed to elicit it are also rare. However, this is not a plausible explanation for the present results. We picked settings and activities that were likely to produce variations in positive and negative emotions. Our observations suggest that strong positive and negative emotions were produced in these settings, but that they were expressed without a consistent relationship to smiling. The positive and negative exclamations during bowling were one indicator of the strength of the emotions produced. Not getting a spare or a strike led to fewer positive exclamations and more negative exclamations, head shakes, and tight-lipped displays. Thus variations in score produced variations in emotional expressions. Although we did not systematically collect data on emotional displays among hockey fans or among pedestrians, our casual observations suggest that hockey fans expressed their excitement, joy, and approval by jumping up and down, clapping, and screaming, whereas pedestrians celebrated spurts of pleasant weather, after days of Ithaca's drizzly gloom or oppressive heat and humidity, by walking with a lilt, whistling, and humming, not by smiling.

A second possibility is that although smiling did indeed have a strong association with happiness, our subjects masked their emotional expressions according to cultural display rules, to confuse their audiences, or to comply with felt normative pressures (Ekman, 1972). They refused to smile when they felt good and used smiling as a mask for negative emotional expressions when they felt bad. However, this too is an unlikely explanation for our results. In both the bowling and hockey settings, subjects were very expressive, probably much more so than in other settings of daily life. For example, the frequency of positive and negative exclamations among bowlers and the strong association of positive exclamations with good scores attest to the freedom of this setting from constraints on emotional expression. In addition, as mentioned earlier, bowlers smiled least when they were alone or were facing away from fellow bowlers and were thus under the least pressure to use display rules.

A third reason why our results linking smiling and happiness were not as strong as we had expected is that we may have been misled by the phrase *emotional expression*. Perhaps because they have underemphasized the functions of emotional displays, earlier commentators, including Darwin (1872/1965) and Ekman (1972), have used the phrase *expression* to mean the outward manifestation of an internal state. While this definition is correct, it is incomplete. To the extent that smiling is linked with happiness, it is an evolutionarily adapted signal that informs other members of the species about the sender's emotional state in order to influence their behavior. Thus, we should expect smiling, like other primate emotional displays such as fear (appeasement) or anger (threat), to be shown to a recipient and to be less frequently seen in the absence of an audi-

ence. This interpretation is consistent with our data. Smiling had its strongest association with emotion-producing conditions when subjects were communicating emotions in the presence of others through additional displays like positive and negative exclamations, but it was only weakly associated with emotion-producing conditions when subjects were socially uninvolved. These tentative results suggest the hypothesis for further research that emotional displays, in general, should be more frequent and more intense in the presence of others, although this trend could be modified by display rules (Ekman, 1972).

A Comparison of the Social and Emotional Hypotheses

The social and emotional hypotheses about the causation of smiling are not incompatible with each other. As we have suggested, social contact may sometimes produce happiness, which in turn may lead to smiling. In addition, people experiencing happiness may show it more in the presence of others.

It is also possible that emotional and social motivations both independently produce facial displays involving what we have termed a smile, but the morphology of the displays may differ. One possibility is that the mouth region has a different shape when expressing happiness, friendliness, or appeasement. Brannigan and Humphries (1972) and Grant (1969) have described several different smiles. Our own attempts to distinguish closed- from open-mouth smiles suggest that they have different underlying motivations, with the closed-mouth smile more purely social and the open-mouth smile more playful and possibly more emotional. Moreover, even if the mouth is the same, other components of the face may differ. The whole face, rather than just the smile per se, carries information about emotion (Ekman et al., 1972; Ekman, Friesen, & Tomkins, 1971). Although the smiling mouth may be the most salient component of a happiness display, other components may often co-occur and differentiate happiness, friendliness, or appeasement smiles. Further research on facial displays needs to move to a finer level of description and categorization in an attempt to differentiate the varieties of smiles and facial displays. Ekman and Friesen's recent work (1978) is a step in this direction.

A Comparison of the Ethological and the Expressive Traditions

The ethological approach to nonverbal communication, from which the social hypothesis derives, differs from other approaches in several ways. Most important, it has remained firmly wedded to evolutionary theory and, as a result, has stayed concerned with the functions of nonverbal displays and their social consequences. This concern with functions and consequences was clearly present in Darwin's original work (1872/1965; see his discussion of the principle of antithesis), but was lost as the study of emotional expression passed through experimental psychology. The expressive approach to nonverbal behavior, from which the emotional hypothesis about smiling derives, has focused on the correspondence between individuals' internal states and their facial and other expression. As a result it has often embedded the study of nonverbal behavior in individualistic psychology by treating individuals as socially encapsulated. It is true that the expressive approach has studied communication in the limited sense of establishing that information about emotions can be transmitted through facial expression. With its almost exclusive reliance on the recognition experiment, however, this approach has not shown that people use information from facial expressions in daily life.

The effects of this neglect, while perhaps unintentional, are large. Ekman et al. (1972), after reviewing more than 10[illegible] studies on the facial expression of emotio[illegible] published in the 100 years since Darwin's original work (1872/1965), found only a few that investigated causation or production of emotional displays and *none* on the effects of emotional expression on subsequent social interaction.

In contrast to the expressive approach, the ethological approach to human nonverbal behavior treats the individual as part of a social network and examines the interactions between people and the effects of nonverbal be-

havior on others. By studying usage, this approach guides investigators toward a careful, descriptive analysis of the situations in which smiling occurs. The types of situation we studied were chosen to compare the emotional and social hypotheses about the causes of smiling, but we, along with others, have not looked at its effects. Indeed, we know of no research on human nonverbal communication from any tradition that has simultaneously studied the social and motivational causes, the morphology, and the social consequences of a human display, as Hooff has done with chimpanzee bared-teeth displays. This holistic approach is the direction, we think, that future work on human nonverbal communication should take.

References

Blalock, H. M. *Social statistics.* New York: McGraw-Hill, 1972.

Blurton-Jones, N. G. Categories of child-child interaction. In N. G. Blurton-Jones (Ed.), *Ethological studies of child behavior.* Cambridge, England: Cambridge University Press, 1972.

Brannigan, C. R., & Humphries, D. A. Human nonverbal behavior, a means of communication. In N. G. Blurton-Jones (Ed.), *Ethological studies of child behavior.* Cambridge, England: Cambridge University Press, 1972.

Connolly, K., & Smith, P. K. Reactions of pre-school children to a strange observer. In N. G. Blurton-Jones (Ed.), *Ethological studies of child behavior.* Cambridge, England: Cambridge University Press, 1972.

Darwin, C. *The expression of the emotions in man and animals.* Chicago: University of Chicago Press, 1965. (Originally published, 1872).

Eibl-Eibesfeldt, I. Similarities and differences between cultures in expressive movements. In R. A. Hinde (Ed.), *Non-verbal communication.* Cambridge, England: Cambridge University Press, 1972.

Eibl-Eibesfeldt, I. The expressive behavior of the deaf- and blind-born. In M. Von Cranach and I. Vine (Eds.), *Social communication and movement.* New York: Academic Press, 1973.

Ekman, P. Universal and cultural differences in facial expressions of emotions. In J. K. Cole (Ed.), *Nebraska Symposium on Motivation* (Vol. 19). Lincoln: University of Nebraska Press, 1972.

Ekman, P. Cross cultural studies of facial expression. In P. Ekman (Ed.), *Darwin and facial expression.* New York: Academic Press, 1973.

Ekman, P., & Friesen, W. V. *Unmasking the face.* Englewood Cliffs, N.J.: Prentice-Hall, 1975.

Ekman, P., & Friesen, W. V. *The facial action coding system.* Palo Alto, Calif.: Consulting Psychologists' Press, 1978.

Ekman, P., Friesen, W. V., & Ellsworth, P. *Emotion in the human face.* New York: Pergamon Press, 1972.

Ekman, P., Friesen, W. V., & Tomkins, S. S. Facial affect scoring technique: A first validity study. *Semiotica,* 1971, *3,* 37–38.

Fraiberg, S. *Insights from the blind.* New York: Basic Books, 1977.

Grant, E. C. Human facial expression. *Man,* 1969, *4,* 525–536.

Hooff, J. A. R. A. M. van. Facial expressions in higher primates. *Symposia of the Zoological Society of London,* 1962, *8,* 97–125.

Hooff, J. A. R. A. M. van. A comparative approach to the phylogeny of laughter and smiling. In R. A. Hinde (Ed.), *Non-verbal communication.* Cambridge, England: Cambridge University Press, 1972.

Hooff, J. A. R. A. M. van. A structural analysis of the social behavior of a semi-captive group of chimpanzees. In M. Von Cranach and I. Vine (Eds.), *Social communication and movement.* New York: Academic Press, 1973.

Izard, C. *The face of emotion.* New York: Appleton-Century-Crofts, 1971.

Kendon, A., & Ferber, A. A description of some human greetings. In R. P. Michael and J. H. Crook (Eds.), *Comparative ecology and behavior of primates.* New York: Academic Press, 1973.

Leventhal, H., & Mace, W. The effects of laughter on evaluation of a slapstick movie. *Journal of Personality,* 1970, *38,* 16–30.

Mackey, W. C. Parameters of the smile as a social signal. *Journal of Genetic Psychology,* 1976, *129,* 125–130.

Spitz, R. A., & Wolf, K. The smiling response: A contribution to the ontogenesis of social relations. *Genetic Psychology Monographs,* 1946, *34,* 57–125.

Sroufe, L. A., & Waters, E. The ontogenesis of smiling and laughter: A perspective on the organization of development in infancy. *Psychological Review,* 1976, *83,* 173–189.

Tomkins, S. S. *Affect, imagery and consciousness: Vol. 1. The positive affects.* New York: Springer, 1962.

Tomkins, S. S., and McCarter, R. What and where are the primary affects? *Perceptual and Motor Skills,* 1964, *18,* 119–158.

Vinacke, W. E. The judgment of facial expressions by three national–racial groups in Hawaii: I. Caucasian faces. *Journal of Personality,* 1949, *17,* 407–429.

Vinacke, W. E., & Fong, R. W. The judgment of facial expressions by three national–racial groups in Hawaii: II. Oriental faces. *Journal of Social Psychology,* 1955, *41,* 184–195.

White, W. F. *Street corner society* (2nd ed.). Chicago: University of Chicago Press, 1955.

Received September 14, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 9, 1554–1560

The Effects of Mediator Rewards and Suggestions Upon Negotiations

James A. Wall, Jr.
School of Business and Public Administration
University of Missouri—Columbia

This study investigated the effects that two mediation techniques—rewarding negotiator concessions and suggesting concessions—have upon a negotiator's bargaining and outcomes. In the study subjects served as the negotiator and opposing negotiator while a confederate acted as the mediator in four separate wage negotiations. The results revealed that (a) mediator rewards result in larger negotiator total concessions, more negotiator–opponent agreements, and higher joint outcomes than do no rewards; (b) mediator suggestions produce larger initial concessions than do no suggestions; and (c) mediator rewards and suggestions interact, in that each technique is more potent in the absence of the other.

Over the years, mediators have developed and employed a myriad of mediation techniques to facilitate negotiation effectiveness. Such methods include establishing effective lines of communication, becoming a spokesman for the weaker side, creating a formal or relaxed negotiation atmosphere, pointing out or exaggerating the costs of no agreement, removing one party's capability to bluff, coordinating concession making, sharing responsibility for the agreement, suggesting and attributing prominence to alternative solutions, and staying out of the way. Even though mediation is noted for its variety, it suffers from a lack of theoretical analysis. Today, just as over a decade ago, "there is relatively little theoretical analysis of the mediation process and even fewer results and conclusions in consequence of such analysis" (Stevens, 1963). Since a prerequisite for such an analysis is a delineation of the mediation process, the following succinct explication is proffered. Subsequently, it will serve as the base for the generation of the hypotheses.

The Mediation Process

The mediation paradigm consists of at least three individuals (mediator, negotiator, and opposing negotiator) and three relationships (mediator–negotiator, mediator – opposing negotiator, and negotiator – opposing negotiator). Consequently, the nature and outcomes of the mediation process can be considered as resulting from the main and interactive influences of (a) the personal characteristics of the three participants, (b) the nature of the three interpersonal relationships, and (c) situational factors.

The negotiation literature documents well the effects of the two negotiators' personal characteristics and their interpersonal relationships (Chertkoff & Esser, 1976; Rubin & Brown, 1975; Yukl, Note 1), and it seems logical to conclude that these effects would retain potency in the mediator's presence. Similarly, the effects of situational factors upon negotiator – opposing negotiator bargaining are thoroughly limned (Druckman,

The author thanks Dennis D. Organ for his critical reading of an earlier draft of this manuscript and the editors and reviewers for their helpful suggestions.

This research was funded by the Indiana University School of Business and the Research Council of the Graduate School, University of Missouri—Columbia.

Requests for reprints should be sent to James A. Wall, Jr., Department of Management, School of Business and Public Administration, University of Missouri, Columbia, Missouri 65211.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3709-1554$00.75

1971), and one can logically assume that most of these effects hold strong in a mediation paradigm.

On the other hand, the effects of the mediator–negotiator and the mediator – opposing negotiator relationships, as well as the mediator's personal characteristics, have not as yet been investigated. And similarly, the interactions of these factors with the situational factors, the negotiator's traits, and his interpersonal relationships have not been studied. The present study focuses upon the mediator–negotiator relationship.

The Mediator–Negotiator Relationship

The relationship between the mediator and each of the negotiators can be considered one of exchange in which both the mediator and negotiator receive outcomes (or rewards) and make inputs (or incur costs). Usually the mediator's primary goal or outcome is agreement between the two negotiators (Pruitt, 1971; Stevens, 1963; Warren, 1954). In seeking this goal, the mediator can concentrate on the settlements and leave the means up to the negotiators. Such techniques include the following: suggesting better alternatives to the negotiators, selecting or defending one party's proposal, overestimating the costs of not settling or the benefits of settling, and making certain viable positions salient. On the other hand, the mediator can adopt a more indirect approach and concentrate upon modifying the negotiator's behavior, expectations, attitudes, perceptions, and so forth so as to conduce or facilitate agreement. Examples of these techniques include controlling the amount of internegotiator communication, advising one party as to the other's intentions, converting overt bargaining to tacit bargaining, suggesting concessions to one negotiator, and persuading one party to undo its commitments. As for the mediator's inputs or costs, they are manifold. The mediator operates under a strict set of norms; he must contribute expertise, objectivity, social approval (or disapproval), humor, and assurances. And occasionally he pays dearly in terms of time, sleep, energy, and relaxation.

Like the mediator, the negotiator experiences rewards and costs in the relationship. The rewards obtained from the mediator include the mediator's suggestions, approvals, payments, insights, and expertise as well as information about the opponent's intentions and forthcoming behavior. As for the negotiator's inputs or costs, they can include his promises, information as to how he will act, honesty and observation of norms vis-à-vis the mediator, and concessions to the opponent.

Because the negotiator wishes to obtain high net outcomes from his relationship with the mediator, his behavior is most likely to be altered by mediation techniques that raise his rewards and lower his costs. Two mediation techniques that do so are direct mediator rewards and mediator suggestions.

Mediator Rewards

A mediation technique that should prove effective in begetting negotiator concessions to an opponent entails rewarding the negotiator for his concessions—a technique successfully utilized by President Carter in his Mideast mediations (Butler, 1978). The predicted effectiveness of this strategy (which involves giving the negotiator a large reward when he makes a large concession to his opponent, a small reward when he makes a small concession, and no reward when he makes a negative concession or no concession) is based upon the reinforcement principle that rewarded behavior is more likely to occur at a later time than is unrewarded behavior. It is also underpinned by the finding (Wall, 1977) that an opposing negotiator, by using his own concessions as rewards, can increase the negotiator's concessions. In view of the effectiveness of rewards when they are used by the opposing negotiator, who is usually perceived as an adversary, it seems that rewards should be highly effective when applied by a mediator, who is usually perceived as a neutral party.

In sum, the mediator's rewards are expected to increase the negotiator's total concessions and consequently to result in more negotiator–opponent agreements and higher joint payoffs. Since the rewards follow the

negotiator's concessions, however, they are not expected to affect his initial concession.

Mediator Suggestions

A mediation technique that should more expeditiously lead to negotiator concessions is the mediator's suggestion of concessions, which provide the negotiator with a face-saving device whereby he can make a concession and move toward agreement without a costly loss of face (Pruitt & Johnson, 1970). If the mediator's suggestion precedes the negotiator's initial concession, it is expected to increase that concession. It is hypothesized, however, that the effect upon the negotiator's subsequent bargaining and its consequences is moderated by the presence or absence of mediator rewards: Whenever the mediator does not reward the negotiator's concessions, the mediator's suggestions retain potency throughout the negotiation because the negotiator retains his need of a face-saving device. That is, he continually requires an excuse or rationalization for his concession making, and the suggestions continually provide these. Therefore, when the mediator gives no rewards, the negotiators receiving mediator suggestions continue to make large concessions to the opponent. Consequently, they make larger total concessions, reach more agreements, and share in larger joint payoffs than do the negotiators not receiving the suggestions.

On the other hand, whenever the mediator does reward the negotiator's concessions, the suggestions (vs. no suggestions) have an insignificant effect upon the negotiator's concession making because they are of little value to the negotiator. He does not need to save face in this situation, because the mediator's rewards provide a rationalization for his concessions; he makes concessions because he receives rewards for doing so.

Hypotheses

In summary, it is predicted that mediator rewards and suggestions have the following effects:

1. Negotiators receiving mediator rewards for their concessions (a) make larger total concessions to their opponent, (b) reach more agreements with their opponent, and (c) share in larger joint negotiator–opponent payoffs than do negotiators not receiving such rewards.

2. Negotiators receiving suggested concessions from a mediator make larger initial concessions to their opponent than do negotiators not receiving such suggestions.

3. An interaction between the mediator's rewards and suggestions is manifested in the negotiator's subsequent bargaining and its outcomes. In the no reward condition, the negotiators receiving mediator suggestions (a) make larger total concessions to their opponent, (b) reach more agreements with their opponent, and (c) share in larger joint negotiator–opponent payoffs than do negotiators not receiving such suggestions. The differences in the reward condition are smaller than those in the no reward condition.

Method

Subjects

The 170 subjects who participated in the experiment were undergraduate male volunteers enrolled in behavioral science courses at Indiana University. For serving as a subject each received 2 hours' credit for experimental participation.

Procedure

Two subjects and a confederate were used in each session. Upon arrival, they were randomly seated in three chairs at a round table that was partitioned into three equal segments of partitions 2 feet high connected at the middle of the table. Each segment was labeled *union representative, mediator,* or *management representative*; the subjects were placed in the two representative positions and the confederate at that of the mediator. In the instructions the subjects were informed that they were going to act as a union or management representative in four negotiations and that the mediator was an MBA student majoring in labor relations who could communicate with them any time he wished. Attached to the union and management representative's instructions was his payoff matrix (and his alone) for the first negotiation. (The management and union payoff matrices for each of the negotiations were constructed so that a gain to one side resulted in an equal loss to the other. The positive payoff areas of

the two matrices overlapped, however, so that in each negotiation it was possible for the management and union representative to agree on a mutually profitable wage. As an example, for the negotiation in which the current wage was $6.00, a settlement at $6.30 yielded a $.20 positive payoff for the management and an $.18 loss for the union. For every $.01 that the wage increased, the union gained $.01 and the management lost $.01. Therefore, if the agreement were $6.48, the management would make $.02 and the union $.00; at $6.49 both would make $.01; and at $6.50 the management would make $.00 and the union $.02.) The instructions also told the representatives that they each had $.50 and that in the first negotiation the management representative would initiate the bargaining. He was to make the first offer on his bargaining sheet and slide it through a slot on his left to the mediator. The mediator, in all conditions, was to look at the offer, make sure that no messages were written on the sheet, and pass the sheet through a slot on his left to the union representative. The union representative, in turn, was to respond with a counteroffer and pass it back to the mediator, who was to return it to the management representative. The management representative then would make his second offer, and so forth. Both representatives were informed that they and their opponent would each be allowed to make four offers in the negotiation and that failure to agree would result in a fine of $.10 to each side. As soon as an agreement was reached or the eight offers were tendered, the mediator raised his hand and the subjects were given their instructions, payoff sheets, and a bargaining sheet for the second negotiation.

The second, third, and fourth negotiations were similar to the first. The management representative made the first offer and both sides were allowed to make four offers. At the conclusion of the fourth negotiation, the subjects were asked to complete brief questionnaires. As soon as they had done so, they were paid and debriefed.

Independent Variables

Previous mediation studies (e.g., Bartunek, Benton, & Keys, 1975; Johnson & Pruitt, 1972; Johnson & Tullar, 1972) have applied the mediation technique to *both* the negotiator and his opponent. Unfortunately, such a methodology provides confounded, equivocal data. The reported effects of their mediation techniques could have resulted from the negotiator's response to the technique, from the opponent's response, or from an interactive negotiator–opponent response. To provide more unequivocal data concerning the effects of the mediation techniques, this study applied each technique only to the negotiator (management representative). In all conditions, the mediator passed notes to the opponent (union representative) on which he had checked the statement, "I have no suggestions for this round."

Mediator rewards. In the reward condition, the mediator gave the negotiator (management representative) 50% of each concession he made to the opponent (i.e., the mediator gave the negotiator $.01 for every $.02 that he conceded to the opponent) and accompanied this reward with a handwritten message that read, for example, "Here is _____ for your last offer." (The blank was filled with the appropriate monetary amount, and the specific wording of this message was varied over the different rounds.) On the sheet passed to the negotiator, the mediator also checked the message, "I have no suggestions for this round."

In the no reward condition the mediator after each round passed the negotiator a note on which he had checked the aforementioned message without adding a handwritten message.

Mediator suggestions. In the suggestion condition, the mediator on his note suggested a predetermined offer—one that exceeded the mean offer made by negotiators in a pilot study—to the negotiator prior to each round. An example of the statements used is "I suggest that you offer about $5.44 this time."

Under the no suggestion condition the mediator in each round checked the no suggestion statement.

Dependent Variables

There were four primary dependent variables under study. The negotiator's initial and total concessions were expressed as the difference between the current wage ($4.00, $5.00, $6.00, or $7.00, with the sequence randomly selected for each negotiator) and his initial and final offer, respectively. The number of agreements in the four negotiations had a possible range of 0 to 4. And the joint negotiator–opponent payoff was the sum of the two bargainers' initial credits ($.50 each) and their payoffs for agreement minus their penalties for deadlocks ($.10 each). It did not include the mediator's rewards to the negotiator.

Design and Analysis

A $2 \times 2 \times 4$ repeated measures design with two levels of mediator reward (reward and no reward), two levels of mediator suggestions (suggestion and no suggestion), and four trials (negotiations) was used to study the negotiator concessions. The number of agreements and the joint payoff were analyzed via a 2×2 design (two levels of mediator reward and two levels of mediator suggestion).

Results

Main Effects

The mediator rewards affected all facets of the negotiators' bargaining and outcomes as predicted in Hypothesis 1. Table 1 reveals that rewarded negotiators made larger total concessions to their opponent than did those

Table 1
Negotiators' Concessions

Condition	Negotiation 1	2	3	4
Reward				
Suggestion				
Initial concession	.24	.22	.22	.25
Total concession	.45	.45	.45	.45
No suggestion				
Initial concession	.13	.19	.25	.23
Total concession	.47	.47	.48	.49
No reward				
Suggestion				
Initial concession	.22	.24	.23	.26
Total concession	.45	.45	.45	.47
No suggestion				
Initial concession	.15	.16	.17	.22
Total concession	.41	.41	.44	.40

not receiving rewards, $F(1, 81) = 3.91$, $p < .051$. Similarly, the rewarded negotiators reached more agreements with their opponent than did those who were not rewarded (Table 2), $F(1, 81) = 5.98$, $p < .02$, and received larger joint payoffs, $F(1, 81) = 8.52$, $p < .005$.

As for the effect of mediator suggestions, Hypothesis 2 was supported. Negotiators receiving suggestions made larger initial concessions than did those receiving no suggestions (Table 1), $F(1, 81) = 8.01$, $p < .006$.

Interactions

The mediator rewards and suggestions interacted as predicted (Hypothesis 3a) to affect the negotiators' total concessions, Table 1, $F(1, 81) = 5.92$, $p < .02$. However, the data reveal a phenomenon in addition to the one predicted. They show that as predicted, the mediator suggestions are slightly more effective in conducing negotiator concessions in the no reward condition. But they also indicate that the mediator rewards are more effective in the no suggestion condition. Thus the two techniques have analogous effects; each is more potent in the absence of the other. Corroborative but nonsignificant evidence for this conclusion is supplied by Table 2. Here it can be seen that the mediator suggestions were effective in facilitating agreements and enhancing joint payoffs only in the no reward condition. Similarly, the mediator rewards were more effective in the absence of mediator suggestions.

Within-Subject Effects

As can be noted in Table 1, the negotiators' initial concessions increased over the four negotiations, $F(3, 82) = 5.45$, $p < .002$; however, their total concessions did not exhibit this trend effect. In addition, the analysis of the negotiator's initial and total concessions revealed no significant Trend × Reward, Trend × Suggestion, or Trend × Reward × Suggestion effects.

Discussion

The principal conclusions to be drawn from this study are threefold. Mediator rewards are effective in facilitating negotiator concessions, negotiator–opponent agreements, and large joint payoffs. Second, these rewards are more effective in the absence of mediator suggestions. And finally, mediator suggestions lead to large initial concessions in both the reward and no reward conditions; however, their effectiveness abates in the reward condition so that their effects upon the total concessions are nonsignificant.

The interaction between the mediator's rewards and suggestions raises an interesting question: Why are the mediator's rewards more potent in the absence of his suggestions and his suggestions more effective in the ab-

Table 2
Negotiator–Opponent Agreements and Joint Payoffs

Condition	Number of agreements	Joint payoff
Reward		
Suggestion	2.7	.81
No suggestion	2.7	.81
No reward		
Suggestion	2.3	.70
No suggestion	1.8	.57

sence of rewards? It appears that mediator suggestions limit the effectiveness of his rewards because the suggestions place a ceiling upon the negotiator's concessions. That is, negotiators seldom will concede more than the mediator suggests. On the other hand, when no suggestions are proffered, such a ceiling is not created. The current data support this conclusion. Only six percent of the negotiators' concessions in the reward–suggestion condition were larger than the mediator's suggestions. However, 25% of the negotiators' concessions in the reward – no suggestion condition were larger than the mediator's suggestion on the corresponding round of the reward–suggestion condition. This difference is significant, $t(41) = 2.73$, $p < .01$.

With regard to the effects of the mediator's rewards upon the potency of his suggestions, the results support the argument that suggestions, in the absence of rewards, provide the negotiators with a face-saving device, allowing them to make concessions without suffering loss of face. However, when the mediator rewards the negotiators' concessions, the suggestions do not provide the negotiators with face because the rewards provide a rationalization for their concession making.

This latter proposition engenders a more thorough consideration of the role played by the mediator's rewards. The reward effects, previously reported, could be solely attributable to the change that the rewards made in the negotiator's net payoffs. Or they could be attributable to a combination of payoff modification and other influences (e.g., face-saving or salience of reward). A 2×2 factorial design with two levels of negotiator reward and two levels of payoff administration would indicate if the payoff modification was solely responsible. As in the present experiment, the reward treatments would consist of one condition (reward) in which rewards are given for concessions and a second (no reward) in which they are not. With regard to the payoff administration factor, in one condition (mediator administered) the mediator, as in the present study, would administer the reward or no reward, and in the other (matrix administered), the payoffs would be dictated entirely by the negotiator's payoff matrices. That is, under the reward – matrix administered condition, the rewards for conceding would be built into the negotiator's payoff matrices, whereas under the no reward – matrix administered condition, the matrices would contain no rewards. If the rewards (vs. no rewards) were found to have the same effect upon the negotiator's concessions in both payoff–administration conditions, one could conclude that the reward effects are attributable only to the change that the rewards make in the negotiator's net payoffs. If a larger effect resulted under the mediator-administered condition, however, one could conclude that other influences also underpin the mediator reward effects.

Since the present study applied the mediation techniques only to the negotiator, it provided an opportunity to investigate the effect that the techniques had, via the negotiator's concessions, upon those of the opposing negotiator. If the increased negotiator concessions, resulting from mediator rewards and suggestions, had generated reduced opponent concessions, then the effectiveness of the mediation would have been reduced. But if increased negotiator concessions had resulted in large reciprocal concessions or had produced no effect, their effectiveness would have been enhanced or unaffected. The analysis of the opposing negotiator's concessions revealed that neither the mediator rewards nor the suggestions altered the opponent's concession making (none of the Fs for the differences approached significance.) That is, the opponents did not reciprocate the negotiator's concessions, nor did they exploit them.

Finally, the reported results attest to the potency of both mediator rewards and suggestions when applied to only one negotiator. Since neither technique produced indirect effects in the other negotiator's bargaining, the effects of each technique would probably be enhanced by its concomitant application to the negotiator and his or her opponent. However, further research is required to determine if this is the case or if joint application would produce an interactive effect.

Reference Note

1. Yukl, G. A. *A review of laboratory research on two-party conflict negotiation*. Unpublished manuscript, Baruch College, 1975.

References

Bartunek, J. M., Benton, A. A., & Keys, C. B. Third party intervention and the bargaining behavior of group representatives. *Journal of Conflict Resolution*, 1975, *19*, 532–557.

Butler, D. Inside Camp David. *Newsweek*, October 2, 1978, 39–49.

Chertkoff, J. M., & Esser, J. K. A review of experiments in explicit bargaining. *Journal of Experimental Social Psychology*, 1976, *12*, 464–486.

Druckman, D. The influence of the situation in interparty conflict. *Journal of Conflict Resolution*, 1971, 523–554.

Johnson, D. F., & Pruitt, D. G. Preintervention effects of mediation versus arbitration. *Journal of Applied Psychology*, 1972, *56*, 1–10.

Johnson, D. F., & Tullar, W. L. Style of third party intervention, face-saving and bargaining behavior. *Journal of Experimental Social Psychology*, 1972, *8*, 319–330.

Pruitt, D. G. Indirect communication and the search for agreement in negotiation. *Journal of Applied Social Psychology*, 1971, *1*, 205–239.

Pruitt, D. G., & Johnson, D. F. Mediation as an aid to face saving in negotiation. *Journal of Personality and Social Psychology*, 1970, *14*, 239–246.

Rubin, J. Z., & Brown, B. R. *The social psychology of bargaining and negotiation*. New York: Academic Press, 1975.

Stevens, C. M. *Strategy and collective bargaining negotiations*. New York: McGraw-Hill, 1963.

Wall, J. A., Jr. Operantly conditioning a negotiator's concession making. *Journal of Experimental Social Psychology*, 1977, *13*, 431–440.

Warren, E. L. Mediation and fact finding. In Kornhauser, A., Dubin, R., & Ross, A. M. *Industrial conflict*. New York: McGraw-Hill, 1954.

Received September 19, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 9, 1561-1575

The Effects of Behavioral Intentions and Consequences on Judgments of the Actor and Other: An S-V-O Analysis

David A. Kravitz and Robert S. Wyer, Jr.
University of Illinois at Urbana-Champaign

An extension of Gollob's subject-verb-object model of social inference was used to investigate the effects of information about behavioral intentions and consequences on judgments of both an actor and the person toward whom the behavior is directed. Participants received one or more pieces of information about some or all of the following factors: an attribute of the actor, the actor's intentions to help or hinder the other, the actual consequences of this action (whether the other is helped or hindered), and an attribute of the other. Judgments of actors' admirableness increased with the favorableness of the adjectives describing them, the favorableness of both their intentions and the consequences of their actions, the justness of their intentions and of the consequences of their actions, and their ability to produce the consequences they intended. Judgments of the other's admirableness depended only on the adjectives describing the other and, when this adjective was not presented, on the consequences of the actor's behavior for the other. Behavioral consequences appeared to affect judgments of both the actor and the other independently of the actor's intentions. A second experiment demonstrated that the effects of information on judgments of the actor depend on the dimension of judgment in predictable ways and suggested that judgments of admirableness may be mediated by perceptions of both virtuousness and competence.

When a person's behavior has consequences for another, subsequent evaluations of both the actor and the other will often vary with the favorableness of these consequences (cf. Lerner, Miller, & Holmes, 1976; Lerner & Simmons, 1966; Walster, 1966). These evaluations are presumably based in part on the judge's assumptions about the implications of the consequences for characteristics of the persons being judged. Unfortunately, the nature of these assumptions and their possible effects have not always been identified in research performed to date. The assumptions may depend on whether the person being judged is the actor or the target of the actor's behavior and also on other available information about these persons that influences the interpretation of the behavioral consequences. Particularly important in this regard may be information about the actor's intentions and about more general personality traits of the parties involved.

The research reported in this paper investigated the nature of these contingencies and the possible assumptions underlying them. Both the actor and the target of his actions were judged on the basis of identical sets of information, and the separate contributions of different informational cues to these judgments were isolated. The theoretical and empirical issues surrounding these matters are discussed below, along with the methodology used to investigate them.

This research was performed as part of a M.A. thesis by the first author under the direction of the second and was supported in part by a University Fellowship in Psychology from the University of Illinois at Urbana-Champaign to the first author and in part by National Science Foundation Grant BNS 76-24001 to the second author.

Requests for reprints should be sent to Robert Wyer, Jr., Department of Psychology, University of Illinois, Champaign, Illinois 61820.

Judgments of the Actor

To the extent that the judge perceives an actor as responsible for the consequences of

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3709-1561$00.75

his/her action, the judge's evaluations of the actor may vary with the desirability of these consequences. However, as Fishbein and Ajzen (1975, pp. 207–213) point out, there are several different criteria for assigning responsibility. One criterion is simply whether the actors are instrumental in producing the consequences that occur, regardless of whether they could have foreseen these consequences. A second criterion is whether the actors intended the consequences of their actions. According to a third criterion, responsibility may be assigned only if the consequences of the actors' behavior are unjustified in terms of social norms or standards of appropriateness. Thus, for example, an actor who intentionally harms a person about to commit a crime would be assigned responsibility according to the first two criteria described above but not necessarily according to the third, since harming a criminal might be considered justified. Fishbein and Ajzen argue that the inconsistent effects of behavioral consequences on evaluations of the actor (cf. Shaver, 1970; Shaw & Skolnick, 1971; Walster, 1966, 1967) may be partially due to a failure to control for the criteria used by judges for assigning responsibility in the experimental conditions employed to investigate these effects.

The above comments suggest that an adequate investigation of the effects of behavioral consequences on judgments of the actor may require a consideration not only of the actor's intentions to produce these outcomes but also of the characteristics of the recipient of these outcomes. That is, if responsibility is assigned simply on the basis of whether the actor caused particular consequences, the consequence information should affect judgments of the actor independently of the actor's intentions, as Walster (1966) reports. However, to the extent that responsibility is assigned on the basis of intentionality, evaluations of the actor may be based on the perceptions of the actor's intentions, independently of the consequences that actually occur. Finally, if judges take into account the justifiability for the actor's behavior, their judgments may also depend on attributes of the recipient that influence their perceptions that the actor's intentions or the consequences of his/her behavior are justified. (Thus, for example, they may evaluate someone who either helps good people or harms bad ones more favorably than someone who either helps bad people or harms good ones.) Note that responsibility could be assigned on the basis of all three criteria simultaneously. To this extent, all three factors described above could contribute independently to judgments of the actor.

In investigating these possibilities, two points must be carefully considered. First, when information is presented about intentions, consequences, and recipient characteristics in combination, configural features other than those described above may emerge and influence judgments of the actor. For example, suppose that people who appear capable of producing the consequences they intend are seen as more competent than those who appear incapable of doing so and that the former are therefore more admired. To this extent, the consistency of the actor's intentions with the consequences of his/her actions may affect judgments in a way that could not be predicted from a consideration of either intentions or consequences in isolation.

Second, the above analysis points out that a given piece of information may contribute to judgments in several different ways, depending on the information accompanying it. Thus information that an actor intends to help another may contribute positively to judgments of the actor based on the assumption that people who intend to help others are admirable. In combination with information that the actor actually harmed the other, however, these helpful intentions might be interpreted as evidence that the actor is incompetent. In combination with information that the other person is a criminal, these intentions may be interpreted as unjustified. In the latter two cases, the actor's intentions would contribute negatively to judgments of him/her. Moreover, these various interpretations of intention information are not incompatible and thus could contribute simultaneously to judgments of the actor. A complete understanding of the influence of behavioral intentions and consequences on judgments of the actor re-

quires that the relative contributions of these factors be isolated and that the manner in which their contributions combine be determined.

Judgments of the Recipient of Behavioral Consequences

At least two theoretical perspectives suggest that the consequences of an actor's behavior may influence judgments of the person who experiences these consequences. Lerner and Simmons (1966) argue that people are motivated to maintain a belief that the world is just and therefore that people deserve the outcomes they receive. As a result, people who receive bad outcomes are assumed to be bad, whereas people who receive good outcomes are considered good. Similar predictions can be derived from the social equity formulation outlined by Walster, Berscheid, and Walster (1973). These considerations suggest that people may be evaluated on the basis of the consequences that befall them, regardless of why these consequences occur.

Judgments may also be affected by the actor's intentions toward the recipient. That is, judges may assume that people whom others intend to help have more favorable attributes than people whom others intend to harm. Moreover, attributes of the actor may also play a role in these judgments. According to balance theory (Heider, 1958), people should be evaluated more favorably (+) if either a highly regarded (+) person responds positively (+) to them or a disliked (−) person responds negatively (−) to them than if a highly regarded person responds negatively or a disliked person responds positively. Thus, the effect of an actor's intentions toward another person on evaluations of the other may be contingent on the favorableness of attributes of the actor as well as on the favorableness of the actor's intentions per se.

The various factors described above may contribute to judgments simultaneously with other aspects of the information presented (e.g., general personality attributes of the person being judged). On the other hand, it is conceivable that judges do not attach equal weight to these factors, but rather base their judgments only on those features of the presented information that they consider to be most reliable. For example, when direct information about the attributes of a person is available, judges may base their evaluations of this person primarily on these attributes while disregarding the less direct implications of others' behavior and intentions toward the person. In this case, these latter implications may have an impact only when direct information about the person's attributes is unavailable. The present study explored the possibilities discussed above.

Description of the Approach

The approach taken to investigate the issues outlined above was stimulated by the subject-verb-object (S-V-O) formulation of social cognition developed by Gollob (1974b). This formulation, which has been applied successfully to a variety of problems in impression formation and attribution (Gollob, 1974a, 1974b; Wyer, 1974, 1975; Wyer, Henninger, & Wolfson, 1975; Wyer & Hinkle, 1976) is described in detail elsewhere (Gollob, 1974b, 1979; Wyer & Carlston, 1979) and so will not be elaborated here. According to this approach, the information in a social interaction situation is analyzed by the judge into a series of *informational cues,* each cue defined in terms of values along one or more dimensions. The use of a given cue is hypothetically based on a specifiable assumption that the judge makes about its implications. Judgments can theoretically be described as a function of the independent contributions of these cues (and thus of the assumptions underlying their use), and the magnitudes of these contributions can be measured and tested statistically. The approach was originally applied to the effects of information along only three dimensions, pertaining to a trait of an actor, the actor's behavior toward another, and a trait of the other. However, it can easily be extended to any number and type of dimensions (cf. Wyer, Henninger, & Wolfson, 1975; Wyer & Hinkle, 1976). Such an extension was made in the present research.

To apply the formulation to the specific conditions of concern, consider a prototypic situation in which an actor engages in behavior that either helps or hinders another person. In such a situation, information relevant to evaluations of the actor and the other might vary along at least four dimensions, pertaining to (a) an attribute of the actor, (b) the actor's intentions to help or hinder the other, (c) the actual consequences of the actor's behavior (the other is helped or hindered), and (d) an attribute of the other. These dimensions will be denoted S, I, C, and O, respectively. If the values along each of the four dimensions are dichotomous (positive vs. negative), 16 configurations of stimulus information can be constructed, each representing a different combination of values along these dimensions.

In making a judgment on the basis of information such as that described above, judges may consider the implications of a number of different, potentially overlapping subsets of values along the dimensions involved. Which subsets they consider, and the effects of these subsets on their judgments, depends on the assumptions they make about their implications and the importance they attach to these implications. Each subset of dimension values, if considered as a basis for making a judgment, functions as a single unit of information, or *informational cue*. It is important to note that not all possible configurations of dimension values are meaningful cues: Only those configurations that have implications for the judgment when interpreted as a single unit function in this manner. Although it is often difficult to predict which of several possible cues a judge will actually use, the set of potentially useful cues can typically be circumscribed a priori.

Judgments of the actor. Based on the theoretical considerations raised above, previous research findings, and intuition, seven informational cues (referred to by Gollob as *biases*) were expected to influence judgments of the actor along a dimension of admirableness. (Admirableness rather than likableness was initially used because it seemed more directly relevant to considerations of the role of justice. However, possible differences between admirableness and other evaluative judgments are considered later in this paper.)

1. *S-bias.* Judges may assume that people with favorable attributes are more admirable than people with unfavorable ones. If this is so, actors should be judged as more admirable when the adjective describing them is favorable than when it is unfavorable.

2. *I-bias.* Judges may assume that people with good intentions (i.e., intentions to help) are more admirable than people with bad intentions (i.e., intentions to hinder). If so, their evaluations of the actor should be more positive when the actor's intentions are good than when they are not, independently of other considerations.

3. *C-bias.* Judges may assume that people who actually benefit others, regardless of their intentions to do so, are more admirable than those who hinder others. If judges make this assumption, their evaluations of the actor should increase with the favorableness of the consequences of the actor's behavior.

4. *O-bias.* Judges may assume that "birds of a feather flock together" and therefore may infer that actors are more admirable when their associates (i.e., the other) have favorable attributes than when these associates have unfavorable attributes. Although this cue is not relevant to the main concerns of this paper, it has been found to make a small but reliable contribution to judgments in previous studies (Gollob, 1968; Wyer, 1975; Wyer & Hinkle, 1976).

5. *IO-bias.* Judges may assume that people whose intentions are just are more deserving of admiration than people whose intentions are unjust. To this extent, judges should rate actors who intend either to help a good person or to hinder a bad person as more admirable than actors who intend to help a bad person or to hinder a good one.

6. *CO-bias.* People who produce outcomes that confirm that judge's perceptions of a just world may be considered more admirable than people who produce outcomes that threaten this perception. To this extent, judgments of actors who actually help a good person or hinder a bad person should be more favorable

than judgments of actors who actually help a bad person or hinder a good one.

7. *IC-bias*. Judges may assume that people who are capable of producing the outcomes they intend are more competent, and thus more admirable, than people who are incapable of doing so. To this extent, actors who actually help someone they intend to help or hinder someone they intend to hinder should be more admired than actors who help someone they intend to hinder or hinder someone they intend to help.

Note that the contributions of I-bias, C-bias, IO-bias, and CO-bias reflect the hypothetical effects of the different criteria for assigning responsibility noted earlier, whereas the contribution of IC-bias reflects the effect of an emergent cue that results when intention and consequence information is presented in combination. S-V-O methodology permits these contributions to be isolated and independently assessed, along with the contributions of more general personality characteristics of the actor and other (S-bias and O-bias).

Judgments of the other. Five informational cues seemed likely on a priori grounds to affect judgments of the person toward whom the actor's behavior was directed. The use of some of these cues is presumably based on assumptions similar to those postulated to underlie judgments of the actor, while the use of other cues is based on quite different assumptions.

1 and 2. *O-bias* and *S-bias*. As suggested above, people with favorable attributes may be seen as more admirable than those with unfavorable attributes, and people who associate with desirable others may be seen as more admirable than those who associate with undesirable others. To this extent, judgments of the other's admirableness should increase with the favorableness of both the adjective describing the other and the adjective describing the actor.

3. *I-bias*. Judges may assume that people generally try to benefit those who are admirable but try to harm those who are not admirable. To this extent, their judgments of the other's admirableness should be greater when the actor intends to help the other than when the actor intends to hinder him/her.

4. *C-bias*. People who believe that the world is just may judge someone who receives good outcomes more favorably than someone who is a victim of misfortune (Lerner & Simmons, 1966). If so, the other should be judged as more admirable if she/he is helped than if she/he is hindered, independently of other considerations.

5. *SI-bias*. Judges may assume that people are more likely to be admirable if good persons intend to help them or bad persons intend to hinder them than if good persons intend to hinder them or bad persons intend to help them. If this assumption is made, judgments of the other's admirableness should be greater in the first two cases than in the second two.

Three things may be noted from the above analysis. First, the contribution of each informational cue described above is reflected by the magnitude of one of the 15 orthogonal contrasts (main effects and interactions) in a standard analysis of variance on judgments as a function of the four informational variables. (For example, the contribution of S-bias is inferred from the main effect of the information given about the actor. Similarly, the contribution of IC-bias is inferred from the magnitude of the contrast corresponding to the interaction of intentions and consequences.) The procedures for calculating these contrasts will be described presently. Second, a single piece of information (for example, the actor's intentions) may be involved in several different informational cues, and thus may contribute to judgments in several different ways. Third, the use of a cue in judging an actor may be based on different assumptions than is the use of the same cue in judging the other.

The studies reported in this paper bear on these possibilities. The first experiment investigated the contributions of informational cues to judgments of the admirableness of both an actor and the other, and the extent to which the contributions of these cues depended upon the availability of other information. The second experiment investigated the effects of similar information on judg-

ments of the actor with respect to two other characteristics, competence and virtuousness (goodness), that were assumed to mediate judgments of admirableness. These data provided further clarification of the cues defined above and the assumptions underlying their use.

Experiment 1

Method

Overview

Forty-eight introductory psychology students judged the admirableness of two persons on the basis of information that varied along four dimensions concerned with (a) an attribute of the actor, (b) an attribute of the other, (c) the actor's intentions to help or hinder the other, and (d) the actual consequences of the actor's behavior for the other (helped vs. hindered). Each type of information was varied independently in a 2^4 design with two different stimulus replications. In addition to the 16 stimuli comprising the complete design, subjects made judgments of stimuli composing all possible subsets of the information dimensions given above. Data were analyzed using S-V-O methodology in a manner described below.

Subjects

Fifty-one undergraduates participated in the study to fulfill a course requirement in introductory psychology. Of these, 3 who failed to provide complete data were subsequently dropped from the analyses. Participants were run in groups of from 8 to 12 in sessions of approximately 1 hour. Subjects were verbally debriefed upon completion of the questionnaire.

Construction of the Questionnaire

Each stimulus in the main section of the questionnaire consisted of from one to four sentences describing an interaction between two people. These stimuli were constructed by factorially varying information along four dimensions concerning an attribute of the actor (S), the actor's intentions (I), the actual consequences (C) of the actor's behavior for the other, and an attribute of the other (O). There were two levels of each factor.[1]

To ensure that the results obtained were not peculiar to a given set of stimulus materials, these materials were constructed at two levels of abstractness. Moreover, different trait descriptions were used to describe the actor and the other in the two sets. Each set of materials was administered to a different group of subjects. The *concrete* materials concerned a specific business situation in which the actor and other were described as either tolerant or intolerant, and the actor's intentions (or consequences) dealt with having the other transferred to another department of the corporation for which they both worked—a transfer that the other either did or did not desire. Thus, for example, the stimulus for the condition in which values along S and I were both positive and values along C and O were both negative was of the form: "A is a tolerant person. A intends to have B transferred to another department, where B wants to work. As a result of A's actions, B is transferred to department where B does *not* want to work. B is an intolerant person."

The *abstract* materials contained information about a less clearly specified situation in which the actor's intentions and the consequences of his/her behavior were described simply in terms of helping and hindering. Moreover, the attributes describing the actor and the other were either honest or dishonest rather than tolerant or intolerant as in the concrete materials. Thus, the stimulus representing the same condition as in the preceding example was of the form: "A is an honest person. A intends to help B. B is hindered by A. B is a dishonest person."

In all, 80 sets of information were presented in each stimulus replication. Of these stimuli, 16 contained information concerning all possible combinations of values along the four dimensions. These stimuli comprised the complete (SICO) stimulus design. The remaining 64 stimuli formed all possible subdesigns, each pertaining to a different subset of information dimensions. Specifically, subjects were presented with the eight stimuli forming each of the SIC, SIO, SCO, and ICO subdesigns, the four stimuli forming each of the SI, SC, SO, IC, IO, and CO subdesigns, and the two stimuli comprising each of the S, I, C, and O subdesigns, where the letter(s) symbolizing each design pertain to the dimension(s) along which information was varied.

Stimuli were presented in three subsections of the questionnaire. One subsection contained those stimuli with information about both I and C (i.e., those comprising the SICO, SIC, ICO, and IC designs); a second subsection contained stimuli with information about C but not I (those comprising the SCO, SC, CO, and C designs); and the third subsection contained stimuli with information about I but not

---

[1] The information presented (with the exception of "help" and "hinder") was chosen on the basis of a pretest. Twenty-three subjects were presented with a number of personality traits, intentions, and consequences, and were asked to rate "how good each individual trait, intention or consequence is." The stimulus pairs used were chosen so that the negative member was as far below the neutral point as the positive member was above it. The intention and consequence information used in the abstract stimulus materials (help and hinder) was chosen to relate to previous work in the area (cf. Gollob, 1974a; Wyer, 1974).

C (those comprising the SIO, SI, IO, and I designs). The remaining stimuli, comprising the SO, S, and O designs, were randomly distributed within each of the three subsections. Stimuli were ordered randomly within each subsection except that the stimuli expected to produce the most extreme judgments were presented first in an attempt to anchor the response scale. These subsections were presented in all six possible orders to control for any demand characteristics that might be engendered by the presence or absence of specific information.[2]

After receiving each stimulus, subjects estimated how much they admired both the actor and the other along 11-point scales that ranged from 0 (not at all) to 10 (extremely much). The order of these estimates was counterbalanced across subjects.

Thus 84 stimuli were contained in the main section of the questionnaire (80 unique stimuli and 4 repetitions; see Footnote 2), with each stimulus requiring two responses.

Results

Judgments of the Actor

Data pertaining to the complete design and to each of the 14 subdesigns were analyzed separately. In each case, analysis of variance was first performed on judgments, collapsed over order of stimulus presentation and order of response, as a function of the four information dimensions presented (S, I, C, and O) and the type of stimulus materials (abstract vs. concrete). The main effect of stimulus materials was not significant in any of the 15 analyses, and few effects of informational variables were significantly contingent on scenario. The only exception to this concerned the interaction of stimulus materials with information about the actor (i.e., the information along S), which was significant in all 8 designs that included this information. These interactions indicated that the effect of the trait assigned to the actor in the abstract materials (honest vs. dishonest) was greater than the effect of the trait assigned to him/her in the concrete materials (tolerant vs. intolerant). This contingency simply reflects the difference in the implications of these traits for admiration and has no theoretical relevance to the issues of central concern in this paper. The few other interactions involving stimulus materials had no obvious psychological meaning and were interpreted as spurious.[3] Subsequent discussion is therefore restricted to data pooled over the two stimulus conditions.

An analysis of judgments as a function of the four informational variables involved in the complete (SICO) design yields 15 possible main effects and interactions. Of these, 7 (the 4 main effects and the two-way interactions between intentions, consequences, and the attribute describing the other) correspond to the information cues expected on a priori grounds to contribute significantly to judgments of the actor (S-bias, I-bias, C-bias, O-bias, IC-bias, IO-bias, and CO-bias). To calculate the orthogonal contrasts assumed to indicate the contributions of these cues, the favorable pole of each informational dimension was assigned a positive (p) valence and the unfavorable pole was assigned a negative (n) valence. The contrast corresponding to a particular effect was then computed by subtracting (a), the mean judgment of stimuli in which the product of the relevant valences was negative, from (b), the mean judgment of stimuli in which the product of the relevant valences was positive. (Thus, for example, the contrast corresponding to the interaction of I and O, which indicates the contribution of IO-bias, was inferred from the difference between the mean judgment based on stimuli in which the valences of I and O were both p or

---

[2] To investigate the possible effects of presentation order, four additional stimuli comprising one of the two-factor designs that appeared in the first subsection of the questionnaire were repeated in the third subsection. The specific two-factor designs used were varied over subjects. Comparisons of these two sets of judgments revealed no significant differences, suggesting that order was not an important factor in this experiment.

[3] In only three designs did stimulus materials interact with any factor other than S. In these cases, the two sets of stimulus materials were analyzed separately. In only one instance did these separate analyses result in a significant effect in the opposite direction from that of the combined data; specifically, the effect of O was negative in the ICO design when the abstract stimulus materials were used. In all other cases, the effect obtained using abstract materials was greater than that obtained using the concrete stimulus materials, but both effects were in the same direction. These interactions included CO in the CO and ICO designs and IO in the IO design.

Table 1
Contributions of Predicted Informational Cues to Judgments of the Actor

Item	Informational cue						
	S-bias	I-bias	C-bias	O-bias	IC-bias	IO-bias	CO-bias
Complete design							
SICO	2.66**	1.73**	1.10**	.14	.38*	.34*	.35**
Subdesigns							
SIC	2.78**	1.96**	1.26**		.49**		
SIO	2.99**	2.69**		.05		.38**	
SCO	3.61**		2.02**	.11			.24*
ICO		2.76**	.86**	−.26	1.18**	.09	.51**
SI	3.02**	3.00**					
SC	3.49**		2.11**				
SO	5.06**			.29			
IC		3.11**	1.24**		.74**		
IO		2.78**		−.26		1.20**	
CO			2.67**	−.15			1.19*
S	4.92**						
I		4.06**					
C			3.04**				
O				.02			

Note. S = an attribute of the actor. I = the actor's intentions to help or hinder the other. C = the actual consequences of the actor's behavior. O = an attribute of the other.
*$F(1, 42) > 4.08$, $p < .05$. **$F(1, 42) > 7.31$, $p < .01$.

n and the mean judgment based on stimuli in which one of these valences was *p* and the other was *n*.) These contrasts are shown in Table 1 for the complete design. Similar contrasts, based upon analyses of each subdesign, are also presented. (Statistical significance in each case is based on a separate analysis of data for each subdesign.)

According to the assumptions hypothesized to underlie the use of each cue and the notational system used to compute its contribution, the contrast associated with each cue should theoretically be positive. In fact, all contrasts but that corresponding to O-bias were significant and positive in the complete design. Moreover, the six cues that contributed significantly to judgments in the complete design contributed significantly to judgments in the subdesigns in which they were presented in 29 of 30 cases. The nonsignificant contribution of O-bias in both the complete design and subdesigns is not too surprising, since the magnitude of this cue was quite small in previous studies (Gollob, 1968; Wyer & Hinkle, 1976). Overall, the 7 contrasts described in Table 1 and interpretable on a priori grounds as reflecting the contributions of informational cues accounted for 97% of the item set variance in the complete design, whereas the 8 interactions that were *not* predicted to contribute to judgments accounted for only 3% of the variance. Moreover, these latter interactions were significant in only 4 of the 13 relevant subdesigns (31%), as opposed to 29 of 37 instances (78%) in the case of the seven predicted effects. None of the significant unpredicted effects qualifies our conclusions concerning the predicted effects of informational cues; in all cases, the contributions of these cues remained in the expected direction, varying only in magnitude.

Nevertheless, the nature of the few unexpected effects may be worth noting briefly. The only unexpected interaction that approached significance in the complete design involved S, I, and C; $F(1, 47) = 8.63$, $p < .01$. This is less than the Bonferroni critical value of 9.73 for 15 comparisons and the .05 significance level (Myers, 1972). To the extent that this interaction is taken seriously, it may best be interpreted as an indication that the contribution of IC-bias is greater when the actor is described by a favorable attribute

(M contrast = .64) than when the actor is described by an unfavorable attribute (M contrast = .12). The use of IC-bias presumably reflects the assumption that people are more admirable if they produce outcomes they intend than if they do not. The possibility that this assumption is made only when the actor has other favorable qualities has some intuitive appeal; if the actor is described by an unfavorable attribute, this may be sufficient to conclude that she/he is not admirable, and judges may thus attend less to other cues. However, since this interaction did not even approach significance in the SIC subdesign, the above interpretation should be treated with caution pending replication of the effect.

Three of the four unpredicted significant effects in the subdesigns pertained to the interaction of S and I. The interpretation of this interaction, which was not significant in the complete design, is unclear. It could reflect a tendency for judges to assume that people are more admirable if their intentions are evaluatively consistent with their general personality attributes (both favorable or both unfavorable) than if they are evaluatively inconsistent. Alternatively, it may also indicate that judges pay more attention to an actor's intentions when the actor is described by a favorable adjective than when he is described by an unfavorable one, and so I-bias has more effect in the former condition than in the latter. (This interpretation is consistent with our tentative interpretation of the SIC interaction.) Finally, the interaction could be due to a floor effect produced by the unavailability of sufficiently extreme responses (on the response scale) when both the adjective describing the actor and his intentions were unfavorable. If this is the case, however, it occurred despite our attempt to stabilize the response scale by presenting the stimuli calling for the most extreme responses first.

Summary. The results described above provide support for several general hypotheses raised in the introduction. First, they suggest that responsibility for behavioral consequences may be assigned according to several different criteria, and that each contributes to evaluations of the actor. That is, responsibility may be assigned simply on the basis of the actor's having committed the act that produced these consequences (leading C-bias to contribute to judgments), on the basis of the actor's intentions (leading I-bias to contribute), and on the basis of the justifiability of these intentions and consequences (leading IO-bias and CO-bias to influence the judgments). In addition, judgments of the actor may be based on other factors, including more general dispositional characteristics (as reflected in the contribution of S-bias) and the ability to produce the outcomes she/he intends (as evidenced by the contribution of IC-bias). It should be noted, however, that while the contributions of these cues may be isolated within any given design in which they occur, this does not imply that the contributions of some cues are unaffected by the availability of others. We shall return to this possibility presently.

Judgments of the Other

Judgments of the other were analyzed in the same manner as were judgments of the actor. Once again, no main effects of stimulus materials were significant, allowing us to pool data over this factor.[4]

Five main effects and interactions, corresponding to the informational cues denoted S-bias, I-bias, C-bias, O-bias, and SI-bias, were postulated on a priori grounds to be significant. The contrasts corresponding to these effects, calculated in the manner described previously, are shown in Table 2 for both the complete design and the subdesigns to which they are relevant. While all contrasts were positive as predicted, only that corresponding to O-bias was significant in the complete design. This suggests that when direct information about attributes of the other is avail-

[4] In only three designs did stimulus materials interact with any factor other than O. In these cases, the two sets of stimulus materials were analyzed separately. In only one instance did these separate analyses yield an effect that was significantly different in direction from those reported in Table 2: The effect of S was in reverse direction in the SIC design with the concrete stimulus materials.

Table 2
Contributions of Predicted Informational Cues to Judgments of the Other

Item	Informational cue				
	S-bias	I-bias	C-bias	O-bias	SI-bias
Complete design					
SICO	.12	.06	.14	3.78**	.08
Subdesigns					
SIC	.06	−.03	.43**		.07
SIO	−.03	.13		3.77**	.19
SCO	.08		.18	3.70**	
ICO		.15	.32**	3.80**	
SI	.29	.29*			.21
SC	.00		.27		
SO	.35			4.96**	
IC		.43**	.80**		
IO		.18		3.99**	
CO			.36	3.91**	
S	.06				
I		.38			
C			.15		
O				5.19**	

Note. S = an attribute of the actor. I = the actor's intentions to help or hinder the other. C = the actual consequences of the actor's behavior. O = an attribute of the other.
* $F(1, 42) > 4.08$, $p < .05$. ** $F(1, 42) > 7.31$, $p < .01$.

able, only this information is used as a basis for judging the person's admirableness. In this case information about other people's intentions and reactions to the person has relatively little influence.

When trait information about the other is not available, however, the actor's intentions and reactions toward him/her appear to have more effect. Specifically, I-bias, the use of which is presumably based on the assumption that people are more apt to be admirable if someone intends to help them than if someone intends to hinder them, contributed significantly to judgments of the other in two of the four subdesigns in which trait information along O was not presented, and was nonsignificantly greater in these four subdesigns (M contrast = 0.27) than in the four designs in which information along O was available (M contrast = 0.13), $F(1, 47) = 1.74$; $p > .10$.[5] Similarly, the contribution of C-bias, presumably based on the assumption that good things occur to good people and bad things occur to bad people, contributed significantly to judgments in three subdesigns, two of which were also those in which trait information about the other was not presented. The mean contribution of C-bias when information along O was not available was nonsignificantly greater (M contrast = 0.41) than its contribution when the information along O was also presented (M contrast = 0.25), $F(1, 47) < 1$ (see Footnote 5). Although these differences are not reliable, their consistency is sufficient to suggest that judges are most likely to infer the admirableness of people on the basis of others' intentions toward them and the consequences of others' behavior toward them when direct information about the attributes of these people is not available.

Although it was statistically significant in only three of eight cases, the contribution of C-bias to judgments of the other was consistently positive in all eight designs in which information along C was presented. (The probability of this result, assuming the strict null hypothesis, is approximately .004.) The contribution of this cue supports Lerner and Simmons's (1966) hypothesis that people infer the characteristics of others from the outcomes the others receive in a way that maintains their image of a just world. In this regard, it is also worth noting that intention information was also presented in all cases in which the contribution of C-bias was significant. Thus, the contribution of consequences information to judgments of admirableness of the other cannot be attributed solely to its mediating effects on beliefs about the actor's intentions to produce these consequences. Rather, consequences information has an effect over and above that of intentions. We shall return to this matter presently.

Of the 10 interactions that were not interpreted a priori as reflecting the contribution of informational cues, only 2 (SICO and CO) were significant in the analysis of the com-

[5] This analysis was performed on the mean contributions of the four relevant designs, with subjects being the replication factor.

plete design. Moreover, the proportion of item set variance accounted for by these 10 interactions was only 2%, as compared to 98% for the 5 effects that were predicted. Finally, the unpredicted interactions were significant in only 1 of 19 instances in which the relevant factors were manipulated in the various subdesigns. None of the 3 unexpected effects had any obvious psychological meaning, and none represented a significant reversal in the direction of the contributions of any informational cue described in Table 2.

Combinatorial Processes

Judgments of the actor, unlike judgments of the recipient, appeared to be based upon several different informational cues. A question therefore arises concerning the manner in which the implications of these cues are combined to influence the judgments. There are at least three obvious possibilities. First, as noted previously, judges may use only a limited number of informational cues as a basis for their responses (cf. Wyer, 1974; Wyer et al., 1975). If this is so, the likelihood that any given cue is used may decrease with the total number of cues available. The other possibilities are suggested by the research on algebraic models of information integration (cf. Anderson, 1971, 1974). That is, judges may first construe the implications of each informational cue separately and then either average or sum these implications to arrive at an overall judgment. If the implications of a given informational cue remain invariant over the stimulus configurations in which the cue is contained, an averaging model also implies that the contribution of this cue will decrease with the number of cues available (since the relative weight given to it will be less). On the other hand, a summative model implies that the contribution of each cue will not be affected by the number of other cues present.

To explore these possibilities, the mean contributions of the informational cues that contributed significantly to judgments of the actor in the complete design were plotted as a function of the number of factors comprising the design. (While the number of dimensions along which information is presented is not a perfect indicator of the number of informational cues available, this criterion was used in order to eliminate the need for a priori assumptions concerning the nature of these cues and whether they would contribute significantly.) These data are shown in Figure 1. In each case, the contribution of the cue decreased in magnitude with the number of informational dimensions comprising the design. This conclusion is supported statistically by the results of linear trend analyses of the data in each panel of Figure 1.[6] These trends were significant in analyses of S, $F(1, 141) = 70.64$, $p < .01$; I, $F(1, 141) = 89.27$, $p < .01$; C, $F(1, 141) = 54.23$, $p < .01$; IO, $F(1, 94) = 9.92$, $p < .01$; and CO, $F(1, 94) = 15.35$, $p < .01$ The linear trend for IC was nearly significant, $F(1, 95) = 3.15$, $.05 < p < .10$. These data rule out a summative process of information integration. However, the other two alternative possibilities raised above

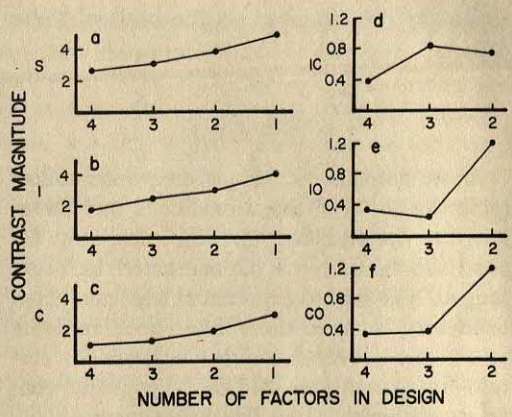


Figure 1. Mean contributions of (a) S-bias, (b) I-bias, (c) C-bias, (d) IC-bias, (e) IO-bias, and (f) CO-bias to judgments of the actor as a function of the number of factors in the design.

[6] These analyses were performed on the mean contribution of the given cue in all of the designs of the same size (1, 2, 3, or 4 factors), with subjects being the unit of analysis. Thus, for example, the value for the contributions of S were based on a single design for Design Sizes 1 (S) and 4 (SICO) and on the mean of three designs for Design Sizes 2 (SI, SC, SO) and 3 (SIC, SIO, SCO).

cannot be distinguished on the basis of these data.[7]

Experiment 2

As we noted earlier, the assumptions underlying the use of informational cues, and therefore the contributions of these cues, may depend substantially on the characteristic being judged. The second experiment was conducted to identify some of these differences. In doing so, it also provided indirect support for certain interpretations of the significant contrasts obtained in the first experiment.

Admiration for people may be based on two general types of criteria, both of which were reflected in the informational cues that contributed to judgments in Experiment 1. On the one hand, a person may be admired because she/he is virtuous, that is, his/her attributes and behavior coincide with social standards of morality and justice. Thus, the contributions of I-bias and IO-bias to judgments of the actor presumably reflected assumptions that people who intend to help others are more admirable than people who intend to hinder others and that people whose intentions are just are more admirable than people whose intentions are unjust. On the other hand, a person may be admired because she/he is particularly competent or has exceptional artistic or athletic skills, quite apart from moral considerations. This would account for the contribution of IC-bias to judgments of the actor in the first experiment, which was presumably based on the assumption that persons who are capable of producing the outcomes they intend are more competent, and therefore more admirable, than are those who are incapable of generating intended outcomes. If this reasoning is correct, it has implications for the contributions of these informational cues to judgments of the actor along dimensions other than admirableness. Specifically, all three cues mentioned above should affect judgments of the actor's admirableness, as was found in Experiment 1. However, suppose contributions of I-bias and IO-bias to admirableness judgments are mediated by their influence on perceptions of the actor's virtuousness, whereas the contribution of IC-bias is mediated by its influence on perceptions of the actor's competence. In this case, I-bias and IO-bias but not IC-bias should contribute to judgments of the actor along a good–bad dimension (assumed to be an index of virtuousness), whereas IC-bias but not IO-bias or I-bias should contribute to judgments of the actor's competence.

Method

Two questionnaires were prepared, one pertaining to competence judgments and the other to general evaluations of the actor along a good–bad dimension. Each form was constructed using the same basic set of stimulus materials, which represented all eight combinations of bipolar values along three informational dimensions pertaining to an actor's intentions (to help vs. to hinder), behavioral consequences (helps vs. hinders), and a general description of the other (good vs. bad). Thus the information presented was comparable to that involved in the ICO subdesign of Experiment 1. This information was presented in items of the form: "S intends to help (hinder) O. O is helped (hindered) by S. O is a good (bad) person." Then, in the form pertaining to competence judgments, two items were constructed with each set of stimulus information. In one item subjects were asked to estimate how likely it was that S was competent, and in the second item they were asked to estimate how likely it was that S was incompetent. In the form pertaining to good–bad judgments, corresponding items asked whether S was likely to be a "good person" or a "bad person."[8]

---

[7] There is, in addition, a statistical problem with these data. They violate the homogeneity-of-treatment–difference–variance assumption (Huynh & Feldt, 1970). Although this does not affect the analyses previously reported, it does force us to further qualify the modeling results. Specifically, when this assumption is not satisfied, the contrasts are not uncorrelated in the population and thus there is no guarantee that the sample contrasts are the appropriate bias weights. The homogeneity-of-treatment–difference–variance test was run on the covariance matrix of the 15 contrasts of the SICO design. The contrasts pertained to judgments of the actor, and the data were pooled over both stimulus replications. For these data: $f = 104$, $d = .8967$, $W = .001$, and $-(df)(d)lnW = 290$ is compared to chi-square with 104 degrees of freedom. The correlations among the contrasts ranged from $-.499$ to $+.421$, and the mean was $-.002$.

[8] Subjects were asked about both poles of the dimension to provide a test of the "relevant bias hypothesis" proposed by Gollob (1974b). This test is beyond the scope of this report; for details, see Kravitz (1978).

In all cases, judgments were made along an 11-point scale, ranging from 0 (not at all likely) to 10 (extremely likely). The 16 items in each form were arranged randomly, except that those items expected to result in the most extreme responses were placed first in an attempt to anchor the response scale.

Fifty-one subjects who did not take part in the first experiment participated in this study to fulfill a course requirement in introductory psychology. Of these, 27 completed the form pertaining to competence judgments, and the other 24 completed the form pertaining to good–bad judgments.

Results

For each set of stimulus information, each subject's judgment of the actor with respect to the negative pole of the relevant response dimension (i.e., either "incompetent" or "bad") was subtracted from his/her judgment of the actor with respect to the positive pole ("competent" or "good"). Each difference score was then treated as a single index of perceived competence or goodness.[9] The two sets of difference scores were then analyzed separately as a function of the three informational dimensions.

The five main effects and interactions that were significant in the first experiment when admirableness judgments were made, corresponding to the contributions of I-bias, C-bias, IC-bias, IO-bias, and CO-bias, were of particular interest. The magnitudes of the contrasts corresponding to these effects, along with the other contrasts involved in the analyses, are shown in Table 3 for both sets of judgments. (In each case, contrasts were calculated in the manner described in Experiment 1.) These data show the pattern expected on the basis of the considerations noted above. That is, of the five informational cues found in Experiment 1 to affect admirableness judgments, only the cue expected to be particularly relevant to competence, IC-bias, contributed significantly to these judgments. However, this cue did not contribute to judgments of the actor along a good–bad dimension. In contrast, I-bias, C-bias, IO-bias, and CO-bias each contributed significantly to judgments of the actor along the good–bad dimension, but not to judgments of the actor's competence. Although the effects of C-bias and CO-bias were not explicitly predicted, their contributions are not incompatible with these predictions. They suggest that people whose actions have either desirable or just consequences are assumed to be better people, regardless of their intentions, than are those whose actions have undesirable or unjust consequences.

Table 3
Contrast Magnitudes for Judgments of the Actor (S) in Experiment 2

Contrast	Dimension of judgment	
	Competent–incompetent	Good–bad
I	.22	2.28**
C	.21	1.27**
IC	2.54**	.32
IO	−.01	.97**
CO	−.04	.74*

Note. S = an attribute of the actor. I = the actor's intentions to help or hinder the other. C = the actual consequences of the actor's behavior. O = an attribute of the other.
*$p < .05$, $F(1, 23) > 4.30$ for competent–incompetent and $F(1, 26) > 4.23$ for good–bad. **$p < .01$, $F(1, 23) > 7.95$ for competent–incompetent and $F(1, 26) > 7.72$ for good–bad.

In summary, the results of this experiment suggest that the effects of the informational cues on judgments of admirableness found in Experiment 1 were mediated by their implications for two different attributes, competence and virtuousness.

Discussion

In combination, the two experiments reported here provide substantial insight into the assumptions and factors that underlie the effects of behavioral consequences and intentions on judgments of the interacting parties. Moreover, they point out important differences between the factors that underlie the

[9] This approach seemed justified on a priori grounds for two reasons. First, it was expected that judgments pertaining to one pole of the dimension would correlate negatively with judgments pertaining to the other pole. Second, the use of both poles would seem to represent the dimension better than the use of a single pole. Supplementary analyses of judgments pertaining to each pole separately do not qualify the pattern of results reported (Table 3).

effects of behavioral consequences on judgments of the actor and the factors that underlie their effects on judgments of the person toward whom the actor's behavior is directed. In the first case, we speculated that judgments of the actor may be mediated by attributions of responsibility at several different levels, ranging from simply the commission of the act that produced the consequences to the actor's intentions to produce the consequences and the justifiability of these outcomes (as defined in terms of social standards of appropriateness). In fact, attributions of responsibility at all three levels appear to contribute to judgments of the actor. In addition, evaluations of the actor appear to be based upon perceptions of the actor's ability to produce the outcomes she/he intends, as well as more general dispositional characteristics implied by the adjectives used to describe him/her. In this regard, results of the second experiment suggest that judgments of the actor's admirableness may be mediated by perceptions of both virtuousness (goodness) and competence, and the contributions of these cues to admirableness ratings may be a result of their more fundamental influence on these mediating perceptions.

Whereas judgments of actors were based upon their intentions and the consequences of their behavior, judgments of recipients were significantly affected by the actor's intentions and behavioral consequences only when more direct information about the recipient's general personality (e.g., trait adjectives) was unavailable. This suggests that judges tend to ignore the implications of others' reactions to a person and the consequences of another's behavior toward this person when more direct indications of his/her personality are provided. Nevertheless, the consistently positive influence of consequences information on judgments of the other provides some support for the hypothesis that judges believe in a just world and therefore evaluate people on the basis of the favorableness of the consequences that befall them (Lerner & Simmons, 1966). Moreover, beliefs in a just world appear to enter into inferences in several additional ways. For example, actors as well as others were admired more when their behavior had good consequences than when it had bad consequences, independently of other considerations. In addition, actors were more admired when their behavior had just consequences (i.e., when they helped good people or hindered bad people) than when it had unjust consequences. This is not to say that behavioral intentions are ignored; the intentions to produce both good outcomes (I-bias) and "just" outcomes (IO-bias) also contributed significantly to judgments of the actor. However, the independent contribution of consequences information is noteworthy and suggests an extension of the role of just world considerations beyond that implied by the original formulation. That is, one's belief in a just world may not only lead one to evaluate others in accordance with the outcomes they receive (thus making one's own perceptions consistent with this belief) but may also lead one to evaluate people more favorably if their behavior has consequences that reinforce this belief.

The effects of consequences and intentions noted above are relevant to a more general issue raised earlier. That is, a given piece of information can contribute to judgments in several ways, depending on the interpretation of its implications. Moreover, these contributions may occur simultaneously. In the present case, for example, the actor's help or hindrance of the other may be considered independently of additional information presented and therefore may contribute to judgments in its own right. However, depending on the additional information available, it may also be interpreted as an indication that the actor behaved justly or unjustly or that she/he is more or less competent. Each of these interpretations has implications for judgments of the actor.

The fact that information can affect judgments in different ways emphasizes the value of S-V-O theory and methodology in conceptualizing these various effects. In this regard, it is worth noting that of the statistical contrasts postulated on a priori grounds to reflect the contributions of psychologically meaningful informational cues, six of seven contrasts contributed to judgments of the actor and two of five contrasts contributed to

judgments of the recipient. On the other hand, very few of the contrasts that were *not* clearly interpretable as informational cues contributed significantly to judgments of either the actor or the other, and these contributions were typically not consistent over the informational configurations (full design and subdesigns) in which the relevant information was presented. Thus, it should be apparent that an S-V-O formulation can serve as a heuristic framework for circumscribing the set of information factors that may influence judgments in a given situation, and for conceptualizing the possible assumptions underlying the effects of these factors.

References

Anderson, N. H. Integration theory and attitude change. *Psychological Review*, 1971, *78*, 171–206.

Anderson, N. H. Cognitive algebra: Integration theory applied to social attribution. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 7). New York: Academic Press, 1974.

Fishbein, M., & Ajzen, I. *Belief, attitude, intention and behavior: An introduction to theory and research*. Reading, Mass.: Addison-Wesley, 1975.

Gollob, H. F. Impression formation and word combination in sentences. *Journal of Personality and Social Psychology*, 1968, *10*, 341–353.

Gollob, H. F. Some tests of a social inference model. *Journal of Personality and Social Psychology*, 1974, *29*, 157–172. (a)

Gollob, H. F. The subject-verb-object approach to social cognition. *Psychological Review*, 1974, *81*, 286–321. (b)

Gollob, H. F. A reply to Norman H. Anderson's critique of the subject-verb-object approach to social cognition. *Journal of Personality and Social Psychology*, 1979, *37*, 931–949.

Heider, F. *The psychology of interpersonal relations*. New York: Wiley, 1958.

Huynh, H., & Feldt, L. S. Conditions under which mean square ratios in repeated measurements designs have exact F-distributions. *Journal of the American Statistical Association*, 1970, *65*, 1582–1589.

Kravitz, D. A. *A subject-verb-object approach to social inferences of actors and recipients*. Unpublished master's thesis, University of Illinois at Urbana-Champaign, 1978.

Lerner, M. J., Miller, D. T., & Holmes, J. H. Deserving and the emergence of forms of justice. In L. Berkowitz & E. Walster (Eds.), *Advances in experimental social psychology* (Vol. 9). New York: Academic Press, 1976.

Lerner, M. J., & Simmons, C. Observer's reaction to the "innocent victim": Compassion or rejection? *Journal of Personality and Social Psychology*, 1966, *4*, 203–210.

Myers, J. L. *Fundamentals of experimental design* (2nd ed.). Boston, Mass.: Allyn & Bacon, 1972.

Shaver, K. G. Defensive attribution: Effects of severity and relevance on the responsibility assigned for an accident. *Journal of Personality and Social Psychology*, 1970, *14*, 101–113.

Shaw, J. I., & Skolnik, P. Attribution of responsibility for a happy accident. *Journal of Personality and Social Psychology*, 1971, *18*, 380–383.

Walster, E. Assignment of responsibility for an accident. *Journal of Personality and Social Psychology*, 1966, *3*, 73–79.

Walster, E. "Second-guessing" important events. *Human Relations*, 1967, *20*, 239–250.

Walster, E., Berscheid, E., & Walster, G. New directions in equity research. *Journal of Personality and Social Psychology*, 1973, *25*, 151–176.

Wyer, R. S. *Cognitive organization and change: An information-processing approach*. Hillsdale, N.J.: Erlbaum, 1974.

Wyer, R. S. Some informational determinants of one's own liking for a person and beliefs that others will like this person. *Journal of Personality and Social Psychology*, 1975, *31*, 1041–1053.

Wyer, R. S., & Carlston, D. *Social cognition, inference and attribution*. Hillsdale, N.J.: Erlbaum, 1979.

Wyer, R. S., Henninger, M., & Wolfson, M. Informational determinants of females' self-attributions and observers' judgments of them in an achievement situation. *Journal of Personality and Social Psychology*, 1975, *32*, 556–570.

Wyer, R. S., & Hinkle, R. Informational factors underlying inferences about hypothetical persons. *Journal of Personality and Social Psychology*, 1976, *34*, 481–495.

Received September 25, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 9, 1576–1588

Self-Directed Attention, Awareness of Bodily States, and Suggestibility

Michael F. Scheier
Carnegie-Mellon University

Charles S. Carver
University of Miami

Frederick X. Gibbons
University of Kansas

Two studies tested the hypothesis that self-directed attention would cause increased awareness of internal states and would thus reduce suggestibility effects. Experiment 1 applied this reasoning to the experience of an emotion. Males viewed moderately arousing slides of female nudes after being led to expect the slides to be either highly arousing or nonarousing. As predicted, ratings of the slides corresponded less with these experimentally-manipulated anticipations when self-focus was heightened by the presence of a mirror than when it was not. Experiment 2 examined a different internal experience: the perception of taste. Some subjects were led to expect a strong flavor as part of a test series, and other subjects were led to expect a weak flavor. Subjects high in private self-consciousness were less affected by this expectancy manipulation and more accurate in reporting their actual internal state than were subjects low in private self-consciousness. Discussion centers on the theoretical implications of the findings.

Casual observation suggests that many persons do not know themselves very well (cf. Nisbett & Wilson, 1977). When asked for a self-description, they provide their questioner with a picture fraught with distortions, ambiguities, and inaccuracies. Moreover, these vagaries in self-report are often unmotivated. That is, the inaccuracies exist simply because people are ignorant of their characteristics, not because of any attempt to deceive. As we all know, these impressions from everyday experience are also substantiated in the empirical literature (cf. Mischel, 1968). The correlations between self-report of behavior and actual behavior, for example, typically hover in the mid-twenties; rarely do they even reach into the thirties.

The authors express their appreciation to the following people, who assisted with data collection: Randall Heller (Experiment 1); Gaye Delevin, Russ Militello, and Steve Faloon (Experiment 2).

Requests for reprints should be directed to Michael F. Scheier, Department of Psychology, Carnegie-Mellon University, Pittsburgh, Pennsylvania 15213.

Recent research stemming from self-awareness theory (Duval & Wicklund, 1972; Wicklund, 1975), however, appears to indicate that this picture of people's self-knowledge need not always be so bleak. In particular, the theory suggests that self-focused attention may be an important determinant of the accuracy of self-reports. Presumably, a person whose attention is directed inward is more cognizant of his or her self-aspects than is a person whose attention is directed elsewhere. The heightened salience of self that occurs as a function of self-focus should be reflected in more accurate self-descriptions.

This seems indeed to be the case. In one study (Pryor, Gibbons, Wicklund, Fazio, & Hood, 1977), undergraduate men were asked for a self-report of their level of sociability; several weeks later their sociable behavior was objectively measured. The correlation between self-report and behavior for this group was .17. In contrast, when the procedure was repeated with a different group who made their self-reports in the presence of a mirror, the correlation approached .70. Related re-

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3709-1576$00.75

search has shown conceptually similar effects for persons who are dispositionally high in private self-consciousness (Scheier, Buss, & Buss, 1978; see also Carver, 1975; Turner, 1978). In brief, self-focused attention appears to be an important moderator variable that affects the accuracy of self-reports.

The primary purpose of the present research was to explore further the relationship between self-focused attention and accuracy of self-reports. In doing so, however, an attempt was made to extend the analysis to a different domain of self-knowledge. Thus, rather than study again the link between self-report and behavior, we chose instead to examine the effect of self-directed attention on knowledge of and reports about a variety of different bodily states.

Similar considerations should apply to awareness of one's bodily experiences as apply to knowledge about one's behavorial tendencies. That is, focusing attention on the self should lead to increased awareness of any experiential dimension that is salient at the time. When attention is focused on one's prior behavior, the accuracy of the self-report of that behavior is enhanced (Pryor et al., 1977; Scheier et al., 1978). If attention is focused on one's bodily state, on the other hand, then the person's self-report about the state should similarly become more accurate. Testing this reasoning was the major objective of the present research.

Suggestibility

The present studies also allowed us to confront another issue as well. We know from previous research that internal sources of information are not the only determinants of self-reports of internal states. To the contrary, many people seem to make extensive use of information from their environments in defining for themselves what their experiences are, even when those experiences are essentially internal ones. For example, people will make use of externally provided information in determining how intense their emotions are (e.g., Barefoot & Straub, 1971; Valins, 1966). To the extent that a person utilizes external rather than internal information when making judgments about his internal states, he is open to a wide variety of suggestibility phenomena. Indeed, a person who judged his bodily states totally on the basis of his internal sensations would be completely impervious to the influence attempts of others. Such suggestibility influences on judgments of bodily states can occur in many forms. For example, phenomena that would fit into this category include "demand" effects (Orne, 1962)—that is, the influence that an experimenter's verbalizations and demeanor can have on subjects' subsequent reports—and "placebo" effects—that is, the influence that misleading information about the actions of an inert substance can have on subjects' subsequent reports. What both these effects have in common is this: The subject accepts some externally provided information indicating the nature or intensity of the subject's recent, present, or impending experience as being accurate when it is not.

According to the reasoning presented above, self-directed attention should minimize suggestibility phenomena in cases where the external suggestion contradicts one's internal experience. That is, increased self-focus should increase the salience of veridical internal information, thus increasing the degree to which the latter is used preferentially in determining the subsequent self-report. As the misleading external information is disregarded in favor of the internal information, the suggestibility effect should diminish.

This reasoning has received one empirical test, in a study on placebo effects (Gibbons, Carver, Scheier, & Hormuth, 1979). Subjects in that study were given a placebo and were either misinformed or correctly informed about the effects of the drug. Misinformed subjects were told that the drug would produce a slight increase in heart rate, sweaty palms, and a tightness in the chest. All subjects were subsequently asked to report the degree of arousal and the specific arousing symptoms that they were experiencing. These reports were obtained either in front of a mirror or not. Neither mirror nor no-mirror subjects reported experiencing the target symptoms when they had been correctly informed about the nature of the drug. Among

misinformed subjects, however, those not exposed to a mirror apparently were taken in by the experimental cues and reported considerable arousal. In contrast, even though the specific target symptoms used in the study could not be directly observed by looking in the mirror, misinformed subjects exposed to a mirror reported minimal arousal. Indeed, the level of drug-produced arousal reported by self-aware subjects in the misinformed condition was no different from the level of drug-produced arousal reported by subjects in the correct information condition. Thus self-attention significantly reduced the placebo effect.

This finding has important theoretical implications, which will be considered more fully below under General Discussion. The data did come from a relatively circumscribed context, however—research on reactions to placebos. Because of that finding's potential importance, it seems highly desirable to assure its generality. Gaining additional information on this question was a second aim of the present research. Therefore, in pursuing our primary objective, which was to study the effects of self-focus on accuracy of self-report of internal states, we did not simply correlate subjects' internal states with their self-reports. Instead, we approached the accuracy issue by examining the role of self-attention in two kinds of suggestibility phenomena. Accuracy, then, is defined as resistance to an attempt to misdirect by experimental suggestion. The first study made use of an emotion as an internal experience. The second study examined perceptions of taste. In both cases experimental manipulations were introduced to influence subjects' anticipations about their upcoming experiences. In both cases, we predicted that heightened self-focus would lead to increased awareness of actual internal experience and would thus lead to enhanced resistance to suggestibility.

Experiment 1

In Experiment 1 we sought to investigate suggestibility in the experience of an emotion. The design of the study was somewhat more complex than that utilized by Gibbons et al. (1979), in the following respects. First, the manipulation used by Gibbons et al. was unidirectional. Experimental subjects were always led to believe that they would be experiencing more of the relevant symptoms than they actually did experience. In contrast, the suggestibility manipulation in the present experiment was bidirectional. All subjects in the present study were exposed to stimuli—slides of nudes—which have been shown to induce a moderate degree of affect—sexual attraction (Scheier & Carver, 1977). Some of the subjects were led to believe that the stimuli would be nonarousing, and others were told that the stimuli would be highly arousing. Thus, an attempt was made to drive subjects' opinions in both directions from a baseline.

In addition, the present research assessed the effects of self-focused attention in two very different ways—a strategy that has been effectively employed previously (e.g., Carver & Scheier, in press; Scheier & Carver, 1977; Scheier, Carver, & Gibbons, in press). Like Gibbons et al. (1979), we included a manipulation of self-directed attention in Experiment 1. Thus some subjects were exposed to their own images in a mirror and others were not. Validation evidence concerning the self-focusing properties of mirror presence have been presented at length elsewhere (Carver & Scheier, 1978). Unlike the subjects of Gibbons et al., however, subjects in the present experiment were also selected on the basis of their chronic tendencies to be self-attentive.

The disposition to focus attention inward has been termed self-consciousness and is measured by the self-consciousness scale (Fenigstein, Scheier, & Buss, 1975). The scale has two major components, private and public, that are only weakly correlated (cf. Fenigstein et al., 1975). Private self-consciousness involves a focus on the more personal and covert aspects of the self. The person high in private self-consciousness is more cognizant of his beliefs, moods, and feelings. Two sample items from the private self-consciousness subscale are "I'm generally attentive to my inner feelings" and "I'm always trying to figure myself out." Public self-consciousness, in contrast, reflects an awareness of the self

as a social object. Persons high in public self-consciousness are aware of their social appearance and the impressions they make on others. Two sample items from this subscale are "I'm concerned about my style of doing things" and "I'm usually aware of my appearance." It is private self-consciousness that is of greatest relevance to present concerns. That is, our prediction was based on the reasoning that self-attentive persons would be especially aware of their veridical emotional experiences. Private self-consciousness involves precisely these types of process: awareness of one's thoughts, feelings, and other internal aspects of self. Consequently, subjects were selected on the basis of private self-consciousness rather than public.

Both construct validity (Carver & Scheier, 1978) and discriminant validity (Carver & Glass, 1976; Turner, Scheier, Carver, & Ickes, 1978) have been established for the private self-consciousness subscale. Moreover, private self-consciousness as a disposition was intended to be similar conceptually to the manipulated state of self-awareness. The fact that on a number of occasions the two have led to empirically similar results indicates that they do converge on a single psychological entity (see e.g., Buss & Scheier, 1976; Carver & Scheier, 1978; Scheier, 1976; Scheier et al., 1978; Scheier & Carver, 1977). In brief, the parallel between manipulated self-awareness and private self-consciousness now seems well documented, and we expected similar results for the two variables in the present research as well. Thus self-focused attention was expected to lessen susceptibility to suggestion, regardless of whether self-focused attention was experimentally heightened by exposure to a mirror or heightened by virtue of the person's propensity to be dispositionally high in private self-consciousness.

Method

Subjects

Subjects were 55 male undergraduates from the University of Miami subject pool. Six additional subjects were deleted prior to data analysis because during debriefing they spontaneously reported the belief that the experimenter had been trying to influence their perceptions of the stimuli. Participation in the research was limited to males because the study used stimuli that had already proven to be somewhat arousing to males under relatively similar circumstances (Scheier & Carver, 1977), whereas we had no comparable knowledge of females' responses to the stimuli. Subjects, tested individually, were randomly assigned to one of the four experimental treatments described below.

Subjects were also divided into groups high and low in private self-consciousness based on their responses to items of the self-consciousness scale (Fenigstein et al., 1975), which had been administered in large group sessions several weeks prior to the study. Subjects were classified as being high in private self-consciousness if their scores fell above the median ($Mdn = 24.5$) of the pretest distribution and as low in private self-consciousness if their scores fell below the median.

Procedure

The procedure used in this study was adapted from that used by Scheier and Carver (1977). The experiment was portrayed as part of a larger project studying reactions to sexually oriented stimuli. Future subjects ostensibly would be wired with physiological recording devices to assess their bodily responses to the stimuli (photographic slides of female nudes). Before beginning that phase of the research, however, it would be necessary to evaluate and classify the stimuli that would be shown to those subjects. In particular, the experimenter continued, it was necessary to determine which slides were very arousing, which ones were minimally arousing, and so on; this was the ostensible purpose of the present sessions. The experimenter went on to explain that each participant would see only a few of the slides, however, in order to prevent overexposure to the stimuli on the part of any of the raters.

The subject was told that he would be viewing each slide twice. The first time he was simply to view it carefully and thoroughly; the second time he was to rate each slide on a separate 15-point scale, the endpoints of which had been labeled "extremely arousing" and "totally *non*arousing." The subject was cautioned not to make his ratings as a function of relative photographic or artistic qualities. Because the research to be done later concerned physiological responses, it was emphasized that all of the present judgments should be made only according to how much of a bodily reaction the slide seemed to produce in the subject, in terms of flushing, muscle tension, slight changes in heart rate, and so forth.

Suggestibility manipulation. After ascertaining that the subject had understood the instructions, the experimenter stood up to start the slide projector. As he did so, he delivered one of the following two statements in an offhanded manner. These statements constituted the instructional set or suggestibility manipulation. To subjects in the nonarousal condi-

tion the experimenter said, "By the way, you might as well know before you even start that the guys we've run so far with this particular set of slides have been reporting that they really aren't arousing at all—I mean, not at all. I've noticed that they've generally been circling points way over there on the left side of the page." To subjects in the high arousal condition the experimenter said "Well, it looks like you're one of the lucky ones. The guys we've run so far with this set of slides have been reporting that they're really pretty arousing. I've noticed that they've generally been circling points way over here on the right side of the page."

Despite these comments on the part of the experimenter, all subjects subsequently viewed precisely the same set of 7 slides. These slides had been selected from a larger group on the basis of pilot data indicating that these slides were all moderately arousing—i.e., that unmanipulated ratings of the slides fell approximately in the middle of the scales used in the main study. Thus both instructional sets constituted manipulations away from what should have been the slides' true effects on the subjects.

After delivering his remark, the experimenter started the slide projector, which automatically displayed each stimulus for 15 seconds, separated by 15-second interstimulus intervals in which no light was projected. After a 30-second interval, the cycle was repeated. During the projection period the experimenter looked away from the subject and said nothing except to inform the subject that he should make his ratings now, as the second cycle was about to begin.

Self-awareness manipulation. The slides were rear-projected onto a freestanding device that stood approximately 1 m in front of the subject's chair. This device was fitted with one of two screens (.5 m × .75 m), which constituted the manipulation of self-attention (cf. Neumann, Carver, & Scheier, 1977).[1] One, used in the no mirror condition, consisted of ordinary rear-projection material. The other, used in the mirror condition, was constructed of 50% transmission – 50% reflection glass, backed by tracing paper. The optical characteristics of the latter screen were such that when a slide was being projected, the subject could see only the slide, but during the interstimulus intervals the subject saw a reflected image of himself. Thus self-attention was enhanced during the slide series, but in such a way that the subject was not distracted from the slides. Following presentation of the slides, the subject was probed for suspicions about the suggestibility manipulation, debriefed as to the purpose of that manipulation, and dismissed.

Results and Discussion

The dependent measure was the subject's summed ratings of the arousal produced by the seven stimulus slides (see Table 1). An analysis of variance of these data revealed, as one might expect, a large main effect for instructional set, $F(1, 47) = 79.83$, $p < .001$. Subjects who expected the slides to be nonarousing reported being less aroused by them ($M = 42.74$) than subjects who expected the slides to be arousing ($M = 74.35$). Thus the expectancy manipulation seemed to have been effective.[2] Of greater interest, however, was the significant interaction between this suggestibility manipulation and mirror-manipulated self-awareness, $F(1, 47) = 4.33$, $p < .05$, of the expected form. Within both set conditions, the presence of a mirror led the slides to be rated more moderately. In short, mirror condition subjects were misled to a lesser degree by the experimenter's suggestions than were subjects in the no mirror condition. As indicated in Table 1, the strength of the mirror's influence on subjects' ratings was relatively equivalent across the two instructional set conditions. Thus, although the interaction term of the analysis of variance was significant, neither simple effect of the mirror was significant by itself.

Table 1
Mean Arousal Totals as a Function of Experimental Treatment (Combined Across Levels of Self-Consciousness), Experiment 1

Instruction	No mirror	Mirror
High arousal	78.69[a]	70.00[a]
Nonarousal	39.85[a]	45.63[b]

Note. The higher the number, the higher the arousal ratings.
[a] $n = 13$.
[b] $n = 16$.

The effects of private self-consciousness were similar in form to the effects of the

[1] Practical considerations made it impossible to prevent the experimenter from knowing which screen was in front of the subject. However, the experimenter did not know the purpose served by the different screens or the predictions being tested in the study.

[2] We had extensively pretested the expectancy manipulation in a pilot experiment, and the results of this pilot experiment showed that the manipulation was quite effective. It was for this reason that no expectancy manipulation check was obtained in either of the main experiments reported above.

mirror, but weaker. Thus, although the private self-consciousness group means were ordered in the expected direction in three of the four cases (in the fourth case the means were the same), the differences were smaller than those produced by the mirror, and no effect involving private self-consciousness reached an acceptable level of statistical significance.[3]

In part because of the weakness of the self-consciousness effect in Experiment 1, we sought to reexamine the effect of private self-consciousness on suggestibility, using a slightly different experimental setting and a more sensitive dependent measure. In Experiment 2, subjects' expectancies about a flavor intensity were experimentally manipulated prior to tasting a solution. Some subjects expected the solution to be strong, and others expected it to be weak. In contrast to Experiment 1, however, each subject in Experiment 2 served as his or her own control. This was done by individually adjusting each subject's rating of flavor intensity from his or her previously established baseline. It was predicted that subjects high in private self-consciousness would be less susceptible to suggestion than would subjects low in private self-consciousness, and would thus be less swayed by the information provided by the experimenter.

There was a second, more important, reason for conducting a second experiment. We have argued that self-focused attention makes a person more aware of his actual internal state and that the heightened salience of the person's internal state makes him less susceptible to influence. Although the findings from Experiment 1 were consistent with this reasoning, they provide only indirect evidence for our hypothesis. That is, no independent evidence was gathered that conclusively showed self-focused subjects to be more aware of their true internal state. We can only infer that they were, based on their reduced susceptibility to the bogus information provided by the experimenter. A similar criticism can be leveled against the earlier research conducted by Gibbons et al. (1979).

A second purpose of Experiment 2 was to provide a more direct test of our reasoning. Thus, in addition to manipulating subjects' expectancies about the flavor intensity they would be tasting, we also varied the actual intensity of the solution. Some subjects tasted a relatively strong solution, and others tasted a solution that was relatively weak. If self-focused attention reduces susceptibility to suggestion because it makes persons more aware of their actual internal states, then ratings of the actual intensity of the solution should be more veridical among persons whose attention is self-focused. Thus a second prediction of this study was that persons high in private self-consciousness would provide more accurate ratings of the actual intensity of the solution than persons lower in private self-consciousness.

Experiment 2

Method

Subjects

Three months prior to the experiment, several hundred undergraduate men and women at Carnegie-Mellon University completed the self-consciousness scale (Fenigstein et al., 1975). The private self-consciousness subscale was used to select subjects. Subjects were divided at the pretest median ($Mdn = 26.1$) into those high and those low in private self-consciousness.

A total of 43 men and 31 women participated in the study. Two men were dropped from the final data analysis: 1 because of suspicion and 1 because he failed to follow instructions. Within levels of self-consciousness, subjects were randomly assigned individually to one of the four treatment conditions described below, with the restriction that there be at least 8 subjects per group (cell *n*s varied from 8 to

[3] In order to determine whether public self-consciousness affected the subjects' ratings, subjects were redivided into eight new groups, based on a median split of the public self-consciousness subscale. A 2 (High vs. Low Public) × 2 (Arousal vs. Nonarousal Instructions) × 2 (Mirror vs. No Mirror) analysis of variance was performed on the data from these reconstituted groups. The results of this analysis revealed the same significant effects reported above. In addition, the main effect for public self-consciousness was also significant, $F(1, 47) = 4.36$, $p < .05$, such that subjects high in public self-consciousness reported being more aroused by the slides than did subjects low in public self-consciousness. No other significant effects were obtained.

10). As in Experiment 1, the experimenter remained blind with respect to the subject's self-consciousness score until after all data had been obtained.

Procedure

The experimenter, who was attired in a white laboratory coat, described the study as part of a larger physiological research project investigating an assortment of variables that influence bodily sensations. The present study ostensibly examined the effect of several of these variables on the experience of taste. After these brief introductory remarks, the subject was told that he/she would be tasting a number of different flavored solutions during the experimental session.

At this point, the experimental procedure was explained. The subject was asked to stand in front of a table, which was located next to a sink. On the table was a tray containing 10 small paper cups with a green colored solution in each. The cups were arranged in two rows. The first row was marked *sample* and contained four cups. The second row contained six cups and was labeled *test*. The subject was told that his/her primary task was to rate the intensity of the six flavoring agents in the test row. In order to acquaint the subject with the experimental procedure, however, he/she was being asked to rate the intensity of the four sample solutions first. In fact, only the first two sample solutions were ever rated. Intensity ratings were made along an 11-point scale, with the end points labeled "extremely weak" and "extremely strong." Inasmuch as the subject needed to have his/her hands free to handle the solutions, the intensity rating scale was posted on the wall in front of the subject so that he/she could make his/her rating verbally. Each rating was recorded by the experimenter.

The experimenter went on to explain that it was important that each person hold the flavoring agents in his/her mouth for the same length of time. Consequently, the experimenter would tell the subject when to put the solution in his/her mouth, how long to hold it there, and when to make the rating. The subjects held the solution in their mouths for 10 seconds before spitting it into the sink. They then waited 10 more seconds before making the intensity rating. The waiting period was to be used by the subject to think about the sensation produced by the solution. Once the rating was made, the subjects rinsed their mouths with water, waited for 60 seconds, and then began the next trial.

The first solution, which consisted of 1.4 parts peppermint extract mixed with 100 parts water, was the same for each subject. The second solution tasted by the subject also consisted of a mixture of peppermint extract and water but, depending upon treatment condition, was either stronger or weaker than the first solution. In the strong-solution condition, the subject tasted a solution consisting of 1.9 parts peppermint extract to 100 parts water. In the weak condition, the solution consisted of .9 parts peppermint extract to 100 parts water. The remaining eight solutions contained only water. Green food coloring had been added to all 10 solutions in order to make them similar in appearance.

In addition to manipulating the actual intensity of the second solution, the subject's expectancy about the intensity of the second solution was also varied. As the subject reached for the second solution, the experimenter casually remarked, "the second solution should be a little stronger (weaker) than the first." Thus, depending upon experimental condition, the subject expected the second solution to be either stronger or weaker than the first one he/she tasted. After the second solution was rated, the subject was probed for suspicion and debriefed.

The procedure required that each subject taste and rate the same first solution. This procedure was instituted so that the subject's rating of the first solution might be used as an anchor to which his/her second rating might be compared. Thus, an intensity change score was derived for each subject by subtracting his/her rating of the first solution from his/her rating of the second. This intensity change score constituted the primary dependent measure. A positive intensity score indicates that the second solution was rated as stronger than the first; a negative score indicated the opposite.

Results and Discussion

Private Self-Consciousness

A 2 (Strong vs. Weak Solution) × 2 (Strong vs. Weak Taste Expectancy) × 2 (High Private vs. Low Private Self-conscious) analysis of variance of the change scores revealed a significant main effect for solution, $F(1, 64) = 33.11$, $p < .0001$.[4] As might be expected, subjects who received a strong second solution rated it higher in intensity ($M = 1.09$) than subjects who received a weak second solution ($M = -.093$). There was also a significant main effect for the expectancy variable, $F(1, 64) = 11.75$, $p < .002$. Subjects who expected a strong second solution rated the second solution as being more intense ($M = .68$) than subjects who expected a weak second solution ($M = -.052$).

[4] Preliminary analyses of the intensity change scores revealed only negligible effects for the gender of the subject. The main effect for gender was nonsignificant, as were all the interactions in which gender was involved. Consequently, gender was dropped as a variable in the design, and the data from men and women were combined for all analyses reported above.

Of greater importance, however, was the significant Expectancy × Private Self-Consciousness interaction, $F(1, 64) = 4.15$, $p < .05$. Consistent with the findings of Experiment 1 and as predicted, subjects low in private self-consciousness were more affected by the expectancy manipulation than were subjects high in private self-consciousness (see Figure 1). The former group rated the second solution as being more intense when a strong solution was expected, and the second solution as less intense when a weak solution was expected, than did subjects higher in private self-consciousness.

There was also a marginally significant Solution × Self-Consciousness interaction, $F(1, 64) = p < .16$. As can be seen in Table 2, the form of this interaction was exactly opposite to that of the Expectancy × Self-Consciousness interaction. Whereas subjects high in private self-consciousness were *less* sensitive to the expectancy manipulation, they tended to be *more* sensitive to the actual intensity of the solution. As predicted, they tended to rate the strong solution as being stronger and the weak solution as being weaker than did subjects low in private self-consciousness.

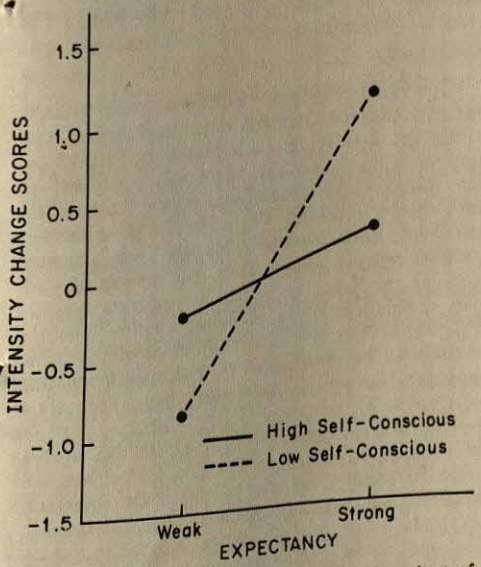


Figure 1. Intensity change scores as a function of private self-consciousness and subject's expectancy, Experiment 2.

Table 2
Intensity Change Scores as a Function of Private Self-Consciousness and Solution Intensity, Experiment 2

Self-consciousness	Weak solution	Strong solution
High private	−1.25	1.26
Low private	− .60	.91

We expected that subjects high in private self-consciousness would be more accurate in rating the actual intensity of the solution than subjects low in private self-consciousness. The form of the Solution × Self-Consciousness interaction was consistent with this hypothesis, but the interaction failed to reach an acceptable level of statistical significance. It is important to note, however, that the Solution × Self-consciousness interaction may not have been the most appropriate test of the accuracy hypothesis. For half the subjects in the study, no conflict existed between their expectancies and the actual intensity of the solution they tasted. In the absence of conflict, subjects high and subjects low in private self-consciousness should have been equally accurate in their ratings of flavor intensity. Only when the subject's expectancy conflicted with his or her actual experience should private self-consciousness have made a difference in the subject's ratings. The high group, whose judgments were presumably based on their actual internal states, should have been very accurate in their assessment of flavor intensity under conflict conditions, whereas the low group, whose judgments were based less on their internal states, should have been misled more by the experimenter's suggestions.

Given these considerations, a slightly different analysis of the intensity change scores was performed. Specifically, subjects were first separated into those who experienced conflict and those who did not experience conflict between expectancy and solution intensity. High and low private self-conscious subjects were then classified into those who were accurate in their ratings of flavor intensity and those who were not. A subject was classified as accurate if the sign of his or her

Table 3
Number of Persons Making Accurate Flavor Intensity Ratings as a Function of Private Self-Consciousness and Presence or Absence of Conflict Between Expectancy and Actual Solution Intensity, Experiment 2

	Conflict absent		Conflict present	
Self-consciousness	Accurate	Inaccurate	Accurate	Inaccurate
High private	12	4	12	5
Low private	17	2	5	15

intensity change score was in the same direction as the actual intensity of the second solution tasted. Thus a subject was classified as accurate if he or she rated the second solution as being weaker than the first and it was in fact weaker, or if he or she rated the second solution as being stronger than the first and it was in fact stronger. A subject was classified as inaccurate either if his or her change score was zero (i.e., if the subject failed to report a change in flavor intensity when one in fact occurred) or if the sign of his or her change score was opposite in direction to the actual intensity of the second solution tasted.

The results of this analysis are presented in Table 3. As can be seen from the left half of this table, when there was not conflict between expectancy and solution intensity, the accuracy of flavor ratings was high for both subjects high and subjects low in private self-consciousness. The difference between the two groups was trivial and nonsignificant, $\chi^2 < 1.0$. A very different pattern of results emerged, however, when accuracy was examined among only those subjects for whom a conflict existed between expectancy and solution intensity. As can be seen from the right half of Table 3, subjects high in private self-consciousness were still quite veridical in their assessment of flavor intensity. Of the 17 subjects high in private self-consciousness for whom conflict existed, 12 rated the intensity of the second solution accurately, and 5 rated it inaccurately. In contrast, of the 20 subjects low in low private self-consciousness who were in conflict, only 5 rated the flavor intensity accurately. This difference in the accuracy of the intensity ratings made by subjects high and low in private self-consciousness who were in conflict was highly significant, $\chi^2(1) = 5.96$, $p < .02$.[5] These findings strongly suggest that high private self-conscious persons are in fact more cognizant of their actual internal states than low private self-conscious persons, as predicted. Moreover, these findings also lend credence to the contention that it is precisely such differences in awareness of internal states between people who are high and people who are low in private self-consciousness that accounts for the differences between them in susceptibility to suggestion.

It should be noted that our purpose in conducting the present research was to document the fact that self-attentive persons are less susceptible to suggestion than persons whose attention is not self-focused, and the findings from both experiments offer strong evidence that this is indeed the case. It is

[5] The above analyses tested the accuracy hypothesis by using nonparametric statistics. However, two 2(High vs. Low Private) × 2(Strong vs. Weak Solution) analyses of variance were also performed on the raw intensity change scores—one analysis for each conflict condition. The results of these analyses paralleled those obtained for the nonparametric analyses reported above. Thus, the analysis of variance of the data for no conflict subjects revealed only a significant main effect for solution intensity, $F(1, 31) = 46.09$, $p < .001$. Subjects rated the strong solution as being stronger than the weak solution, regardless of whether they were high, $F(1, 31) = 18.37$, $p < .01$, or low, $F(1, 31) = 28.22$, $p < .01$, in private self-consciousness. In contrast, the analysis of variance performed on the data for subjects who were in conflict revealed a significant interaction between self-consciousness and solution intensity, $F(1, 33) = 5.61$, $p < .03$. When in conflict, subjects high in private self-consciousness still rated the strong solution as being stronger than the weak solution, $F(1, 33) = 6.77$, $p < .02$, but those low in private self-consciousness did not, $F < 1.0$.

also true, however, that self-attention did not entirely eliminate suggestibility effects. That is, even self-aware subjects in Experiment 1 differed in their ratings of the stimulus slides as a function of the arousal instructions (see Table 1), and even high private self-conscious subjects in Experiment 2 differed in their ratings of solution intensity as a function of the expectancy manipulation (see Figure 1). What is important, though, is that self-attentive subjects were significantly less susceptible to both of these manipulations than were subjects who were less self-attentive.

Subsidiary Analysis

In order to determine the effect of public self-consciousness on intensity ratings, subjects were redivided into eight new groups, based on a median split of the public self-consciousness subscale. A 2 (High vs. Low Public) × 2 (Strong vs. Weak Solution) × 2 (Strong vs. Weak Taste Expectancy) was then performed on the intensity changes scores. The previously mentioned main effect for solution was still significant, $F(1, 64) = 29.81$, $p < .001$, as was the main effect for the expectancy variable, $F(1, 64) = 13.38$, $p < .001$. Neither the main effect for public self-consciousness, $F(1, 64) = 2.46$, $p < .05$, nor any interactions involving public self-consciousness were significant. Thus, the effects of public self-consciousness in Experiment 2 were negligible.

General Discussion

Experimental Findings

The idea that attention plays an important role in the perception of bodily states is not a new one. Schachter and Rodin (1974), for example, have discussed the importance of this variable in their obesity research, drawing a distinction between internal and external personality types. They point out that "externals" tend to rely upon external cues for information about their internal states, whereas "internals" respond more to internal cues and appear to be more in touch with their actual bodily states. The present research takes this notion a step farther, however. The results of both studies appear to indicate that self-attention can lead to heightened awareness of internal experience even when the experience directly contradicts experimentally induced expectancies.

Moreover, the greater accuracy of self-report that comes with increased attention is not limited to reports of bodily states. That is, previous research has shown that self-reports of behavioral tendencies are most accurate when obtained under conditions of experimentally heightened self-awareness (Pryor et al., 1977). Similarly, persons high in private self-consciousness have been shown to provide more veridical self-reports of their behavioral tendencies than do persons lower in private self-consciousness (Scheier et al., 1978). In brief, self-directed attention appears to be an important determinant of the extent to which a person is aware of his or her bodily states, attitudes, and dispositions. As such, it is also an important determinant of the accuracy with which those states, attitudes, and dispositions are reported.

Related Suggestibility Findings

We have shown in the present research that self-focused attention can heighten a person's cognizance of his or her internal state and thereby make that person *more* resistant to the suggestions and influence of others. In contrast, previous research has shown that heightened self-focus can also lead a person to be *less* resistant to the influence of others and more likely to conform (Carver, 1974; Duval, 1976; Wicklund & Duval, 1971). For example, Wicklund and Duval (1971) found that subjects who listened to their own voices on a tape recorder conformed more with the modal opinion of a positive reference group than did subjects who listened to the tape-recorded voice of another person.

There are at least two ways, however, in which these previous studies differ from the present research, and these differences between the two sets of studies may account for the different sets of findings. The first difference involves the specific self-aspect toward which the influence attempt was di-

rected. Our own research and the earlier research by Gibbons et al. (1979) attempted to influence the person's belief about his or her internal bodily state. This research uniformly showed greater resistance to suggestion as a function of increased self-focus. In contrast, the earlier research showing decreased resistance to suggestion either attempted to influence the person's opinions, beliefs, or standing on some attitude topic (Wicklund & Duval, 1971, Experiment 1) or attempted to influence his or her assessment of an ambiguous external stimulus (Duval, 1976).

Thus, in each of these earlier studies, an attempt was made to influence the person's opinion about some aspect of self (attitudes or beliefs) or some aspect of the external environment (an ambiguous stimulus) for which there was no clear right or wrong position. That is, in each case the correctness of the person's position required extensive social validation (cf. Festinger, 1954). In contrast, the correctness of one's perceptions of one's own bodily states may require less social validation. Persons have direct access to information that enables them to determine for themselves whether they are aroused or not, whether something tastes strong or weak, and so on.[6] In brief, we suggest that important differences exist in the extent to which a person's judgments about himself or herself require social validation. For some self-aspects, like bodily states, little social validation may be necessary, and it may be for only those self-aspects, like bodily-states, that self-focused attention will increase resistance to suggestion.

A second possible difference between the two sets of studies rests on the distinction that has been drawn between private and public self-consciousness. Although the private–public distinction has so far been applied solely to individual differences in self-consciousness, only a small extrapolation is needed to apply the distinction to manipulations of self-awareness as well. Indeed, Buss (in press) has already suggested that self-awareness manipulations might be usefully categorized in terms of whether or not they are likely to induce awareness of the private or public aspects of self. He argues, for example, that exposure to a mirror is likely to induce private self-awareness, and the parallels that have been obtained between the mirror and private self-consciousness would seem to substantiate this claim (see e.g., Carver & Scheier, 1978; Scheier, 1976; Scheier & Carver, 1977). On the other hand, a variety of manipulations might induce public self-awareness and make the person more cognizant of himself as a social object. Examples of manipulations that might fall into this category are exposure to an audience, exposure to a TV camera, and hearing one's own tape-recorded voice.

Although the effect of public self-consciousness in the present research was negligible, it should be remembered that we attempted to influence the person's perception of his or her bodily state. The fact that public self-consciousness had little effect in this circumstance is not necessarily unexpected. That is, perhaps public self-consciousness is relevant to suggestibility only when the influence attempt involves self-aspects that require more extensive social validation. In this regard, it is interesting to note that virtually all of the earlier research in which self-focused attention increased suggestibility used manipulations that might have heightened public self-awareness—for example, the person's tape-recorded voice (Wicklund & Duval, 1971) and a TV camera (Duval, 1976). In addition, the influence attempts in those studies were aimed at self-dimensions that typically required validation from others—an attitude (Wicklund & Duval, 1971) and an ambiguous judgment (Duval, 1976). Thus, the manipulations used in those studies may have increased

---

[6] It should be noted that these comments about social validation and ambiguity are meant *only* to apply to the person's ability to identify in himself the presence or absence of a particular physiological state—e.g., identifying the presence or absence of arousal. They are not meant to apply to the person's ability to provide a name for his state—e.g., labeling the arousal as fear (cf. Schachter & Singer, 1962). Nor are they meant to apply to the person's ability to identify the source or cause of his state—e.g., knowing that the arousal was produced by running (cf. Zillman, Johnson, & Day, 1974).

the person's awareness of self as a social object, with a concomitant increase in concern over the appraisal and evaluations of others, in contexts where others' opinions are needed to evaluate one's own position. If so, the fact that these earlier studies found increased suggestibility due to heightened self-focus becomes less surprising.

Theoretical Implications

As was noted earlier in this article, the finding that self-directed attention can weaken suggestibility phenomena has important theoretical implications. These implications concern two alternative interpretations that might be invoked to explain findings of previous research in which mirrors were used as manipulators of self-focus. One of those interpretations is based on drive theory, the other on the notion of experimental demand.

In its simplest form, a drive-theory interpretation of mirror effects holds that the presence of a mirror leads to heightened drive or arousal, which in turn causes dominant responses to be emitted with increased frequency. Thus subjects' behavior is more in line with salient behavioral standards when a mirror is present than when it is absent (cf. Liebling, Seiler, & Shaver, 1974). The second potential explanation of the effect of a mirror is that it makes subjects more sensitive to demand characteristics (Orne, 1962) that are inherent in the experimental situation. According to this argument, subjects in the presence of a mirror do what they think the experimenter wants them to do to a greater degree than do subjects with no mirror.

Although either of these two explanations might provide an adequate account for some previous mirror-induced effects (e.g., Carver, 1974, 1975; Gibbons, 1978; Scheier, Fenigstein, & Buss, 1974), there are a number of other findings that neither approach can explain (e.g., Duval & Wicklund, 1973; Gibbons, et al., 1979; Carver & Scheier, Note 1). Nor can they account for the present findings. If a mirror increases arousal, all of the subjects in the present Experiment 1 who responded in the presence of a mirror should have experienced—and thus reported—more arousal than those with no mirror. Instead, the mirror manipulation interacted with the bogus information provided the subjects. Alternatively, if a mirror increases the emission of dominant responses, the interaction should have taken a form opposite to that which did occur, as subjects emitted the responses that had been rendered dominant by the experimenter's suggestions. Finally, if a mirror increases responsivity to experimenter demand, then self-aware subjects in the present research should have been more influenced by the experimenter's comments. The fact that they were less swayed by the suggestions of the experimenter than were less self-attentive subjects renders a demand-based interpretation of the present and previous findings much less tenable.

It is also worth noting that any attempt to apply drive or demand interpretations to mirror effects seemingly implies that such interpretations should be applicable to self-consciousness effects as well. The self-consciousness scale, however, has been found to be unrelated to measures of test anxiety, social anxiety, achievement motivation, general emotionality, impulsivity, intelligence, and social desirability (Carver & Glass, 1976; Turner et al., 1978). Drive or demand interpretations of self-consciousness effects would seem to require that some of these correlations be positive. In brief, the most viable interpretation of the accumulated research involving both manipulated self-awareness and dispositional self-consciousness appears to be based on attention. It is the only explanation that can simultaneously account for the effects of each of these variables, and it is the only explanation that can account for the entire set of findings.

Reference Note

1. Carver, C. S., & Scheier, M. F. *The self-attention-induced feedback loop and human motivation: A control-systems analysis of social facilitation.* Manuscript submitted for publication, 1979.

References

Barefoot, J. C., & Straub, R. B. Opportunity for information search and the effect of false heart rate

tive support for age-related changes regarding some of the levels (a Levels × Populations interaction), although their results are open to alternative interpretations.

Perhaps the most fundamental criticism relates to the adequacy of the stories used. First, there is doubt as to whether at least some of the stories reflect the criteria that they are intended to represent, especially at the foreseeability level (Sedlak, 1978). Second, factors such as effort that have been shown to affect AR (Raymond & Dillehay, Note 1) are confounded with the attribution criteria. Third, each level is presented in different situational contexts, a variable that affects AR according to Shaw and Sulzer's (1964, p. 45) qualitative observations. The inclusion of only two stories at each level may be insufficient to prevent the confounding of level and story context. Finally, the use of the same story character is likely to increase subject reactivity. These kinds of factors possibly produced anomalous findings such as a decrease in children's AR from foreseeable (but accidental) to intentional negative outcomes. Later research in this tradition (e.g., Shaw & Reitan, 1969; Shaw & Schneider, 1969; Sulzer & Burglass, 1968) is similarly plagued by basic problems of interpretation, and even though a new more satisfactory set of Perry stories has been devised, they have to date only been used with children in a cross-cultural study (Shaw & Iwawaki, 1972).

In an attempt to meet the above criticisms, Harris (1977) used a single situation to present Heider's levels to five different age groups (7-, 9-, 12-, and 13.5-year-olds and college students). Utilizing a between-subjects design, he obtained partial support for Heider's theory. Although an Age × Stimulus Level interaction was found for the perception of causality, this interaction only approached significance ($p < .09$) for AR. Furthermore, no distinction was found between causality and foreseeability despite a successful check on the manipulation of this variable. Finally, levels of the stimulus event did not affect 9-year-olds' perceptions of naughtiness but did have an effect on adjacent age groups. No explanation is offered for this result. Perhaps the strongest support for the age-related nature of Heider's levels stems from a recent study by Sedlak (1979). Despite an exceptionally small sample, she showed via multidimensional scaling that 8-year-olds, 11-year-olds, and adults utilized Heiderian criteria as a dimension in their cognitive representations of stimulus stories. Moreover, the salience of this dimension (relative weighting) was directly related to age, supporting the notion of increasing AR differentiation with development. However, even adults did not distinguish Heider's last two levels of intentionality and justifiability, thus replicating Shaw and Sulzer's (1964) earlier results. It thus appears that although previous research is adequate for testing aspects of Heider's (1958) ideas, the relevant attribution criteria have not been adequately operationalized, nor have developmental differences in AR been fully investigated. For example, the implicit assumption that AR is cumulative because higher levels imply lower levels of attribution (e.g., perceived intentionality implies perceived foreseeability, causality, etc.) has never been evaluated. Consequently, the status of Heider's last two levels (intentionality and justifiability) is unclear, and it is not known whether age differences in AR actually reflect developmental stages.

One aspect of Heider's theory, the development of AR based on intention, appears to have been established by Piagetian research. Both early and recent methodologically innovative studies show that the degree to which intentional perpetrators are held responsible for outcomes increases with the age (e.g., Gutkin, 1972; Hebble, 1971). As Karniol (1978) notes, however, these studies tend to vary intentionality in the context of accidentally produced outcomes and hence are not valid tests of Heider's (1958) two levels. In contrast, some investigations have not found the traditional age shift in evaluations of intentional versus accidental acts (e.g., Armsby, 1971; Farnill, 1974; Sedlak, Thompson, & Sands, Note 2). Whatever the developmental status of intention information utilization, this literature is clearly limited to testing only a single aspect of Heider's theory.

Research in Heiderian and Piagetian traditions displays several common features. In both attempts to understand, developing no-

tions of AR occur via the child's evaluation of another's actions. However, Piaget (1932) clearly stated that children's moral judgments of themselves and others would differ. In respect to the self, children succeed much earlier (at age 3 or 4 years) "in differentiating intentional faults from involuntary breaches of the moral code" (p. 180). A second limitation relates to the almost exlusive focus on a single attribution measure. With the exception of Harris (1977), all studies assess only moral attribution. But as Shaver (1975, p. 102) points out, attribution of causality and responsibility are not necessarily synonymous, although they are often assumed to be. Finally, attempts to obviate subject reactivity have often resulted in the use of between-subjects designs, whereas a within-subjects design is necessary for investigating the cumulative nature of Heider's AR levels.

Experimental Overview

The present study attempts to clarify certain conceptual issues and meet several of the methodological problems outlined above. Conceptually, it was considered that Heider's (1958) analysis of AR was incomplete at the highest levels. Although a facilitative cause is held to diminish responsibility (justifiability), no mention is made of a converse negative environmental force. Theoretically, such a force should, in accordance with Kelley's (1971) augmentation principle, increase responsibility, and hence a sixth level was hypothesized. In addition, it seems plausible that a more accurate developmental picture may be yielded by investigating AR in relation to the self. Consequently, negative-outcome stimulus stories representing six attribution criteria were administered to five developmental groups (6–7.5, 8–9.5, 10–11.5, 12–13.5 years, and adults) under two stimulus conditions; the actor was either a hypothetical other or the self.

Methodologically, this study differed from previous experiments in two major respects. First, a Latin square design was used to control for story context. Hence levels were varied within a single story theme. To retain the advantages of a within-subjects design (cf. Sulzer, Note 3) and in order to evaluate the scalability of Heider's levels without increasing subject reactivity, six story contexts were generated yielding 36 stimulus items. Subjects were thus presented with a different level from each of the story themes. Second, two different attribution measures were used. One required an ascription of causality, whereas the other asked for an assessment of blame.

It was hypothesized that the attribution measures would evoke two distinct attribution styles leading to different response patterns. More specifically, an interaction between age and story character (self versus other) was predicted for the moral judgment measure on the basis of Piaget's (1932) theory. Younger children were expected to evidence a difference in these two conditions, whereas older ones were not. No such interaction was predicted in regard to perceived causality. In terms of Heider's (1958) theory, it was anticipated that both attribution measures would evidence an effect for stimulus level, the subject's age, and the interaction of the two factors. Generally, younger children were expected to show a high and more global level of attribution across the stimulus levels in contrast to older subjects' differentiated attributions, which should show greater differences across the levels.

However, it was anticipated that the nature of the interaction between age and stimulus level would be different for the two attribution measures. In regard to blame, a convergence of age groups was hypothesized to occur only at the higher levels, including intentionality, whereas attribution of causality might coincide at the levels of causality or foreseeability. This prediction was based on the belief that judgment of causality is not primarily an evaluative event and may therefore result in an earlier consensus. Also, attribution of blame at the adult level should only be maximal at the highest stimulus levels, whereas the less differentiated responses of children may show such blame at lower levels.

Finally, it was held that Heider's AR model should display the characteristics of a Guttman scalogram as suggested by Fishbein

and Ajzen (1973). However, previous evidence for the developmental nature of these AR criteria is based solely on quantitative group scores and is thus open to question. No attempt has been made to look at individual response patterns in a developmental context.

Method

Subjects

Two hundred and forty persons participated in the study. Children were drawn from two Oxford schools and fell into one of four age groups determined by whether it was their second (for age, $M = 82.75$ months, $SD = 4.83$), fourth ($M = 105.55$ months, $SD = 5.05$), sixth ($M = 132.47$ months, $SD = 3.43$) or eighth ($M = 150$ months, $SD = 4.79$) year of formal education. To control for socioeconomic class, the schools selected were sited on the same council housing estate (predominantly lower middle class) within 600 m of each other. The adult group comprised both graduates and undergraduates at the University of Oxford as well as students from Oxford Polytechnic.

Equal numbers of male and female subjects were tested at each age level. The sample was almost exclusively white; only one or two black children were present in each age group.

Stimulus Stories

Stimuli comprised stories reflecting six familiar contexts (e.g., riding a bicycle) within which the attribution levels were varied. The construction of the stories followed three stages. Initially, 10 graduate students in psychology rated the stimuli on each of Heider's extended attribution criteria. On the basis of these responses, a second modified set of stories was constructed and again rated. Informal discussion following this procedure led to further but less extensive modifications. The full set of 36 stories was then presented to 6 senior PhD students in social psychology who had not participated in the previous phase. The students were asked to classify each story according to level after they had read Heider's (1958, pp. 113–114) description of the levels and the rationale for including a further stage. Because unanimity was obtained, no further structural changes were made. Finally, the stimuli were informally discussed with two teachers who were asked to comment on the linguistic complexity of the stimuli and on the children's familiarity with the words used. A few minor alterations consisting mainly of word changes were made at this stage.

Because the precise manner in which Heider's levels of responsibility are operationalized is of some importance, the structural features of each are presented below. It should be noted that an attempt was made to follow the exact general examples given by Heider:

Level 1 (association). At this level the actor was associated with the outcome in two ways; by spatial and temporal contiguity and by ownership. Typically, a victim is accidentally hurt in the actor's presence and by an object belonging to the actor.

Level 2 (causality). Here the actor merely causes the outcome. In a situation where competence can reasonably be assumed, the actor accidentally hurts a victim of whose presence he or she is not even aware (e.g., the actor slips while running).

Level 3 (foreseeability). Although the harm is still caused accidentally, foreseeability is achieved by a slight change of conditions. The actor experiences a lack of competence (e.g., the actor continually slips while running along a wet surface) and is aware of being a potential harmdoer (the actor sees the victim come and run too close to him or her.)

Level 4 (justifiability). Justifiability was represented by an "eye for an eye" situation. Hence when the actor saw someone prepare to attack him or her in some way (e.g., run past and knock the actor over while running), the actor responded to the provocation by doing to the other person exactly what would have been done to the actor as the victim.

Level 5 (intentionality). The actor behaves in the same manner as at the previous level, but there is no provocation. The actor thus appears as a free, uncoerced agent.

Level 6 ("supererogation"). This level is exactly the same as the previous one, except that the actor contravenes a request by the victim's mother to be careful "as my boy (girl) gets hurt very easily." (Notice that the order of Heider's, 1958, Levels 4 and 5 has been reversed to create a ranking of increasing internality of causation.) [1]

A characteristic common to all the stories was that they represented a realistic description of an action sequence; they only contained information that would have been available to an observer of the episode. Hence attributional criteria (e.g., intentionality) were not explicitly mentioned in any of the stories. However, an attempt was made to keep the demands made on social inference abilities at a bare minimum to avoid results that merely reflected age-related differences on this variable. A final feature of the stories was the use of only two similar outcomes. In three of the story contexts, the consequence of the actor's behavior was a bleeding nose, whereas in the other three he/she produced a bleed-

[1] Attribution of responsibility is a function of both the observers' response level and the stimulus level. Heider orders his criteria according to the former, whereas the reversal is made on the latter. The present attempt to order the levels according to increasing attribution was done to allow the possible emergence of a Guttman scale response pattern.

ing lip. These outcomes were chosen because they were judged to be of medium intensity.

Stimulus stories were combined to produce six sets of stimuli that contained one attribution level from each story context. The procedure followed is detailed in Winer (1971, p. 689). Since a Greco-Latin square design, which would have allowed assessment of any order effect in addition to the context variable, was impracticable in terms of subject number, a fixed order of presentation was followed. The exact order was randomly determined. It should be noted that any order effect should have worked against the hypotheses of the study.

Procedure

All subjects were administered one of the six stimulus story sets. To facilitate younger children's understanding of the task requirements, 6- and 8-year-olds were read the stories in small groups of 4 so that individual attention could be given to each subject. Owing to practical difficulties, all other subjects read the stories in normal class (lecture) periods (25–30 persons). An attempt was made to avoid comprehension difficulties among the older children by specifically excluding subjects known to have reading difficulties or learning problems of any sort. In addition, three adults were freely available to assist whenever any difficulty was encountered by a subject. Their help was, however, rarely required, as almost all the children found the task requirements fairly simple. A further precaution was taken in constructing the test booklets. Each story was presented on a separate page followed by the appropriate rating scale alongside which appeared a clear line drawing of the central object in the story. Any comprehension difficulties should have led to random responding and would therefore not have supported the experimental hypotheses.

Before proceeding with the task, each group was told that sincere and honest individual responses were important even though the study was concerned with group differences. Subjects were assured of complete anonymity to allay any further anxiety, and the importance of the results was emphasized. In the self–actor condition, subjects were told that they would be presented with imaginary stories about themselves and were asked to respond to the situation as if it had actually occurred. The other–actor condition stories were described as situations that had occurred, and subjects were asked to give their evaluations. Some time was spent explaining the use of the rating scales and ensuring that the concepts of blame and cause were understood. Where necessary, examples were given. It should be emphasized that both the language used and extent of elaboration given were varied as a function of the subjects' age and apparent comprehension. The latter was considered more important than adherence to a rigidly worded set of instructions. Finally, subjects were given two very explicit practice items and their correct use of the rating scale was checked by supplying appropriate feedback.

Each subject's responsibility attribution was assessed by asking, "How much should ACTOR be blamed for the boy (girl) getting hurt?" (in the experimental situation the appropriate name or pronoun was used). Subjects indicated their answers on a 7-point rating scale comprising a single empty rectangle and six Xs of increasing height. The rectangle was labeled "none," the midpoint, "medium," and the last X, "totally." The first and fifth Xs were labeled "very little" and "very much," respectively. The rectangle was used to emphasize the qualitative distinction between attributing blame and no blame, in contrast to the quantitative differences in the amount of blame assigned, represented by increasing Xs.

Subjects were then asked, "How much was ACTOR the cause of (reason for) the boy (girl) getting hurt?" Again, subjects indicated their answers by ticking an appropriate scale point. Darkened rectangles replaced the Xs in an attempt to ensure that the two questions that followed each story were answered on the appropriate scale. The order of the questions was not varied, since it was judged that the use of only two questions was unlikely to result in any fatigue or similarly distracting effects.

Results

An initial 2 (condition) × 5 (age) × 6 (level) × 2 (attribution measure) analysis of variance was conducted to determine whether separate analyses for the two dependent measures were justified. As hypothesized, a significant main effect for attribution measure, $F(1, 230) = 6.52$, $p < .05$, was found. Hence perceived blame and causality served as dependent measures in separate univariate analyses.

Because the self-versus-other as actor was hypothesized as an interaction effect, analyses for sex and story context differences were conducted on the data for each condition separately.

Sex Differences in Attribution

To test for sex differences, a 5 (age) × 6 (level) × 2 (sex) analysis of variance was carried out. In respect to both perceived blame and causality, sex produced neither a significant main effect nor an interaction in either actor condition ($p > .10$). Consequently data for both sexes were combined in subsequent analyses.

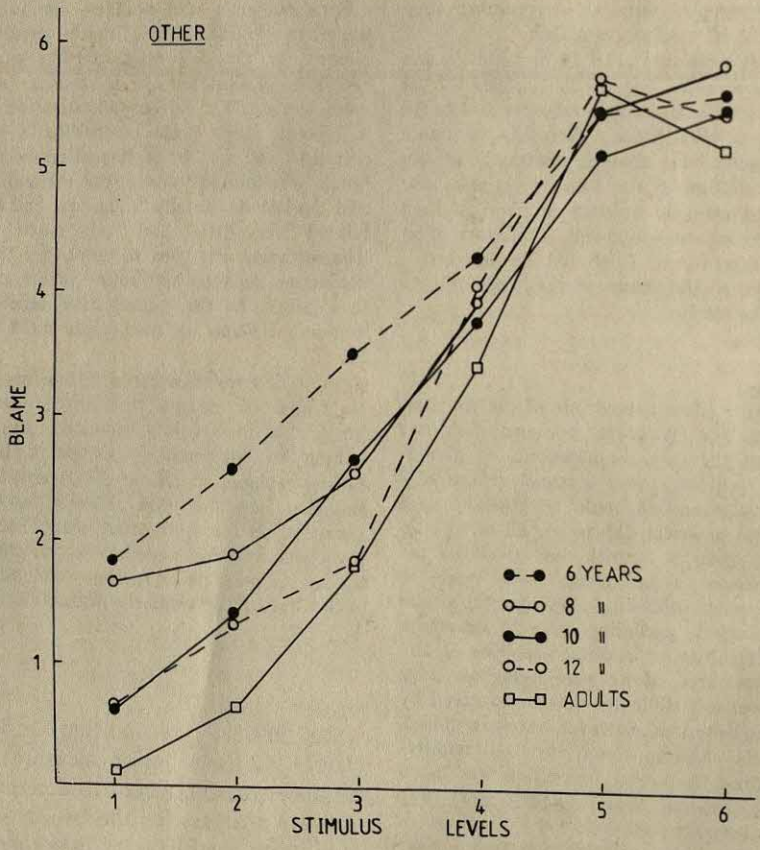


Figure 1. Mean moral attribution as a function of stimulus level for other as actor.

Story Context Effects

In evaluating possible story context effects, age was not included as a separate factor because stable differences were considered unlikely to emerge with only four subjects per cell. A 6 (story) × 6 (level) analysis of variance showed that story context did not emerge either as a main effect or in interaction across both dependent measures and administration conditions ($p > .10$). All further analyses therefore also ignored story context.

Measure of Perceived Blame

It was hypothesized that subjects' blame judgments would be affected by the subjects' age, the stimulus level, and the interaction of these two factors, whereas administration condition (self vs. other) was held to be important only at a younger age. Subjects' mean scores in both conditions are displayed in Figures 1 and 2. These data curves suggest a configuration consistent with the predicted results. An analysis of variance performed on the data revealed that the effects of subject's age, $F(1, 230) = 13.92$, $p < .001$; stimulus level, $F(5, 1150) = 396.79$, $p < .001$; the interaction of these two factors $F(20, 1150) = 4.92$, $p < .01$; and the interaction of story character with age, $F(4, 230) = 2.73$, $p < .05$, were all significant. No other significant results were found.

To further investigate the relationship between the effects of subjects' age, the stimulus level, and their interaction, simple main effects analyses of variance were computed following Kirk's (1968, pp. 284–294) recommended procedure. The results were consistent with the prediction that younger sub-

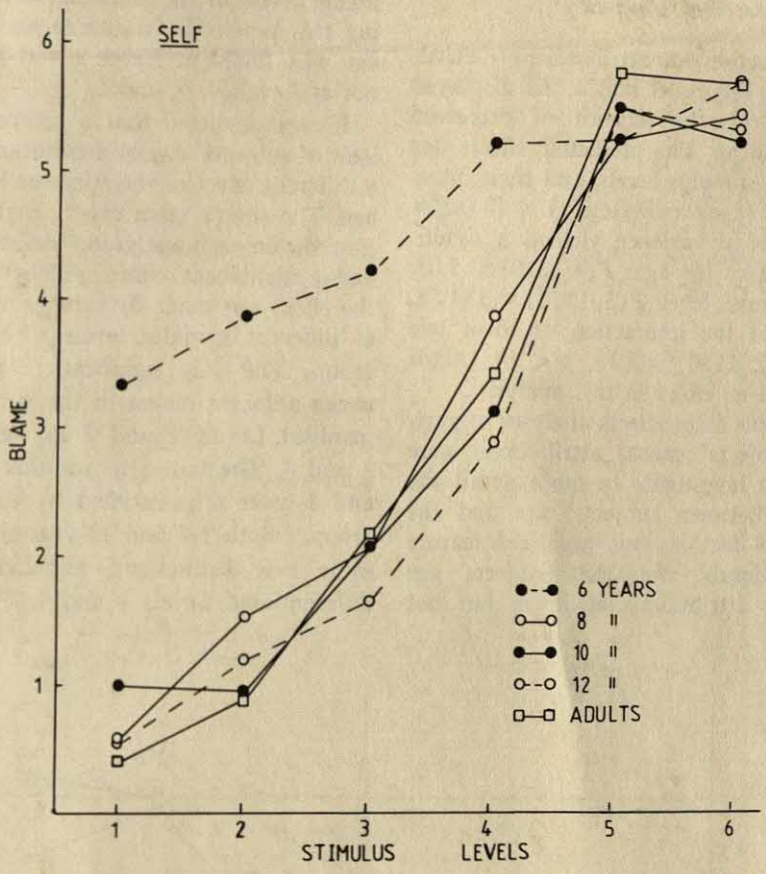


Figure 2. Mean moral attribution as a function of stimulus level for self as actor.

jects would attribute more blame at lower levels of the stimulus factor than adults. A simple main effect for age was found in all but the last two levels (in these and subsequent analyses the .01 level was adopted unless otherwise stated).

A second predicted feature of the age–stimulus level relationship related to possible differential effects of the stimulus level at various ages. The simple main effects analysis showed a significant stimulus level effect at all ages. To probe these effects in greater detail, comparisons between adjacent levels of the stimulus factor were made. The pattern of results supports the notion of increased differentiation of AR criteria with age. While 6-year-olds significantly distinguished only accidental from intentional outcomes (Levels 3 and 4), 8-year-olds in addition differentiated Levels 4 and 5. Ten-year-olds drew a further distinction in clearly separating Levels 2 and 3, an additional distinction not made by the 12-year-old group. Finally, adults clearly distinguished all levels except the first and last pairs.

A similar main effects analysis of variance was conducted to determine the exact nature of the Story Character × Age interaction. It was found that only the youngest age group was affected by actor type in their attribution of blame. Further investigation showed that this difference was manifest at the association, causality, and justifiability levels. However, the direction of the difference was opposite to that expected; greater blame was attributed in the self- than in the other condition.

Measure of Perceived Causality

The mean scores for attribution of causality appear in Figures 3 and 4. As displayed in these figures, the pattern of perceived causality confirms the predicted effects for subjects' age, stimulus levels, and their interaction. A 5 (age) × 6 (levels) × 2 (actor type) analysis of variance yielded a significant main effect for age, $F(4, 230) = 5.03$, $p < .01$; stimulus level, $F(5, 1150) = 184.51$, $p < .001$; and the interaction of these two factors, $F(20, 1150) = 3.25$, $p < .01$. Actor condition had no effect in this analysis.

Again, simple main effects analyses of variance for subjects' causal attributions were performed to investigate in more detail the relationship between subjects' age and the stimulus level factors. One predicted feature of this relationship was that subjects' age would affect attributions at lower but not higher levels of the stimulus factor. Supporting this prediction, a simple main effect for age was found at Level 2 and Level 3 but not at Levels 4, 5, and 6.

It was predicted that a second characteristic of subjects' causal attributions would be a different effect of the stimulus level at each age. The simple main effects analysis showed that within each age group the stimulus level had a significant effect. Further analysis of the effect was made by comparing the means of different stimulus levels within each age group. The only significant differences between adjacent means in the youngest group involved Levels 1 and 2 as well as Levels 3 and 4. The latter, in addition to Levels 4 and 5, were differentiated by the 8-year-old group. Both 10- and 12-year-olds made the same two distinctions, although they also differentiated Levels 1 and 2. For the adult

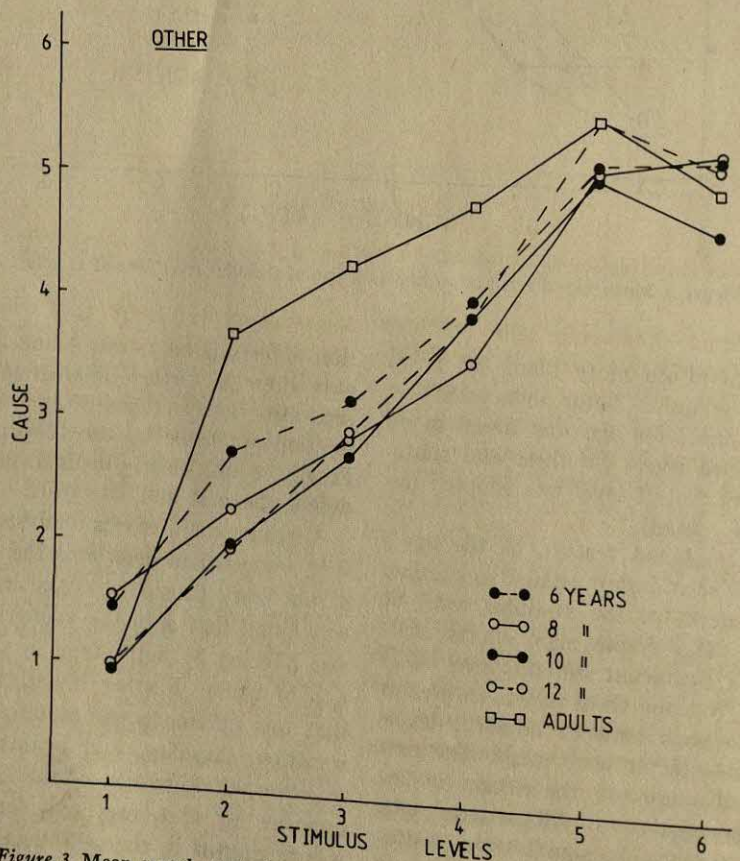


Figure 3. Mean causal attribution as a function of stimulus level for other as actor.

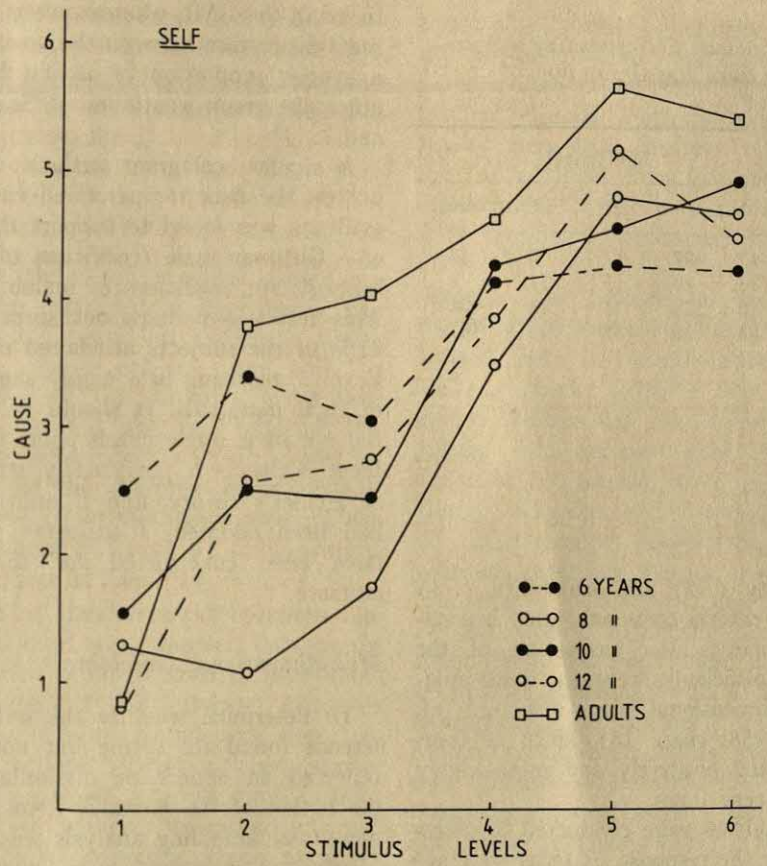


Figure 4. Mean causal attribution as a function of stimulus level for self as actor.

group, significant differences were obtained for Levels 1 and 2 and Levels 4 and 5. In sum, distinctions were mainly made between Level 1–2, 3–4, and 4–5 pairs, while no group distinguished Levels 2–3 and 5–6.

Developmental Stages in Attribution of Responsibility

A Guttman scalogram analysis was performed on the complete set of AR data gathered in the study. Although an Actor Condition × Age interaction was found, two separate analyses were not conducted for two reasons. First, the interaction obtained did not affect the order relationship between blame attributions and the various stimulus levels. Second, deterministic models for categorical data scaling tend to capitalize on chance variation, making the use of samples as small as 120 subjects a "dangerous extravagance" (Torgerson, 1958, p. 324). In addition, only Heider's original five levels were included in the analysis, since no age group distinguished Level 5 and Level 6, and only one subject attributed blame in the latter case but not the former.

An initial scalogram analysis showed that Heider's levels displayed the expected ordering but also revealed that the distribution of responses at Level 5 was skewed beyond an acceptable level. Because this is one of the auxiliary criteria that can lead to a spuriously high coefficient of reproducibility, this statistic was recalculated on the basis of Levels 1 to 4 data only (Level 5 was, however, still retained in order to distinguish the 13.3% of scale types who attributed blame at this level but not at the previous stimulus levels). The coefficient of both

Table 1
Percentage of Subjects Corresponding to Pure Scale Types in Each Age Group for Perceived Blame

Scale type	Age (years)				
	6[a]	8[b]	10[c]	12[d]	Adult[e]
5	47	27	12	8	3
4	13	10	15	14	8
3	10	23	25	22	31
2	24	25	38	42	47
1	3	12	10	14	11
0	3	3	0	0	0

[a] $n = 38$.
[b] $n = 40$.
[c] $n = 40$.
[d] $n = 36$.
[e] $n = 38$.

reproductibility (.92) and scalability (.73) exceeded the criteria commonly used in evaluating Guttman scales. Inspection of the correlation coefficients between items suggested a unidimensional latent structure (cf. Torgerson, 1958, chap. 13), since adjacent items correlated positively and more highly than nonadjacent items.

Further analyses were conducted to determine whether the response patterns were age related. The scale scores, based on all five Heiderian levels, differed in the expected manner; they were inversely related to age, $F(4, 235) = 6.32$, $p < .01$. However, multiple-means comparisons showed that the four oldest age groups did not significantly differ from each other, although they all differed from the youngest group. Table 1 shows the age distribution of the 86.25% of responses conforming to perfect scale types. From the data displayed in this table it can be seen that the age effect may possibly derive from younger subjects' dominance at Scale Type 5. This tendency is in part due to bias introduced by the self-as-actor manipulation, and hence scale scores for the other-as-actor condition were analyzed. No significant effect for age was found, $F(4, 115) = 1.58$, $p > .10$. Moreover, inspection of the age-by-scale-type distribution for this data showed no strongly discriminable age-related pattern except that (a) some younger children were undifferentiated in their AR, whereas every adult made some distinction between the levels; and (b) a greater proportion of adults than of any other age group was found at Scale Types 2 and 3.

A similar scalogram analysis was carried out on the data for perceived causality. No evidence was found to support the existence of a Guttman scale (coefficient of reproducibility $< .90$; coefficient of scalability $< .60$). This result is perhaps not surprising, since 81% of the subjects attributed causality by Level 2, resulting in a highly skewed spread of item marginals. It should be noted that the use of a single cutoff point for building these scales is a particularly stringent test of Heider's theory, and if multiple criteria had been adopted, a scalogram might also have been constructed for the causality measure.

Multidimensional Unfolding

To determine whether the self/other difference found for blame but not for cause reflected an underlying dissimilarity in the perception of the stimulus items, a multidimensional unfolding analysis was attempted. Degenerate solutions were repeatedly obtained owing to a lack of constraint in the data. To increase response variability, adult and child data were combined, but again, unstable solutions emerged. Inspection of the raw data showed that there was much consensus in the response patterns across the various ages (e.g., no subject showed a systematic decrease in AR with increasing level), and hence this analysis was abandoned.

Attribution to Intentional Versus Unintentional Behavior

In terms of Piaget's (1932) theory, the major moral development in middle childhood involves a shift from outcome-based to intention-based judgments. Comparisons were thus made between Level 2 (an uncoerced unintentional act) and Level 5 (an uncoerced intentional act) to determine whether subjects differentiated intentional from unintentional events in their attributions. These comparisons were made for both attribution

measures and within each age group. Significant differences were found on both variables for all the groups ($p < .05$), indicating an emergence of this change earlier than is traditionally expected.

Discussion

In general the results provided strong support for the experimental hypotheses. For both moral and causal attributions a strong interaction between subjects' age and the characteristics of the stimulus event was found. As predicted, the data reflected significant age differences in mean scores at lower levels of the stimulus factor but not at higher levels (for the moral measure the convergence was at Level 5; for perceived causality, it was at Level 4).

The form of the interaction between age and stimulus level was, however, only partly consistent with a developmental interpretation of Heider's (1958) theory. Although increasing age was associated with more significant distinctions between adjacent stimulus levels, this result requires some qualification. Almost without exception all age groups attributed increasing blame as actor behavior became more internally directed. Hence even the youngest children appeared on the average to show distinctions in the hypothesized direction between Heider's criteria for AR. With development these distinctions became more marked. Sedlak's (Sedlak, 1979) finding with respect to the age-related importance of Heider's criteria in cognitive representations directly supports the present results.

In addition to the above, it was found that even the 8-year-old group distinguished between intentionality and justifiability, the second significant differentiation to emerge with age. This contrasts with Shaw and Sulzer's (1964) study, where only adults made a slight distinction between these two criteria, leading to the suggestion that they may only be fully appreciated by "mature" senior citizens. The early emergence of this sophisticated attribution process is perhaps not surprising for three reasons. First, the justifiability level was operationalized in terms of a familiar, commonly occurring situation for children (self-defense), whereas parents often model AR judgments on justification. Second, outcome valence and intensity were held constant within the context of a single story presentation. Third, development of AR may be fairly rapid in such familiar interpersonal situations with undesirable outcomes; casual observation suggests that explanations are more often sought after something bad has happened. Indeed there is recent, though admittedly limited, support to show that even kindergarten children can take mitigating circumstances into account when evaluating an aggressor (cf. Karniol, 1978, p. 84), whereas Darley, Klossom, and Zanna (1978) found no age differences between 6-, 9-, and 24-year-olds regarding the use of contextual information in evaluating intentional behavior. In any event, significant distinctions between these two criteria did not emerge in a sequence consistent with a developmental interpretation of Heider's (1958) levels. Moreover, not even adults made a significant distinction between the theoretically easiest level pair (Levels 1 and 2), a result also found by Harris (1977). It would thus seem that the difference between actor-caused accidental events (Level 2) and accidents not produced by actors (Level 1) may not be as important in AR as other Heiderian criteria.

Finally, a scalogram analysis based not on the intensity of responsibility attributed but rather on its absolute ascription showed that Heider's levels formed a Guttman scale. Within the age range tested, only limited and rather ambiguous support was found for age-related response patterns. Under conditions that did not prejudice the youngest group's attributions, no age differences were found in scale scores. However, inspection of the data conforming to scale types did show that undifferentiated responses were exclusive to the child sample and were especially evident in younger children. In addition, adults tended to dominate in the groups that based AR on intentional but justifiable acts and on foreseeable actor-produced accidents. Since these results reflect individual response pat-

terns rather than between-subjects mean scores, they may prove to be a more appropriate test of the developmental support for Heider's theory. In the absence of previous research utilizing such an approach, this result requires replication before the developmental status of Heiderian criteria for attributing responsibility can be fully evaluated.

It should be noted that although Heider's attribution criteria were strongly confirmed, no evidence was obtained in this study to support the additional level hypothesized. Nonetheless, abandoning this criterion is felt to be premature for two reasons. First, it seems to be an extension logically implied by Heider's model. Second, it may not have been appropriately tested in the present circumstances. Because of a medium-intensity negative outcome, a maximal environmental force against the action may already have been broken at the intentionality level leading to the ceiling effect observed. This could be the case for all negative-outcome events, and hence the distinction may only emerge under neutral- or positive-outcome conditions.

To test this hypothesis the same within-subjects Latin square design was used to administer stories representing both positive and negative outcomes at Levels 4, 5, and 6 to 45 adults. Although no substantial evidence was obtained to support Level 6 (differences obtained were, however, in the predicted direction), the interaction between level and outcome was mainly owing to the extreme evaluations of negative outcomes at Levels 5 and 6. In fact subjects estimated that very few persons would have acted as the story character did at these levels (8% and 9%, respectively), supporting the idea of a ceiling effect for internal causation. In contrast, it was held that the behavior of large percentages of people would be similar to that of the story character for positive outcomes (Level 5 = 55%, Level 6 = 42%). These estimates of the action occurring given the situational circumstances suggest that Level 6 was not adequately operationlized to produce maximal internal causality in this condition. It appears that while an extension of Heider's scheme for negative outcomes may be questionable, its utility for positive events remains to be adequately evaluated.

A third aim of the present study was to investigate attribution differences relating to the other and the self as actor. As predicted, this variable interacted with age, but the direction of the difference was contrary to that expected. Inspection of the raw data showed that almost half of the children attributed maximum blame to themselves regardless of level, a phenomenon not found in the other-as-actor condition. In view of this anomaly, the study was repeated on an independent sample of 40 6-year-olds. Since it was suspected that the previous results might reflect a conservative simplifying strategy adopted in the presence of any uncertainty regarding the stimulus story conditions, children were individually tested, while finger puppets were used to further clarify the story presentation. Although the general pattern of increasing AR with level was found, no self/other differences emerged. This result is consistent with Keasey's (1977) finding that children in kindergarten but not in first grade show greater use of intention information in relation to the self than in relation to a hypothetical other. It thus appears that the difference initially found is an artifact that, as mentioned, may merely indicate a simplifying strategy.

A further purpose of the study related to differences between two measures of subjects' attribution, which are often assumed to represent a single entity. Strong evidence to support the hypothesized distinction was found. Whereas all but one group correctly distinguished association and mere causality in terms of perceived causality, no group made this distinction in attributing blame. In addition, responses to perceived blame formed a Guttman scalogram, whereas perceived causality did not. These two findings represent an interesting qualification to Heider's (1958) theory, because both measures have been cited as indications of responsibility (e.g., Harris, 1977).

The utility of these two attribution measures for developmental research was also

illustrated. While adult distinctions regarding blame were achieved by 10 years of age, no child group showed the adult response pattern for causality attribution. Children tended to perceive differences in causality where adults did not, suggesting that causality and responsibility were not as clearly distinguished. Furthermore, increasing age was associated with the emergence of a clear pattern in the use of the two attribution measures. Whereas younger children sometimes ascribed greater blame than causality, adults rarely did so (6 years = 28% of total responses, 8 years = 24%, 10 years = 20%, 12 years = 17%, and adults = 10%). Any response order effect due to regression would not alter the above result, which is consistent with the age-related changes in the data curves for both dependent measures and suggests a similar increasing cognitive differentiation between perceived blame and causality with age.

The results of this study also support an emerging belief that even young children differentiate intentional from unintentional behavior and are capable of more subtle distinctions in their attributions than are traditionally ascribed to them. That children as young as 6 years clearly distinguished accidental and intentional outcomes is consistent with recent research (e.g., Farnill, 1974) showing that children do use intentionality information when tested under appropriate conditions that do not overburden their cognitive processing capacities.

In sum, the present study provides some support for a developmental interpretation of Heider's (1958) levels of AR. When quantitative group scores were considered, all ages made theoretically consistent distinctions between the levels, although a greater number of significant differences was associated with increasing age. Similarly, certain response patterns dominated at particular ages, but no evidence was found for developmental stages. Insofar as developmental differences do exist, a potentially fruitful question for future research to answer is whether they represent age-related variations in evaluative criteria or social perception.

Reference Notes

1. Raymond, J. S., & Dillehay, R. C. *Perceived effort and incidental outcome as determinants of moral judgment in children.* Paper presented at the meeting of the Midwestern Psychological Association, Chicago, May 1973.
2. Sedlak, A. J., Thompson, V. D., & Sands, R. E., with Teddie, C., Hudson, S., Jaffe, J. & de Carufel, A. *Effects of outcome, motive and intentionality on children's moral judgments and lie decisions.* Manuscript submitted for publication, 1978.
3. Sulzer, J. L. *Heider's "levels model" of responsibility attribution.* Paper presented at the Symposium on Attribution of Responsibility Research, Williamsburg, Virginia, July, 1971.

References

Armsby, R. E. A re-examination of the development of moral judgment in children. *Child Development,* 1971, *42,* 1241–1248.

Darley, J. M., Klosson, E. C., & Zanna, M. P. Intentions and their contexts in the moral judgments of children and adults. *Child Development,* 1978, *49,* 66–74.

Farnill, D. The effects of social judgment set on children's use of intent information. *Journal of Personality,* 1974, *42,* 276–289.

Fishbein, M., & Ajzen, I. Attribution of responsibility: A theoretical note. *Journal of Experimental Social Psychology,* 1973, *9,* 148–153.

Gutkin, D. C. The effect of systematic story changes on intentionality in children's moral judgments. *Child Development,* 1972, *43,* 187–195.

Harris, B. Developmental differences in the attribution of responsibility. *Developmental Psychology,* 1977, *13,* 257–265.

Hebble, P. W. The development of elementary school children's judgment of intent. *Child Development,* 1971, *42,* 1202–1215.

Heider, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

Karniol, R. Children's use of intention cues in evaluating behavior. *Psychological Bulletin,* 1978, *85,* 76–85.

Keasey, C. B. Young children's attributions of intentionality to themselves and others. *Child Development,* 1977, *48,* 261–264.

Kelley, H. H. *Attribution in social interaction.* Morristown, N.J.: General Learning Press, 1971.

Kirk, R. E. Experimental design: *Procedures for the behavioral sciences.* Belmont, Calif.: Wadsworth, 1968.

Piaget, J. *The moral judgment of the child.* New York: Harcourt Brace, 1932.

Sedlak, A. J. Developmental differences in understanding plans and evaluating actors. *Child Development,* 1979, *50,* 536–560.

Shaver, K. G. *An introduction to attribution processes.* Cambridge, Mass.: Winthrop, 1975.

Shaw, M. E., & Iwawaki, S. Attribution of responsibility by Japanese and Americans as a function of age. *Journal of Cross-Cultural Psychology,* 1972, *3,* 71–81.

Shaw, M. E., & Reitan, H. T. Attribution of responsibility as a basis for sanctioning behavior. *British Journal of Social and Clinical Psychology,* 1969, *8,* 217–226.

Shaw, M. E., & Schneider, F. W. Negro–white differences in attribution of responsibility as a function of age. *Psychonomic Science,* 1969, *16,* 289–291.

Shaw, M. E., & Sulzer, J. L. An empirical test of Heider's levels of attribution of responsibility as a function of age. *Journal of Abnormal and Social Psychology,* 1964, *69,* 39–46.

Sulzer, J. L., & Burglass, K. B. Responsibility attribution, empathy and punitiveness. *Journal of Personality,* 1968, *36,* 272–282.

Torgerson, W. S. *Theory and methods of scaling.* New York: Wiley, 1958.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1971.

Received October 16, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 9, 1603-1616

Multiple Discovery and Invention: Zeitgeist, Genius, or Chance?

Dean Keith Simonton
University of California, Davis

The occurrence of independent contributions by two or more scientists can be interpreted in terms of zeitgeist, genius, or chance. The relative adequacy of these three theories was examined by hypothesizing four critical empirical tests. These tests focus on (a) the general and intradisciplinary probability distribution of multiples and (b) the relationship of individual eminence with multiple production and priority. An analysis of 579 multiples and of 789 scientists and inventors lends the most support to the chance theory, followed by the zeitgeist theory. The results are integrated into a single probabilistic perspective that incorporates some of the major features of all three theories.

What mathematician does not know that the calculus was independently devised by both Newton and Leibniz? Who among biologists has not heard that Darwin's laborious documentation of evolutionary theory was almost forestalled by an independent abstract by Wallace? And what psychologist has not observed to his or her students that the James–Lange theory of emotion was actually advanced independently by James and Lange? Certainly the occurrence of multiple discoveries constitutes one of the most fascinating phenomena in the history of science and technology. Such multiples become particularly dramatic when they provoke fierce priority disputes such as the debates that surrounded the independent prediction of the planet Neptune by Adams and Leverrier. But why do multiple discoveries and inventions happen at all?

The traditional explanation of multiples is founded on what can be called the *zeitgeist* theory of creativity. According to this social deterministic view, the individual creator is largely irrelevant or epiphenomenal to the cultural progress represented by the inevitable accumulation of scientific knowledge and technological expertise. Rather, it is the sociocultural system as a whole, embodied as the spirit of the times, which is ultimately responsible for any given technoscientific advance. Thus, if neither Adams nor Leverrier had predicted the existence of Neptune, someone else would certainly have done so. The failure of gravitational astronomy to explain the anomalies in the orbit of Uranus became so obvious and irksome that Neptune's forecast was in the air. Because the times were ripe for the prediction of the eighth planet, it was really immaterial who actually picked the fruit; any competent astronomer might have done the same, had Adams and Leverrier both died in the crib. Not surprisingly, the zeitgeist account of multiples is frequently defended by those anthropologists and sociologists who desire to minimize the role of the individual as a causal agent of sociocultural events. For example, in an article entitled "Are Inventions Inevitable?" Ogburn and Thomas (1922) list nearly 150 multiples to prove that any given discovery or invention has a probability near unity of appearing, the presence or absence of any specific genius notwithstanding. And Merton (1961) takes this interpretation so much as given that he once asserted, "It is the singletons—discoveries made only once in the history of science—that are the residual cases, requiring special explanation" (p. 477). But psycholo-

This research was made possible by a Faculty Research Grant from the University of California.

Requests for reprints should be sent to Dean Keith Simonton, Department of Psychology, University of California, Davis, California 95616.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3709-1603$00.75

gists are not always ill-disposed to the zeitgeist interpretation of multiples, as we can readily witness in Boring's (1963) citation of multiples as evidence that the "genius" is merely an agent of the zeitgeist, an effect or a symptom rather than the cause of the times.

Still, there is no doubt that not all psychologists would completely endorse the zeitgeist position. The assumption implicit in the research on creativity in the 1950s and 1960s was that substantial and sustained individual differences in creativity determine whether any given person can make a contribution to society (e.g., Taylor & Barron, 1963). In short, many psychologists would feel much more comfortable with what can be styled the *genius* theory of creativity. According to this perspective, scientific discoveries and technological inventions are produced by great scientists and inventors who possess abilities, personalities, and backgrounds that set them apart from their colleagues. On first blush, it would seem that the genius theory could not adequately explain the appearance of independent contributions, but this supposed inadequacy is only partly true. In fact, genius theory can explain one aspect of multiples that must prove an embarrassment to zeitgeist theory—the frequent occurrence of rediscoveries. Not all multiples are produced simultaneously. The various independent contributions are sometimes separated by long time spans during which the first invention or discovery passes into temporary oblivion. Merton (1961) has reported that even though 20% of the multiples he studied occurred within a 1-year span, some 34% entailed a time lapse of 10 or more years. Classic instances of rediscoveries include Mendel's formulation of the laws of heredity, some of Cavendish's unpublished electrical discoveries, much of Gauss's unpublished mathematical work, and the prodigious notebooks of one of the greatest anticipators of all, Leonardo da Vinci. How can someone like Mendel come to be 35 years ahead of his time? Genius theory has less difficulty on this point: A major genius is defined as one who transcends the limitations of the zeitgeist and one who may even anticipate movements far into the future. Long delayed rediscoveries can be adopted, therefore, as evidence for genius theory. For example, even if Stevinus merely rediscovered hydrostatic principles already known to Archimedes, the fact would remain that it took almost 2,000 years for someone to be born who could replicate Archimedes' achievement.

Besides the genius and zeitgeist theories, a third position is a credible contender for understanding multiples: *chance* theory. Since most behavioral scientists seem to go out of their way to disprove the null hypothesis predicated on the random model (Greenwald, 1975), it is perhaps unsurprising that this theory has not always received the study it deserves. Nonetheless, chance or "luck" definitely plays some part in creativity, as the phenomenon of serendipity well illustrates (Cannon, 1940). In addition, probabilistic or stochastic models have been so successfully applied to other facets of creativity that there is no reason not to approach multiples in the same manner. To illustrate, probability models have explained individual differences in creative productivity (Price, 1976; Simon, 1955), the correlation between productivity and eminence (Dennis, 1954), and the correlation between productivity and age (Dennis, 1966; Simonton, 1977a). More specifically, there have been several investigators who have proposed some kind of probability model for multiples (e.g., Price, 1963, chap. 3; Schmookler, 1966, chap. 10; Simonton, 1978). As an example, many adherents of the zeitgeist perspective believe that the near simultaneity of many inventions or discoveries precludes explanations based on mere chance (e.g., Kroeber, 1917). Such multiples as the liquification of oxygen by Pictet and Cailletet, the discovery of the stereochemistry of carbon by Van't Hoff and Le Bel, the electrolytic process for aluminum manufacture of Hall and Heroult, and even the rediscovery of Mendelian genetics by De Vries, Correns, and Tschermak all took place not simply within a year but often within a month or so. For chance theory to handle this simultaneity, we have only to assume that once a given contribution is made, the probability of a duplication declines rapidly due to communication termi-

nating any research that might head in the same direction.

Each of the foregoing three theories can be offered as partial interpretations of multiples in science and technology. But which theory best explains the data? In this article I propose to examine each of the three principal theories so as to derive a set of distinctive empirical predictions that can serve as critical tests. In deriving these distinguishing repercussions of the three theories, I will treat each position as a pure or ideal type. In other words, I will try to make each theory as distinctive as possible so that the contrasting empirical implications can best be highlighted. Only at the close of this article, after the results of the tests are known, will I attempt to synthesize the three theories into a single view. With this qualification in mind, we can advance the following four tests.

Test 1

If the zeitgeist theory holds, then multiples should be very common relative to singletons, and the distribution of the various grades of multiples (i.e., doublets, triplets, quadruplets, quintuplets, etc.) should be approximated by a binomial distribution with a mean appreciably greater than 1. If the genius theory holds, then multiples should be virtually nonexistent, yielding a degenerate binomial with a mean of 1 and a variance of 0. If the chance theory holds, then multiples should be rare relative to singletons, and the distribution of the various grades of multiples should be approximated by a Poisson distribution with a mean less than 1.

Let us suppose that for any given invention or discovery there are n individuals capable of producing it, and each has a probability p of success, where $n \geq 1$ and $0 < p \leq 1$. Then, as we know from any introductory statistics text (e.g., Snedecor & Cochran, 1967, chap. 8), we can employ the binomial distribution to obtain the proportion of the various numbers of successes. Now suppose we apply this simple probability model to the three theories. If the zeitgeist theory is valid, then the number of

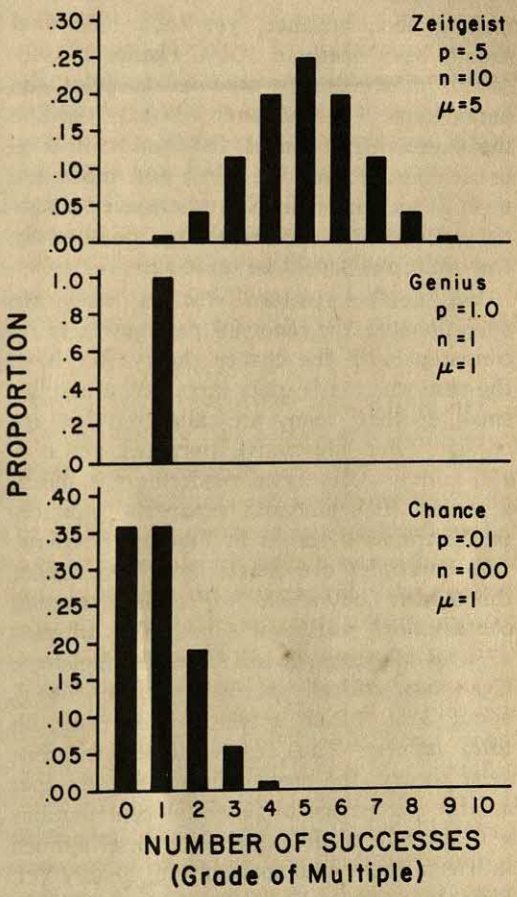


Figure 1. Predicted probability distributions obtained by interpreting the three alternative theories in terms of the binomial distribution with different values for n and p.

creators should be very large and the probability of success somewhat small, so that genius per se is irrelevant. Assume for the sake of argument that $p = .5$ and $n = 10$, that is, that for any potential contribution there are 10 people capable of making it, although each enjoys only a 50–50 chance of success. Then we should obtain a binomial distribution with a mean of $np = 5$ and with the shape shown in Figure 1. Under this distribution almost 99% of all contributions to science and technology will involve two or more independent creators, and the modal multiple will be of Grade 5 (i.e., quintuplets). Hence, inventions and discoveries must be virtually inevitable.

If the genius theory is correct, there should be only a small number of people capable of

making each advance, yet each individual should have nearly a 100% chance of success. For argument's sake let us posit the parameters of $n = 1$ and $p = 1.0$, yielding the degenerate binomial distribution in Figure 1. Notice that the mean and mode are now 1, with no multiples whatsoever occurring. Hence the genius theory does imply that multiples should be quite rare.

Yet another picture emerges when we operationalize the binomial parameters to be consistent with the chance theory. We have the case where n is quite large but p is quite small, so that "many are called but few are chosen." For illustrative purposes, let $n = 100$ and $p = .01$, again rendering the mean, μ, unity. The binomial expansion produces the distribution shown in Figure 1. Observe that multiples are much more rare under this model: Only about 26% of all attempted contributions will involve multiples, whereas 37% of all attempts will produce singletons. Even more critically, according to this model, about 37% of all attempts will result in total failures. Thus, quite unlike the zeitgeist theory, the chance theory implies that a large percentage of potential contributions will never be made and thereby inserts much indeterminancy into sociocultural history (cf. Simonton, 1978). Furthermore, if the chance theory is true, then we may go one step further and assert that the frequency distribution of multiples of various grades will be closely approximated by a Poisson distribution. The latter is the exponential limit of the binomial distribution when n is very large and p is very small (Haight, 1967, p. 15). For example, the correlation between proportions derived from a Poisson distribution with $\mu = 1$ and those from a binomial distribution with $n = 100$ and $p = .01$ departs from unity only after the fourth decimal place. Also, since the mean and variance are equal in the Poisson case, the Poisson distribution is easier to estimate from limited data, a valuable asset, given that we do not know how many total failures there have been in the history of science and technology (i.e., how many "nulltons"). For this first test the Poisson distribution will therefore be adopted as the hypothesis to be disproven. Only if we are empirically compelled to abandon the Poisson model with a mean less than 1 will we feel obliged to examine the two alternative theories.[1]

Test 2

According to the zeitgeist theory, the probability of multiples occurring should be highest in the mathematical and physical sciences and lowest in the biological sciences. If the genius theory is correct, the reverse should be true. But by the chance theory, the probability that multiples will occur should be constant across scientific disciplines.

Some scientific endeavors can be said to engage more zeitgeist than others. As Kuhn (1970) pointed out, the physical sciences often have well-developed paradigms that logically generate the puzzle-solving research of "normal science." Thus the problems of classical physics were largely defined in terms of the repercussions of Newtonian mechanics and Maxwellian electromagnetic theory. Other, "softer" sciences do not feature such clean theoretical foundations, and consequently the zeitgeist is more diffuse. Hence we would expect the more codified and mathematically constructed disciplines to generate a disproportionate number of multiples. The prediction of genius theory goes almost in the opposite direction. Genius theory maintains that the major figures of human history tend to transcend the spirit of their times. Indeed, a study of more than 2,000 thinkers in Western civilization has

[1] Actually, we could argue that a more reasonable probability distribution would be the negative binomial (Simonton, 1978). This "contagious" or "compound" Poisson allows the parameter $\mu = np$ to be described by the gamma distribution, and accordingly the probability that a contribution will appear does not have to be a constant (Fisz, 1963, pp. 164–168). Yet if the chance theory holds, we must recognize that the Poisson distribution is also the exponential limit of the negative binomial, and consequently we do not gain additional predictive power (Feller, 1968, p. 281). In fact, there would be a loss due to the inability to apply the usual estimation techniques for the truncated negative binomial to the current data set (e.g., Brass, 1958).

shown that there is a negative relationship between the fame of a thinker and the extent to which he or she fits the prevailing beliefs of the time (Simonton, 1976d). Accordingly, those disciplines that attract the most geniuses should have a smaller proportion of multiples. But how do we know which scientific enterprises attract the greatest minds? Roe (1952) found in her study of 64 eminent scientists that physicists tend to score higher on standard intelligence tests than do biologists. Similarly, Harmon's (1961) survey of doctorates discovered that physicists and mathematicians had higher intelligence quotients than biologists, with chemists and engineers falling in between. Therefore, if we are willing to grant two crucial assumptions, we would predict in accordance with Test 2. The first assumption is that the distribution of intelligence in contemporary science can be projected a few centuries into the past. Partial support for this extrapolation can be gleaned from Cox's (1926) IQ estimates for select scientists born as far back as the fifteenth century: Mathematicians have an average IQ of 183, physicists 178, medical researchers 172, chemists 172, biologists 171, and astronomers 170. The second assumption is that psychometric IQ has something to do with cultural eminence. While classic studies of contemporary populations by Terman and Oden (1947) and of historical populations by Cox (1926) have attempted to demonstrate this relationship, the evidence is far from secure. For instance, the correlation of .26 that Cox found between eminence and IQ has been shown to be a methodological artifact (Simonton, 1976a). Nevertheless, without this assumption we cannot derive a unique prediction for the genius theory. Finally, turning to the chance theory, we see that the random model is the simplest, for it merely affirms that the probability of the coincidences labeled as multiples is uniform across all disciplines. A recent reanalysis of data published by Ogburn and Thomas (1922) seems to endorse this latter prediction (Simonton, 1978). Nonetheless, the sample size may have been too small to permit rejection of the null hypothesis.

Test 3

The zeitgeist theory implies that there should be a positive relationship between the eminence of a given scientist or inventor and the number of multiple discoveries or inventions he or she participates in. In contrast, the genius theory implies that the relationship between eminence and multiples participation should be negative. The chance theory predicts that this correlation should be zero once variations in productivity are controlled.

Boring (1963) believed that certain historical figures are singled out as "eponyms" because they epitomize the times in which they live. If multiples are also interpreted as zeitgeist-inspired events, then it follows that the eponyms of history should engage in more multiples than those outcasts from the mainstream of history. Thus Merton (1961) has maintained that persons of "great scientific genius will have been repeatedly involved in multiples" (p. 484). The genius theory would advocate the exact opposite. To employ Kuhn's (1970) terminology again, the lesser figures may be wrapped up in the puzzle-solving activities dictated by the predominant paradigms of normal science, but the principal figures are generating influential scientific revolutions that will change the zeitgeist. On the other hand, chance theory takes a stance somewhere between these extremes. As Merton (1961) noted, the more eminent scientists produce more multiples in part "because the genius will have made many scientific discoveries altogether" (p. 484). So just by chance the more productive scientist will create more multiples. Yet creative productivity has been repeatedly shown to be correlated with eminence (e.g., Dennis, 1954; Simonton, 1976b). Hence, there may very well be a positive correlation between eminence and participation in multiples, but that correlation may be the spurious result of a common cause, creative productivity. Therefore, if we partial out the source of spuriousness, the relationship should vanish, at least according to chance theory.

Test 4

If the genius theory is justified, there should be a positive relationship between individual eminence and temporal priority in a multiple. If either the zeitgeist or the chance theory holds, there should be no relationship, positive or negative.

As already noted, the genius not only is independent of the zeitgeist but may also be a precursor of the future zeitgeist. Hence, the more eminent scientist produces inventions or discoveries far in advance of his or her time, contributions that lesser minds may take decades if not centuries to rediscover. Unfortunately for this prediction, past research is not in complete accord with the precursive conception of genius. The study of more than 2,000 thinkers mentioned earlier found that the more eminent thinkers were actually behind their times, as if they were more interested in consolidating the past than in intimating the future (Simonton, 1976d). If this result applies to science and technology too, then the major figures may be those who rediscover the work of lesser known predecessors. In any case, we would not expect any consistent relation between priority and eminence, should the zeitgeist or chance theories prove valid.

Quite clearly the above four tests do not all have the same secure foundation. The conceptual basis for Tests 1 and 3 is perhaps more firm than for Tests 2 and 4. Even so, something is definitely gained by stipulating both strong and weak criteria for the three theories. Although the three approaches do not always imply three separate predictions for each test, each of the theories is uniquely characterized when all four tests are taken together. If both weak and strong tests point in the same direction, then the preferred theoretical explanation can be better identified. In addition, the outcome of all four tests may provide the basis for a synthesis of the three theories into a single, more comprehensive view.

Empirical Tests

How Probable Are Multiples?

Clearly our first task is to obtain an extensive list of multiples. To do this I began with the previously published lists of Ogburn and Thomas (1922), Kroeber (1963), and Stern (1927, chap. 3).[2] This preliminary list was then augmented by consulting a large number of additional sources, especially Darmstaedter (1908), Taton (1964, 1965, 1966), Asimov (1972), and Williams (1974). I also attempted to verify all purported multiples to make sure they actually involved independent efforts of two or more researchers (using, e.g., Debus, 1968; *Encyclopædia Britannica*, 1974; Gillispie, 1970; Singer, Holmyard, Hall, & Williams, 1958). A critical theoretical issue arose at this verification stage: The case for the occurrence of multiples is not as strong as one is led to believe by zeitgeist enthusiasts. One problem is that many putative instances rest on vague generic terms that may encompass far too much (Schmookler, 1966, chap. 10). For instance, Ogburn and Thomas (1922) credit both Boerhaave and Hales with the independent contribution of the "beginnings of modern organic chemistry" (p. 98), an extremely general attribution, to be sure. Another difficulty is that even more specific terms, such as the "microphone," may cover inventions or discoveries that are often comparable only regarding function, each contribution employing quite different means to attain roughly the same end. Indeed, it may even be argued that true multiples are virtually nonexistent: The calculus of Newton was not the same as that of Leibniz, nor was Priestley's oxygen generated the same way as Scheele's. Even the eighth planet differed from Adams to Leverrier, since each predicted different orbits for Neptune (both of which just happened to coincide fairly well in 1846). Yet even granting some abstraction in categorization, many multiples still prove to be mirages upon closer inspection. Often what is described as independent work is not independent at all. Thus Ogburn and Thomas (1922) should not list both Henry and Morse as inventors of the telegraph, for it was Henry's freely given advice that had allowed a painter to succeed. Similarly, Morton learned of the anesthetic properties of ether

[2] I attempted to obtain Merton's (1961) sample as well, but his data are no longer available (Merton, Note 1).

from Jackson, and Fulton well knew Jouffroy's work on the steamboat and actually attended the launch of Symington's steamboat (cf. Ogburn & Thomas, 1922). In short, many multiples do not involve independent efforts but rather consecutive developments in the evolution of science and technology (cf. Constant, 1978). A final problem is that a multiple mentioned in one source might not be mentioned in any other source and hence would stand unconfirmed.

To deal with the foregoing problems, I chose to collect two separate samples of multiples (excluding cases prior to 1500) from the general fields of mathematics, astronomy, physics, chemistry, biology, medicine, and technology. The first, "exclusive" sample included only those multiples mentioned in two separate sources, without contradiction by any other source. The second, "inclusive" sample consisted of all multiples mentioned in at least one source as long as no contradictory evidence could be found in any other source. The exclusive sample contained 199 multiples, more than the 148 collected by Ogburn and Thomas (1922) but less than the 264 gathered by Merton (1961). The more liberal definition added another 380 multiples to produce the 579 multiples of the inclusive sample. All principal hypotheses were then tested on *both* samples. Fortunately, the substantive conclusions do not materially differ for the two samples, and I will therefore focus on the inclusive sample (since a larger sample size permits parameter estimates with smaller standard errors).

General probability (Test 1). One secure finding is that multiples are extremely rare events. It may sound impressive to assert the existence of 579 potential multiples, but the number of so-called singletons is immensely larger. To illustrate this point, we may note that Darmstaedter's (1908) chronology, probably the most comprehensive in any Western language, lists 12,300 contributions in the seven fields represented in our sample (Sorokin, 1957, Table 5). So at best multiples constitute a mere 5% of all contributions. I say "at best" because the Darmstaedter chronology is evidently more selective than the undefined sample from which the multiples were apparently derived; most of the purported multiples were not sufficiently important to be cited in that source. Consequently, the zeitgeist theory seems disproven: As we can conclude from Figure 1, singletons are supposed to be far less common than multiples if the zeitgeist theory holds. In fact, the results thus far appear to partly bolster the genius theory.

An advocate of the zeitgeist position, however, may claim that it is most unfair to compare singletons to multiples. Merton (1961) presented 10 reasons why he believes that multiples are considerably underestimated and that many singletons may actually be multiples incognito. On first glance, this accusation, if true, would seem to invalidate any empirical investigation and would thereby transport the whole issue to the realm of armchair social philosophy. But inquiry can be salvaged if we ignore the singletons altogether and instead concentrate on the probability distribution of multiples of various grades (Simonton, 1978). We have only to assume that the relative proportions of doublets, triplets, quadruplets, and so on remain constant after being sifted through the selection process. If there are twice as many doublets as triplets in the real world (i.e., the statistical population), there will presumably be twice as many doublets as triplets in any sample, once we allow for sampling error. Given this standard assumption of statistical sampling, we can then try to fit a truncated Poisson distribution using the two-moments method of estimating the mean, μ, and the standard error, *SE* (Haight, 1967, p. 89; Patil, 1962). The result is displayed in Table 1, where I give the observed frequency of multiples of Grades 2 through 8 along with the fitted values according to a Poisson model (using the tables in Molina, 1942).

The fit is fairly close. A goodness-of-fit test gives $\chi^2(3) = 6.07$, $p > .1$ that any discrepancies are due to mere chance. Although this test does not strictly prove the chance theory, we can be very confident that both the zeitgeist and the genius theories are disconfirmed. This conclusion is reinforced when we inspect the estimated mean, which is less than 1. This mean implies that contributions are made less than once, on the average. Accordingly, a large proportion of potential

Table 1
Observed Frequencies of 579 Multiples and Frequencies Predicted by a Poisson Model With $\mu = .82$

Multiple grade	Observed frequency	Predicted frequency
0	—	1361
1	—	1088
2	449	435
3	104	116
4	18	23
5	7	4
6	0	0
7	0	0
8	1	0

Note. Poisson parameters were estimated using the two-moments method. The standard error of the estimate is .12.

contributions never appeared at all. Given a mean of about .8 we would expect, by the Poisson Model, that some 45% of all potential contributions would become nondiscoveries or noninventions, about 36% becoming singletons, and the remaining 19% multiples. Given this outcome, scientific or technological advance is not very inevitable, to say the least.

I hasten to point out that it would be very difficult for zeitgeist theory to save the day by simply finding more multiples. It is my experience after two years of collecting multiples that the more multiples one teases out of history, the worse the picture becomes for any social deterministic outlook. What is required is more high-grade multiples—more triplets on up—in order to give the distribution a larger mean, less right skewness, and an approximately inverted U shape instead of the monotonically decreasing form. But that is not what happens. The quest for more multiples tends to dig up more doublets, rendering the case for the chance model stronger rather than weaker. Thus if we examine just the exclusive sample of 199 multiples, we find that there are 144 doublets, 42 triplets, 9 quadruplets, 3 quintuplets, and 1 octuplet. The two-moments estimate of the mean is 1.06 with a standard error of 1.33, $\chi^2(3) = 3.37$, $p > .25$. Even though this mean falls within the interval estimate for the inclusive sample, it still is larger. Hence the addition of more multiples actually lowered the mean, making the distribution more skewed right by the disproportionate addition of doublets.

I should also observe that the Poisson distribution proves to be an acceptable fit only if the singletons are excluded. Otherwise, I know of no probability model that will fit the data. We therefore have two choices. On the one hand, we can accept the observed relative frequencies of singletons and multiples as being realistic, in which case the genius theory is favored. On the other hand, we can ignore the singletons and look only at the probability distribution of multiples, in which case the chance theory is favored. The zeitgeist theory could not explain these results without adding some post hoc interpretation.

Probabilities within disciplines (Test 2). If we subdivide the multiples into the seven fields, we discover that there are 53 multiples in mathematics, 28 in astronomy, 115 in physics, 78 in chemistry, 33 in biology, 184 in medicine, and 88 in technology. By comparison, Darmstaedter (1908) lists 329 contributions in mathematics, 478 in astronomy, 1,511 in physics, 2,469 in chemistry, 1,415 in biology, 1,268 in medicine, and 4,830 in technology (Sorokin, 1957, Table 5). Hence, as an approximation we may assert that 16% of mathematics, 8% of astronomy, 8% of physics, 3% of chemistry, 2% of biology, 14% of medicine, and 2% of technology involves multiples. Except for the high proportion in medicine, these percentages seem to support the zeitgeist position in Test 2: The more codified and mathematical disciplines do appear to have a larger proportion of multiples. But again, this conjecture hinges on the assumption that the comparison of multiples and singletons is justified. Do we obtain collaboration when the probability distributions are studied?

Table 2 shows the observed frequency distributions as well as the Poisson statistics and goodness-of-fit tests for each of the seven fields. Observe that the Poisson model again provided an adequate fit, even though in no case can we assert that the chance theory is strictly proven. The means also vary from one field to another, but not in a manner consis-

Table 2
Multiple Frequencies and Poisson Statistics Within Disciplines

Item	Seven fields						
	Mathematics	Astronomy	Physics	Chemistry	Biology	Medicine	Technology
Multiple grade							
2	41	21	94	65	29	130	69
3	9	6	21	10	2	41	15
4	2	0	0	3	2	8	3
5	1	1	0	0	0	5	0
6	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0
8	0	0	0	0	0	0	1
Statistics							
μ	.84	.89	.46	.57	.56	1.02	1.61
SE	.42	.61	.18	.25	.38	.27	.36
x^2	.40	.26	3.57	1.24	.94	4.63	1.61
df	2	2	2	2	1	3	2
p	>.75	>.75	>.10	>.50	>.25	>.10	>.25

tent with either the zeitgeist or the genius theories. In fact, the standard errors of the estimates are so large that we cannot reject the hypothesis that all the sample means come from the same population mean.[3] Further reason for doubting the substantive significance of any mean differences comes from the fact that the magnitudes and ordering of the means in Table 2 are not very consistent with those found in a secondary analysis of the Ogburn and Thomas (1922) data (Simonton, 1978). Hence, this part of Test 2 seems to endorse the chance theory. If we reject the proportion of multiples to singletons as meaningless, then the probability that a multiple will appear is roughly the same across all disciplines.

Who Participates in Multiples?

Eminence, multiples, and creative productivity (Test 3). To scrutinize Test 3 empirically, we must first obtain a sample of creators. This end was accomplished by taking all those scientists and inventors with entries in either Asimov (1972) or Williams (1974) who were born on or after 1500 and who died on or before 1909. (The latter date was selected to maximize the number and quality of sources available for operationalizing creative productivity.) The result was a sample of 789 scientists and inventors. The eminence of each was determined by using the *Encyclopædia Britannica* (1974), a source shown by previous studies to yield reliable estimates of individual eminence (e.g., Simonton, 1976a, 1976d, 1977b). A scientist or inventor was given 3 points if included in the selective *Macropædia,* 2 points if given a "major entry" in the more inclusive *Micropædia,* 1 point if only given the terse "subsidiary entry" in the *Micropædia,* and 0 points if not included at all. Creative productivity was measured by counting the number of contributions listed in Darmstaedter (1908), a source shown to be very reliable for the period covered (Simonton, 1975, 1976b, 1976c). Finally, for each scientist or inventor I tabulated the number of multiples in which he or she participated. When we calculate the zero-order Pearson product–moment correlation coefficient between eminence and the number of multiples produced by an individual we find the statistically significant value of .25 ($p < .001$). So far, the zeitgeist theory is supported. Nonetheless, creative productivity correlates .32 ($p < .001$) with eminence and .42 ($p <$

[3] Even though we did not prove the Poisson model to be the best possible fit, the data are definitely far from being normally distributed. Therefore, it seems far more reasonable to base statistical tests on the standard error of estimate for each Poisson μ.

.001) with the number of multiples. Because the more eminent creators are more productive regarding both singletons and multiples, the correlation between eminence and multiples may be spurious, as chance theory predicts. Calculation of the first-order partial correlation between eminence of multiple productivity controlling for creative productivity yields a value of .14 ($p < .001$). We are now in something of an interpretative quandary. On the one hand, this partial correlation remains highly significant statistically, and consequently the zeitgeist theory seems justified. The more eminent creators are more likely to participate in multiples even after controlling for individual differences in the number of notable contributions. On the other hand, the substantive significance of the partial correlation is small, given that only 2% of the variance in eminence and multiple participation is shared after controlling for creative productivity. So in a substantive sense, the effect of zeitgeist is so small that we might adopt the chance theory as the main explanation.

Since we are forced to interpret such small effects, I decided to attempt a replication using different operationalizations of eminence and creative productivity. This time eminence was defined using the earlier edition of Asimov (1964), which makes finer distinctions regarding relative fame. A creator was given 3 points if he or she had a "major entry" in Asimov, 2 points for a "subsidiary entry," 1 point for only a listing in the index, and 0 points for no listing at all. This measure of eminence correlated .45 with the one previously derived from the *Encyclopædia Britannica* (1974). This interitem reliability coefficient compares quite favorably with intertest reliabilities for creativity and intelligence measures on selective populations (see, e.g., Wallach & Kogan, 1965). The alternative measure of creative productivity was not taken from a single source, but rather consisted of all contributions listed in a large number of separate chronologies (viz. Auerbach, 1923; Boyer, 1968; Dannemann, 1928; Daumas, 1957; Feldhaus, 1904; Gabriel & Fogel, 1955; Garrison, 1929; Grun, 1975; Hilditch, 1911; Langer, 1972; Mayer, 1949; Mumford, 1934). This second measure correlated .79 with that derived solely from Darmstaedter (1908). When we calculate the correlation between the two new measures of eminence and multiples we obtain .26 ($p < .001$), a value hardly differing from that obtained earlier. Moreover, the new measure of creative productivity correlates .37 ($p < .001$) with the new fame measure and .48 ($p < .001$) with the number of multiples, values only modestly larger than those calculated formerly. Therefore, it should come as no surprise that the partial correlation between eminence and multiple participation controlling for creative productivity is reduced only slightly to a still statistically significant .10 ($p < .05$). So the results we have obtained are extremely robust. In fact, no matter which combination of fame and creative productivity measures we employ, the outcome is essentially unaltered. Further, the same basic results emerge even if we employ only the exclusive sample of multiples.

It is manifest that the genius theory cannot explain the results. In addition, since the correlation between eminence and multiple production remains even after controlling for creative productivity, the zeitgeist theory has some credence. Nevertheless, partialing out creative productivity reduces the correlation to such a small value that the chance theory is not altogether ruled out as long as a modest zeitgeist influence is permitted.

Eminence and priority (Test 4). Are the more famous creators more likely to get there first, leaving their lesser colleagues the frustration of rediscovering the already known? To answer this question we first must be able to distinguish the relative eminence of each participant in a multiple. We have already shown how Asimov (1964) and the *Encyclopædia Britannica* (1974) could be used for this purpose, and the scores from these sources can be summed to produce an eminence measure ranging from 0 to 6. However, since many multiples involve obscure individuals, this composite score would not discriminate very well among the virtual nonentities. Therefore, possession of an entry in the extremely comprehensive reference edited by Debus (1968) added another point to the

composite score, generating an index ranging from 0 to 7.

Since the possible rank orderings vary depending on the grade of the multiple, it makes more sense to apply this eminence measure to each grade separately. An examination of doubles reveals that of the 449 cases, the data were insufficient in 63 to estimate the relationship between eminence and priority (usually because the two contributors were equally obscure). Of the remaining cases, the contributions were simultaneous in 197, the more eminent creator had priority in 79, and the less eminent creator had priority in 110. The preponderance of simultaneous contributions may be taken as evidence against the genius theory. I strongly suspect, however, that this proportion would be reduced if dissertation-sized research could be devoted to each case (e.g., the proportion of simultaneous multiples is much smaller for the exclusive sample that involves better documented cases). Much more curious is the fact that the priority assignments, when not simultaneous, are definitely contrary to the genius theory, at least as presumed in Test 4. The more eminent creators are more likely to be the rediscoverers, whereas the less eminent creator is more likely to have priority, $\chi^2(1) = 5.08$, $p < .01$. Perhaps the eminent scientist, like the eminent philosopher, is behind the times (cf. Simonton, 1976d). In the process of consolidating past advances, the major scientist may inadvertantly rediscover much of the work of various predecessors. Therefore, this outcome might be taken as support for the genius theory if we change our perspective on the genius from precursor to consolidator. Nonetheless, even if we grant this post hoc amendment, the impact of genius must be small, given the prevalence of simultaneous contributions.

Because the higher grade multiples are so few in number, inferences are necessarily less secure. Among triplets, the contributions are simultaneous in 36, the most eminent creator first in 15, second in 13, and third in 17 (the remaining 23 being unclassified for reasons given above). For quadruplets, 4 are simultaneous, and the most eminent creator is first 5 times, second twice, third twice, and last once (with 4 cases unclassified). Among quintuplets, there are no simultaneous multiples, the most eminent creator is first in 3 instances, and the third, fourth, and fifth places have 1 case each (and one unclassified). Finally, the single octuplet can only be considered a simultaneous multiple due to the plus or minus on the dating. Hence the whole picture suggests some tendency for the more eminent creators to have priority for the higher grade multiples, but none of these comparisons involve enough cases to permit statistical confidence in any conclusions.

Conclusion

The position that best meets all critical tests is the chance theory. Only the chance theory predicts the characteristic probability distribution of multiples and the absence of any probability differences from discipline to discipline. Although chance theory had to yield some ground to zeitgeist theory in Test 3, control for creative productivity considerably reduced the common variance between eminence and multiple participation. So on the whole, chance theory may be the best general explanation of the phenomenon. Despite the apparent explanatory superiority of the chance theory, however, the other two theories are not completely overturned. I think it is significant that the more eminent scientist is more likely to rediscover independently the work of his or her less eminent colleagues. Like the great philosopher, the eminent scientist appears to be a consolidator, synthesizing past achievements (cf. Simonton, 1976d). This consolidative process may sometimes entail some sociocultural redundancy in the form of rediscovery or reinvention. Nonetheless, the fact remains that there exist individual differences in priority, differences that cannot be explained by either chance or zeitgeist theories.

When we turn to zeitgeist theory, we cannot ignore its explanatory value either. We must obviously acknowledge that the zeitgeist is probably a necessary even if not a sufficient determinant of discovery or invention. There can be no denying that some contributions are prerequisites to other contribu-

tions. Certainly the discovery of bacteria presupposes a microscope, and the discovery of Jupiter's moons presumes the existence of a telescope. Nevertheless, the invention of these instruments does not automatically cause the corresponding discoveries for which they are prerequisite. The discoveries may or may not occur after the preconditions have been fulfilled: Galileo's discovery of Jupiter's satellites took place only 1 year after the telescope was invented, whereas Leeuwenhoek's discovery of bacteria happened about 200 years after the simple microscope appeared. So there remains some degree of indeterminancy in the influence of the zeitgeist, an indeterminancy consistent with the chance model.

Allow me to expand the last point: Not only can the zeitgeist and chance theories be seen to provide complementary accounts of multiples, but in addition, all three theories can be made compatible to some extent. Admittedly, I have sometimes oversimplified one or another theoretical outlook in the interest of drawing distinctive empirical predictions. Although advocates of the pure forms for each theory definitely exist (e.g., Kroeber, 1917), many researchers opt for some more moderate combination (e.g., Merton, 1961). By adjusting all three theories in the light of the current data, a more unified perspective on multiples can emerge. This possibility can be readily seen by more closely examining what I consider to be the two principal empirical results of this study.

The first main finding concerns the intercorrelation among individual eminence, creative productivity, and participation in multiples. It is definitely consistent with genius theory to notice a high correlation between the fame of an individual scientist or inventor and the number of contributions he or she makes (Albert, 1975). It is also in line with zeitgeist theory to observe that the more eminent contributor is more likely to produce more multiples. Yet in accord with the chance theory, a large part of the multiple production of more eminent creators is a probabilistic consequence of their greater creative productivity. Hence, all three theories tend to converge to explain the same phenomenon, at least so long as each undergoes some accommodations. The genius theory must permit the eminent scientist to be a major proponent of the current trends in his or her field—and not a dissenter or rebel. Similarly, the zeitgeist and chance theories must each yield some ground in explicating the relation between eminence and participation in multiples. Eminent scientists engage in more multiple production both because they are more productive and because they are more representative of the spirit of their times.

The potential compatibility of the three theories becomes even more evident when we scrutinize the second principal finding, namely, that the frequency of various multiple grades can be described by a Poisson distribution. To simplify the argument, let us suppose the mean of the distribution is somewhat close to unity. Then it may be argued that for any given potential contribution about 25 people are capable of the achievement, each having 1 chance out of 25 succeeding (i.e., $np = 25 \times 1/25 = 1$). These parameters can be made consistent with the genius model if we define the genius as a rare individual who has a probability greater than 0 of making notable contributions. It would be no slight achievement to be one out of a mere 25 persons in all history who are able to create the calculus or evolutionary theory or any other contribution. And even though chance plays a larger role than a pure genius theory might suggest, this interpretation still states that only a select few have any opportunity whatsoever. So serendipity may be ever present, but it is an elitist serendipity when so few have the capacity to be so lucky. Moreover, even if the probability that a creator would produce any given contribution were only around 1/25, the probability of that same creator producing *some* notable contribution would be high. On the average, after a little more than two dozen trials the odds would favor the creator "chancing" upon some major discovery or invention.

A chance model with the parameters of $n = 25$ and $p = 1/25$ could also be compatible with a zeitgeist explanation with some amendments in its traditional posture. If the spirit of the times is influential, then certainly we would expect that the course of scientific

progress would not depend on just one individual. And the existence of more than two dozen potenial contributors for any given discovery or invention obviously would provide the needed reserves. On the other hand, zeitgeist theory must relinquish the idea that discoveries or inventions are inevitable. Under the current parameters more than one-third of all potential contributions will be missed by all. Nonetheless, matters could be a lot worse: The mean could be much smaller than unity, so that the proportion of total failures could exceed half or more. As it stands, with a mean approaching 1, contributions are made once on the average. By permitting multiples to compensate for complete misses, discovery and invention can be said to have at least a probabilistic variety of inevitability.

In conclusion, after various mutual concessions all three perspectives can be said to enhance our overall understanding of multiples. A small group of highly productive individuals are most likely to participate in multiples, including independent rediscoveries. These same geniuses, as it were, are also unusually intimate with the technoscientific zeitgeist and are perhaps equally gifted with an inordinate amount of good luck.

Reference Note

1. R. K. Merton. Personal communication, April 19, 1976.

References

Albert, R. S. Toward a behavioral definition of genius. *American Psychologist,* 1975, *30,* 140–151.

Asimov, I. *Biographical encyclopedia of science and technology.* Garden City, N.Y.: Doubleday, 1964.

Asimov, I. *Biographical encyclopedia of science and technology* (new rev. ed.). Garden City, N.Y.: Doubleday, 1972.

Auerbach, F. *Entwicklungsgeschichte der modernen Physik.* Berlin: Springer, 1923.

Boring, E. G. *History, psychology, and science* (R. I. Watson & D. T. Campbell, Eds.). New York: Wiley, 1963.

Boyer, C. B. *A history of mathematics.* New York: Wiley, 1968.

Brass, W. Simplified methods of fitting the truncated negative binomial distribution. *Biometrika,* 1958, *45,* 59–68.

Cannon, W. B. The role of chance in discovery. *Scientific Monthly,* 1940, *50,* 204–209.

Constant, E. W. On the diversity and co-evolution of technological multiples: Steam turbines and Pelton water wheels. *Social Studies of Science,* 1978, *8,* 183–210.

Cox, C. *The early mental traits of three hundred geniuses.* Stanford, Calif.: Stanford University Press, 1926.

Dannemann, F. *Vom werden der naturwissenschaftlichen probleme.* Leipzig: Engelmann, 1928.

Darmstaedter, L. *Handbuch der geschichte der naturwissenschaften und der technik.* Berlin: Springer, 1908.

Daumas, M. (Ed.). *Encyclopédie de la Pléiade: Histoire de la science.* Paris: Librairie Gallimard, 1957.

Debus, A. G. (Ed.). *World who's who in science.* Chicago: Marquis Who's Who, 1968.

Dennis, W. Bibliographies of eminent scientists. *Scientific Monthly,* 1954, *79,* 180–183.

Dennis, W. Creative productivity between the ages of 20 and 80 years. *Journal of Gerontology,* 1966, *21,* 1–8.

Encyclopedia Britannica (15th ed.). Chicago: Encyclopædia Britannica, 1974.

Feldhaus, F. M. *Lexikon der erfindungen und entdeckungen auf den gebieten der naturwissenschaften und technik.* Heidelberg: Winter, 1904.

Feller, W. *An introduction to probability theory and its applications* (3rd ed., Vol. 1). New York: Wiley, 1968.

Fisz, M. *Probability theory and mathematical statistics* (3rd ed.). New York: Wiley, 1963.

Gabriel, M. L., & Fogel, S. (Eds.). *Great experiments in biology.* Englewood Cliffs, N.J.: Prentice-Hall, 1955.

Garrison, F. H. *An introduction to the history of medicine* (4th ed.). Philadelphia, Pa. Saunders, 1929.

Gillispie, C. C. (Ed.). *Dictionary of scientific biography* (14 vols.). New York: Scribner's, 1970.

Greenwald, A. G. Consequences of prejudice against the null hypothesis. *Psychological Bulletin,* 1975, *82,* 1–20.

Grun, B. *The timetables of history.* New York: Simon & Schuster, 1975.

Haight, F. A. *Handbook of the Poisson distribution.* New York: Wiley, 1967.

Harmon, L. R. The high school backgrounds of science doctorates. *Science,* 1961, *133,* 679–689.

Hilditch, T. P. *A concise history of chemistry.* New York: Van Nostrand, 1911.

Kroeber, A. H. The superorganic. *American Anthropologist,* 1917, *19,* 163–214.

Kroeber, A. H. *Anthropology: Culture patterns and processes.* New York: Harcourt, Brace & World, 1963.

Kuhn, T. S. *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago, 1970.

Langer, W. (Ed.). *An encyclopedia of world history.* Boston: Houghton Mifflin, 1972.

Mayer, A. *Annals of European civilization 1501–1900.* London: Cassell, 1949.

Merton, R. K. Singletons and multiples in scientific discovery: A chapter in the sociology of science. *Proceedings of the American Philosophical Society*, 1961, *105*, 470–486.

Molina, E. C. *Poisson's exponential binomial limit*. Princeton, N.J.: Van Nostrand, 1942.

Mumford, L. *Technics and civilization*. New York: Harcourt, Brace, 1934.

Ogburn, W. K., & Thomas, D. Are inventions inevitable? *Political Science Quarterly*, 1922, *37*, 83–93.

Patil, G. P. Estimation by two-moments method for generalized power series distributions and certain applications. *Sankhyā*, 1962, *24B*, 201–214.

Price, D. *Little science, big science*. New York: Columbia University Press, 1963.

Price, D. S. A general theory of bibliometric and other cumulative advantage processes. *Journal of the American Society for Information Science*, 1976, *27*, 292–306.

Roe, A. *The making of a scientist*. New York: Dodd, Mead, 1952.

Schmookler, J. *Invention and economic growth*. Cambridge, Mass.: Harvard University Press, 1966.

Simon, H. A. On a class of skew distribution functions. *Biometrika*, 1955, *42*, 425–440.

Simonton, D. K. Invention and discovery among the sciences: A *p*-technique factor analysis. *Journal of Vocational Behavior*, 1975, *7*, 275–281.

Simonton, D. K. Biographical determinants of achieved eminence: A multivariate approach to the Cox data. *Journal of Personality and Social Psychology*, 1976, *33*, 218–226. (a)

Simonton, D. K. The causal relation between war and scientific discovery: An exploratory cross-national analysis. *Journal of Cross-Cultural Psychology*, 1976, *7*, 133–144. (b)

Simonton, D. K. Interdisciplinary and military determinants of scientific productivity: A cross-lagged correlation analysis. *Journal of Vocational Behavior*, 1976, *9*, 53–62. (c)

Simonton, D. K. Philosophical eminence, beliefs, and zeitgeist: An individual–generational analysis. *Journal of Personality and Social Psychology*, 1976, *34*, 630–640. (d)

Simonton, D. K. Creative productivity, age, and stress: A biographical time-series of 10 classical composers. *Journal of Personality and Social Psychology*, 1977, *35*, 791–804. (a)

Simonton, D. K. Eminence, creativity, and geographic marginality: A recursive structural equation model. *Journal of Personality and Social Psychology*, 1977, *35*, 805–816. (b)

Simonton, D. K. Independent discovery in science and technology: A closer look at the Poisson distribution. *Social Studies of Science*, 1978, *8*, 521–532.

Singer, C., Holmyard, E. J., Hall, A. R., & Williams, T. I. (Eds.). *A history of technology* (5 vols.). Oxford: Oxford University Press, 1958.

Snedecor, G. W., & Cochran, W. G. *Statistical methods* (6th ed.). Ames: Iowa State University Press, 1967.

Sorokin, P. A. *Social and cultural dynamics* (abridged ed.). Boston: Porter Sargent, 1957.

Stern, B. J. *Social factors in medical progress*. New York: Columbia University Press, 1927.

Taton, R. (Ed.). [*The beginnings of modern science*] (A. J. Pomerans, Trans.). New York: Basic Books, 1964.

Taton, R. (Ed.). [*Science in the nineteenth century*] (A. J. Pomerans, Trans.). New York: Basic Books, 1965.

Taton, R. (Ed.). [*Science in the twentieth century*] (A. J. Pomerans, Trans.). New York: Basic Books, 1966.

Taylor, C. W., & Barron, F. (Eds.). *Scientific creativity*. New York: Wiley, 1963.

Terman, L. M., & Oden, M. H. *The gifted child grows up*. Stanford, Calif.: Stanford University Press, 1947.

Wallach, M. A., & Kogan, N. *Modes of thinking in young children*. New York: Holt, Rinehart & Winston, 1965.

Williams, T. I. (Ed.). *A biographical dictionary of scientists* (2nd ed.). New York: Wiley, 1974.

Received October 16, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 9, 1617-1630

Value Correlates of Conservatism

N. T. Feather
The Flinders University of South Australia, Bedford Park, Australia

Stepwise multiple regression analyses that involved measures of value importance, age, sex, education, and income as the independent variables were conducted to discover the major predictors of conservatism. On the basis of past discussions of the ideal conservative, it was hypothesized that conservative people would emphasize values concerned with attachment to rules and authority and ego defense (e.g., security, cleanliness, obedience, politeness, and salvation) and downgrade other values concerned with equality, freedom, love, and pleasure, as well as open-minded, intellectual, and imaginative modes of thought. This hypothesis was confirmed by the results from two independent surveys involving a sample of families in metropolitan Adelaide (Sample 1, 1972) and the families of a sample of students at Flinders University (Sample 2, 1976/1977). Respondents in both samples answered the Rokeach Value Survey and the Wilson and Patterson Conservatism Scale and provided background and demographic information. In addition to the values, age and sex were significant predictors (especially age), with older respondents tending to be more conservative than younger ones and females more conservative than males. Education and income of the heads (Sample 1) and fathers (Sample 2) of families played a minor role in prediction. Results are interpreted as supporting both the cognitive learning and psychodynamic explanations of value/attitude relationships. The need for developmental studies in this area using new sophisticated methodologies along with causal models is emphasized.

Does the person who has conservative views over a wide range of issues have a different pattern of values from the person whose attitudes are more liberal? The present article reports research that was designed to answer this question. For present purposes conservatism will be defined by reference to the analysis by Wilson and his colleagues (Wilson, 1973). Wilson (1973) claims that conservatism is "a general factor underlying the entire field of social attitudes much the same as intelligence is conceived as a general factor which partly determines abilities in different areas" (p. 3). The "ideal" conservative is seen as a person who tends to have a fundamentalist religious orientation and whose political leanings are likely to be proestablishment and supportive of the status quo; who tends to insist upon strict rules and punishments; who is likely to favor militarism; who tends to be ethnocentric and intolerant of minority groups; who tends to prefer what is conventional, traditional, and familiar; who is likely to be antihedonistic in outlook and to favor restriction of sexual behavior; who tends to oppose scientific progress and "new-fangled" ideas; and who is likely to be fatalistic and superstitious.

Wilson and his colleagues have used a questionnaire, the Conservatism Scale (C Scale), to measure general conservatism. This scale involves items that were selected to relate to the main characteristics of the ideal conservative that have just been mentioned. Each item consists of a word or phrase representing a familiar or controversial issue (e.g., legalized abortion, death penalty, Bible

This research was funded by grants from the Australian Research Grants Committee and from Flinders University.

Requests for reprints should be addressed to N. T. Feather, Discipline of Psychology, School of Social Sciences, The Flinders University of South Australia, Bedford Park, South Australia 5042.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3709-1617$00.75

truth), and the respondent registers approval, disapproval, or uncertainty. This simple item format is assumed to overcome many of the problems inherent in more traditional item formats employing attitude statements.

Wilson (1973, chaps. 4 & 5) draws upon factor analytic evidence in support of his claim that the C Scale measures a general factor called conservatism.[1] One should note, however, that his claim has not, as far as the C Scale is concerned, gone undisputed (Boshier, 1972; Robertson & Cochrane, 1973 —but see also Kirton, 1978; Wilson, 1972). The issue of the structure of conservatism is controversial and complex and is unlikely to be resolved on the basis of factor analytic findings alone (Feather, 1975a). It is clearly necessary to develop theories about the nature of general conservatism and its behavioral correlates and about how conservative attitudes develop. Then one can proceed to test these theories by using a wide range of empirical procedures.

Wilson's own theoretical stance vis-à-vis the conservative attitude syndrome is that it is intimately related to genetic and environmental factors that determine feelings of insecurity and inferiority. The common basis for all of the various components of the syndrome is assumed to be "a generalized susceptibility to experiencing threat or anxiety in the face of uncertainty" (Wilson, 1973, chap. 17). The conservative individual tends to avoid both stimulus and response uncertainty, and this avoidance is reflected in the verbal attitudes that are expressed as well as in other aspects of behavior. Conservative attitudes are therefore assumed by Wilson (1973) to

> . . . serve a defensive function. They arise as a means of simplifying, ordering, controlling, and rendering more secure, both the *external* world (through perceptual processes, stimulus preferences, etc.) and the *internal* world (needs, feelings, desires, etc.). Order is imposed upon inner needs and feelings by subjugating them to rigid and simplistic external codes of conduct (rules, laws, morals, duties, obligations, etc.), thus reducing conflict and averting the anxiety that would accompany awareness of the freedom to choose among alternative modes of action. (pp. 261–264)

These ideas are not dissimilar to those involved in the earlier influential research into the authoritarian personality (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950; Kirscht & Dillehay, 1967; Ray, 1973; Sanford, 1973)—for example, in the emphasis upon the defensive functions of conservative attitudes—although Wilson's analysis of conservatism stresses order and control as a means of reducing anxieties relating to feelings of insecurity and inferiority, whereas the theoretical account of authoritarianism draws upon Freudian concepts and deals more with the control of impulses that may be personally and socially unacceptable, such as sex and aggression. Sanford's (1973) recent overview of research into the authoritarian personality impresses with its subtle awareness of the complex personality processes that may underly authoritarian attitudes and the important role played by the family and the wider social context. It is noteworthy, however, that the main personality dispositions of the authoritarian as described by Sanford (1973, pp. 143–146) overlap with the characteristics of the ideal conservative that were previously listed. Descriptively, therefore, the concept of conservatism as employed by Wilson has much in common with the earlier concept of authoritarianism. And at the level of explanation, there is some agreement also that these attitude syndromes may serve the function of ego defense.

In the present article we make no attempt to develop a theory about the structure and dynamics of conservatism. Rather our main interest is to discover whether conservative attitudes are related in predictable ways to more general values. Very little is known about these relationships, but it is reasonable to expect that attitude–value linkages will occur and that their emergence will depend upon the individual's developing ability to form abstract concepts that capture the

[1] For example, a principal-components analysis of the C Scale yielded a first unrotated factor that accounted for much more of the total variance (18.67 percent) than did subsequent factors and that involved a wide variety of items from different content areas (e.g., race, religion, and sex). The high loading items were approximately equally divided as to direction of scoring for conservatism, suggesting that the factor could not be interpreted in terms of response set. Moreover, loadings on the first factor were very similar to the item–whole correlations.

regularities or consistencies in experience, an ability that improves with increasing maturity. It is likely that specific attitudes come first and that general values begin only later to emerge as abstractions from personal experience, especially experience of one's own attitudes and actions, their relationship to the attitudes and behaviors of significant others, and the judgments of significant others about what is and what is not desirable. In time these values become organized into value systems according to their relative importance. Once formed, values and value systems serve as frames of reference or criteria that may be used to guide thought and action in many different contexts. They represent a personal working over of experience in terms of the individual's own internal cognitive capacities and personality dynamics. They may be modified and refined in light of new and discrepant experience, but throughout life they remain as important components of the self-concept. These points have been discussed and elaborated in previous publications (Feather, 1975b, 1978, in press; Rokeach, 1973).

What are some of the values that one would expect to be associated with the conservative attitude syndrome? The previous discussion suggests that values that are commonly held to involve ego defense would be involved. Rokeach (1973, p. 16) singles out a number of them. Research on the authoritarian personality and on religion, he argues, suggests that an overemphasis "on such modes of conduct as cleanliness and politeness and on such end-states as family and national security may be especially helpful to ego defense" (p. 16) and that "religious values more often than not serve ego-defensive functions" (p. 16). At the same time, Rokeach (1973) recognizes that the motivational base of values is complex and that in the final analysis "all of a person's values are conceived to maintain and enhance the master sentiment of self-regard—by helping a person adjust to reality, defend his ego against threat, and test reality" (p. 15).

The values mentioned by Rokeach are contained in his Value Survey, a test that lists 18 terminal values (concerned with goals or end states of existence) and 18 instrumental values (concerned with means or modes of conduct). There are other values from these lists, however, in addition to those related to ego defense, that one would expect to be associated with general conservatism. Thus, a conservative tendency to overglorify authority, rules, and punishment would probably be associated with a more extreme emphasis upon such instrumental values as being obedient and polite and with a downgrading of terminal values such as equality and freedom. Strict adherence to the familiar, the traditional, and the conventional, as demonstrated, for example, in anti-intellectual and antiscientific biases, would probably be associated with diminished importance attached to such instrumental values as being broadminded, imaginative, intellectual, and logical. It is likely that the antihedonistic attitudes assumed to be characteristic of the ideal conservative would be associated with downgrading such terminal values as mature love and pleasure. Finally, the ethnocentric bias of the ideal conservative might appear in an upgrading of cleanliness values and a downgrading of equalitarian values.

In summary, it was predicted that general conservatism as measured by the C Scale would be associated with a more extreme emphasis upon values concerning salvation, security, cleanliness, and obedience, and a less extreme emphasis upon values concerning equality, freedom, pleasure, love, and modes of conduct that involve acting and thinking in broadminded, imaginative, intellectual, and logical ways. These predictions are all consistent with the descriptive and theoretical accounts of conservatism and authoritarianism that were mentioned previously and that have been extensively discussed by others (e.g., Kirscht & Dillehay, 1967; Sanford, 1973; Wilson, 1973).

In addition to providing evidence relating to these predicted relationships, the present study also included background and demographic variables (age, sex, education, income) in the analysis with a view to investigating their relationships to conservatism as well. It is already known that conservatism as measured by the C Scale tends to increase with age and to be higher for females (Feather, 1975b, 1977b; Wilson, 1973). The relationship of conservatism to socioeconomic status is, however, less clear. In a recent

study of the structure of social attitudes, Eysenck (1975) presented evidence for two conservatism factors—general conservative versus radical ideology and socioeconomic conservatism versus socialism—in addition to the tough-mindedness versus tender-mindedness dimension that appeared in his earlier work (Eysenck, 1954). He found that middle-class people tended to be more radical in their general attitudes but more conservative in their economic attitudes when compared with working-class people (see also Lipset, 1960). These results suggest that C Scale scores should tend to decrease with increased level of education, given the fact that the C Scale appears to be measuring the type of general conservatism factor to which Eysenck refers.[2] Thus, it was further predicted that general conservatism will not only increase with age and be higher for females than for males but will also be lower for higher levels of education.

A final aim of the present study was to discover the most important value predictors of general conservatism by using stepwise multiple regression procedures and to compare the level of prediction achieved by these values with that achieved by the values, the background, and demographic variables taken together, and with that achieved by the background and demographic variables alone. Such a comparative analysis does not involve prior specification of a causal model, but it does throw light on the question as to whether information about a person's value priorities provides better prediction of general conservatism than information about age, sex, education, and income and whether information about both sets of variables (values plus background and demographic information) yields an even better result. One can then determine just how much is added by each set of variables when they are included in the predictive equation. The analysis has theoretical significance in that it provides a rough clue as to whether the relationships between value priorities and general conservatism are simply the products of socialization and other experiences associated with age, sex, and socioeconomic status or whether these relationships are more complexly determined and involve not only the molding effect of experience in particular situations but also the operation of internal personality dynamics that affect the way environmental input is coded and transformed (see Sanford, 1973, for a good discussion of this point).

A distinctive feature of the present investigation is that it involved two separate data sets collected at different times with different samples. Hence it was possible to discover how far results with one sample were replicated by the results of the other sample.

Method

Subjects and Procedure

Sample 1 (Adelaide metropolitan). The data for this sample were obtained from an extensive family survey conducted in metropolitan Adelaide in 1972. Details about subjects and procedure have been presented previously (Feather, 1975b, pp. 121–125; 1978). The survey involved dwellings that included children 14 years of age or more. The sample was selected by using a multistage cluster sampling frame of the Adelaide Statistical Division maintained by the School of Social Sciences at Flinders University. An attempt was made to survey all adult members in each dwelling, including children 14 years of age or older. A trained interviewer conducted the survey. Respondents were assured that their answers would be confidential, and names were not required on the questionnaire. They were requested to complete the questionnaire independently, without discussing it with other members of the family. Out of a total of 659 possible respondents from the dwellings sampled, there were 72 refusals, leaving 587 respondents who answered some or all of the complete survey questionnaire (an 89% response rate). These respondents were made up of 147 who described themselves as heads of household (117 male, 30 female), 145 wives, 152 sons, 126 daughters, and 17 from other members of the dwellings.

Sample 2 (Flinders University). This sample involved students who were enrolled in a third year (senior) course on social motivation offered at Flinders University in 1976 and again in 1977, together with members of their families when they were available. These students completed a questionnaire containing the relevant measures. The

[2] The factor analytic studies discussed by Wilson (1973) do not provide evidence for a factor of socioeconomic conservatism. However, this may simply reflect an absence of relevant items in the C Scale.

majority of them then took the questionnaire home and administered it to their parents and to siblings who were 14 years of age or older living in the dwelling. A small number of older students administered it to their wives (or husbands) and to those of their children who were 14 years of age or older and were living in the dwelling. It was emphasized that the questionnaires should be answered independently by family members without any consultation with each other and that answers were confidential. Names were not required on the questionnaire. Data were obtained from 358 respondents consisting of 74 fathers, 83 mothers, 92 sons, and 109 daughters. The response rate from available family members was very high, exceeding 90%.

Questionnaire. The first part of the questionnaire was designed to obtain background and demographic information (e.g., status in household, age, sex, birthplace, income, religion, etc.). This was followed by Form D of the Rokeach Value Survey, containing the two sets of terminal and instrumental values. In this version each value with its short definition is printed on a removable gummed label, and the values in each set are presented in alphabetical order. One set of 18 terminal values concerns general goals or end states of existence (e.g., freedom, happiness, salvation, wisdom). The other set of 18 instrumental values concerns modes of conduct (e.g., being broadminded, honest, loving, responsible). Respondents ranked the 18 terminal and then the 18 instrumental values in their "order of importance to *you,* as guiding principles in *your* life," using the standard instructions (Rokeach, 1973, pp. 357–361). Respondents arrived at their final rank orders by rearranging the gummed labels within each value set so that the final rank orders indicated "how you really feel."

As in other studies in the Flinders program (Feather, 1975b, pp. 23–24) the ranks from 1 (most important) to 18 (least important) were transformed to z scores corresponding to a division into 18 equal areas under the normal curve. This transformation was made on the assumption that it would be easier for respondents to discriminate between the relative importance of values at the extremes of the scale than between those ranked in the middle of the scale (see Cohen & Cohen, 1975, p. 265, and Hays, 1967, pp. 35–39, for justification of this procedure). These transformed ranks for the 18 terminal and the 18 instrumental values constituted the measures of value importance for each respondent, and they could range from +1.91 (most important) to −1.91 (least important).

Respondents also completed the Conservatism Scale devised by Wilson and Patterson (1968) and discussed in Wilson (1973, 1975). The C Scale consists of 50 items, and as mentioned previously, each item relates to some familiar or controversial issue or concept (e.g., death penalty, evolution theory, chastity, church authority, colored immigration, disarmament, strict rules, and so forth). Respondents were asked, "Which of the following do you favor or believe in?" and they answered each item by circling *yes, ?,* or *no.* The C Scale is balanced for acquiescence response set, since *yes* answers to *odd*-numbered items are scored to reflect a conservative response and *no* answers a liberal or nonconservative response, whereas *yes* answers to *even*-numbered items are scored to reflect a liberal response and *no* answers a conservative response. Responses to each item were scored in the usual manner (liberal response = 0, ambiguous response = 1, conservative response = 2), so that total conservatism scores could range from 0 to 100 in the direction of increasing conservatism. There were 575 respondents in Sample 1 and 358 respondents in Sample 2 who completed the C Scale.[3]

Form of Analysis

Product–moment correlation coefficients were computed between the transformed rankings for each value and general conservatism as measured by the total score on the C Scale, and between each of the demographic and background variables and this total score. Subsequently, stepwise multiple regression analyses were conducted using the Statistical Package for the Social Sciences (Nie, Hull, Jenkins, Steinbrenner, & Bent, 1975, chap. 20).

In these analyses sex was coded 1 (male) and 2 (female). Education and income were both coded in relation to the nominated head of each household for Sample 1 (the 1972 metropolitan Adelaide family sample) and in relation to the father of the family in Sample 2 (the 1976/1977 Flinders University family sample). The scores obtained were used for all members of that household so as to provide general measures of the socioeconomic status of each family. In Sample 1 education was coded from 1 to 4 using the following categories: primary or elementary school education (1); secondary school education without matriculation (2); matriculated at secondary school and/or obtained trade qualifications (3); tertiary education (4). Income was coded from 1 to 5 using the following categories for reported annual income: less than $2,000 (1); $2,000–$3,999 (2); $4,000–$5,999 (3); $6,000–$7,999 (4); above $8,000 (5). Both education and income levels were higher in Sample 2 than in Sample 1, as might be expected from the nature of the sample and from the fact

[3] Medians of the ranks for each of the terminal and instrumental values and means for the total conservatism scores for Sample 1 are presented for different groups in Feather (1975b). Correlations between value importance and total conservatism with age partialed out are presented in Feather (1977b). Item means for the 50 items from the C Scale for Sample 1 can be found in Feather (1975a, 1977a). Corresponding information for Sample 2 has not yet been published.

Table 1
Product-Moment Correlations (Simple rs) Relating Independent Variables to Total Conservatism for Samples 1 and 2

Independent variable	Total conservatism	
	Sample 1 *r*	Sample 2 *r*
Value importance, terminal values		
A comfortable life	.04	−.06
An exciting life	−.27***	−.31***
A sense of accomplishment	−.04	−.03
A world at peace	−.04	.03
A world of beauty	−.16***	−.15**
Equality	−.25***	−.18**
Family security	.36***	.35***
Freedom	−.28***	−.22***
Happiness	.02	−.03
Inner harmony	.01	−.06
Mature love	−.20***	−.18**
National security	.30***	.33***
Pleasure	−.16***	−.23***
Salvation	.41***	.56***
Self-respect	.13**	−.05
Social recognition	−.01	−.07
True friendship	−.11*	−.23***
Wisdom	.08	.07
Other variables		
Age	.48***	.47***
Sex	.24***	.09
Education	−.11*	.01
Income	−.13*	.02

Independent variable	Total conservatism	
	Sample 1 *r*	Sample 2 *r*
Value importance, instrumental values		
Ambitious	.09*	−.04
Broad-minded	−.39***	−.43***
Capable	.03	−.05
Cheerful	−.09*	−.10
Clean	.30***	.32***
Courageous	−.01	.00
Forgiving	.02	.21***
Helpful	.08	.06
Honest	.15***	.11*
Imaginative	−.32***	−.44***
Independent	−.23***	−.35***
Intellectual	−.16***	−.18**
Logical	−.19***	−.05
Loving	.03	.07
Obedient	.33***	.45***
Polite	.27***	.30***
Responsible	.12**	.06
Self-controlled	.04	.15**

Note. Because of missing data, the *N*s for the correlations in this table for Sample 1 (metropolitan Adelaide) ranged from 558 to 575 except for education and income, where the *N*s were 457 and 430, respectively. For Sample 2 (Flinders University), the *N*s ranged from 357 to 358 except for education and income, where *N* was 331.
* $p < .05$. ** $p < .01$. *** $p < .001$.

that data were collected some 4 or 5 years after Sample 1 had been surveyed. For example, 27% of fathers in Sample 2 had or were receiving tertiary education, whereas the corresponding percentage for heads of household in Sample 1 was 17%; and 46% of fathers in Sample 2 had incomes greater than $10,000, whereas the corresponding percentage for heads of household in Sample 1 was 7%. Because of these higher levels of education and income in Sample 2 and because of the smaller sample size, it was not feasible to subdivide Sample 2 for education and income by using the Sample 1 categories. Instead, education was coded as follows: secondary school education without matriculation and/or primary or elementary school education (1), matriculated at secondary school and/or obtained trade qualifications and/or tertiary education (2). Income for Sample 2 was coded as follows: below $10,000 (1), above $10,000 (2). It should be noted, therefore, that analyses involving education and income as variables were on safer ground for Sample 1, where it was possible to distinguish more categories and where the sample was more representative of the population at large.

Results

Simple Correlations of Variables With Conservatism

Table 1 presents the product–moment correlations relating each of the independent variables to total conservatism for Sample 1 and Sample 2.

In both samples there were statistically significant positive relationships between total conservatism and the relative importance assigned by respondents to the following values: family security, national security,

salvation, being clean, being honest, being obedient, and being polite. Also, in both samples there were significant negative relationships between total conservatism and the relative importance assigned by respondents to the following values: an exciting life, a world of beauty, equality, freedom, mature love, pleasure, true friendship, being broad-minded, being imaginative, being independent, and being intellectual. The significant relationships were remarkably consistent from sample to sample not only in their direction but also in their magnitude, and the majority of them were predicted in advance.[4]

The correlations for both samples also showed that total conservatism tended to increase with age and to be higher for females than for males. These relationships were also consistent with predictions.

Total conservatism scores were negatively related to the level of education and income of the head of household in Sample 1, supporting predictions. These relationships were not strong, however, even though they were statistically significant. Nor did they occur in Sample 2, probably because this sample was more highly selected (families of Flinders University students) and, as noted before, less differentiated in the categories of education and income that were employed in the analysis.

A small number of significant value/conservatism correlations were not replicated from sample to sample. Some of these may involve the effects of sample differences and some may be chance events. They will not be discussed further.

Table 2
Multiple Correlations (R*s*) *and Proportion of Variance Explained* (R^2) *From Multiple Regression Analyses Predicting Total Conservatism From Age, Sex, Education, and Income*

Independent variable	Sample 1		Sample 2	
	R	R^2	*R*	R^2
Age + sex	.52	.27	.48	.23
Education + income	.14	.02	.02	.00
Age + sex + education + income	.53	.28	.48	.23

Multiple Regression Analyses

The relationships presented in Table 1 are further clarified by the results of the multiple regression analyses presented in Tables 2, 3, 4, and 5.[5]

Table 2 shows that the combination of age, sex, education, and income was able to account for 28% of the variance in conservatism scores in Sample 1 and 23% in Sample 2, with age and sex as the most important predictors (especially age—see Table 1).

At the 10th step of the stepwise solution, the set of terminal values alone accounted for 45% of the variance in conservatism scores in Sample 1 and 53% in Sample 2; the instrumental values alone accounted for 31% of this variance in Sample 1 and 44% in Sample 2 at the 10th step (see Table 3).[6]

The combination of terminal values with age, sex, education, and income accounted for 50% of the variance in conservatism scores in Sample 1 and 59% in Sample 2

[4] The Sample 1 correlations between value importance and total conservatism and between value importance and age presented in Tables 1 and 6 respectively are marginally different from those presented in a related report (Feather, 1977b) because of very slight differences in the *N*s between both analyses due to the fact that the present study used pairwise deletion of missing data.

[5] In the absence of specific hypotheses about curvilinear trends and interaction effects, it was decided to use a linear regression equation without interaction terms. Moreover, given the large number of independent variables for final inclusion in the equation, this seemed to be a reasonable decision. One exception was the inclusion of Age × Sex as an interaction term in an exploratory analysis involving age, sex, and Age × Sex as predictors. Inclusion of this interaction term made virtually no difference to the multiple *R*s for Sample 1 and Sample 2 when compared with the *R*s obtained by using the equation involving age and sex alone.

[6] Results in Tables 3, 4, and 5 are reported for the first 10 steps for ease of comparison. The step beyond which there was no further significant ($p < .05$) step-by-step increase in R^2 is noted, either in the table or in the text. In all of the analyses reported in Tables 2, 3, 4, and 5 the "adjusted" R^2 values were virtually identical to the actual R^2 values (Nie, Hull, Jenkins, Steinbrenner, & Bent, 1975, p. 358).

Table 3
Multiple Correlations (Rs), *Proportion of Variance Explained* (R^2), *and Standardized Regression Coefficients* (βs) *From Stepwise Multiple Regression Analyses Predicting Total Conservatism From Terminal and Instrumental Values*

Sample 1					Sample 2				
Step	Value	R	R^2	β	Step	Value	R	R^2	β
				Terminal values					
1	Salvation	.41	.171	.42***	1	Salvation	.56	.315	.56***
2	Family security	.53	.283	.24***	2	Family security	.65	.419	.21***
3	National security	.58	.341	.24***	3	National security	.69	.473	.22***
4	Equality	.61	.375	−.13***	4	Equality	.70	.497	−.16***
5	Mature love	.64	.406	−.15***	5	Social recognition	.71	.509	.08
6	Freedom	.65	.418	−.08*	6	Mature love	.72	.516	−.09*
7	Self-respect	.65	.426	.12***	7	An exciting life	.72	.521	−.09*
8	A comfortable life	.66	.432	.12**	8	A comfortable life	.72	.525	.07
9	Wisdom	.66	.439	.10**	9	A sense of accomplishment	.73	.528	.05
10	Happiness	.67	.445	.08*	10	Freedom	.73	.530	−.04
				Instrumental values					
1	Broad-minded	.39	.149	−.28***	1	Obedient	.45	.206	.15**
2	Imaginative	.46	.212	−.18***	2	Broad-minded	.55	.301	−.31***
3	Clean	.49	.245	.15***	3	Imaginative	.61	.368	−.28***
4	Obedient	.52	.268	.11**	4	Independent	.63	.397	−.21***
5	Independent	.53	.280	−.16***	5	Clean	.64	.407	.09*
6	Cheerful	.54	.293	−.11**	6	Cheerful	.64	.416	−.11*
7	Loving	.55	.297	−.08*	7	Responsible	.65	.424	−.10*
8	Logical	.55	.301	−.07	8	Loving	.66	.430	−.08
9	Helpful	.55	.304	−.05	9	Helpful	.66	.432	−.06
10	Responsible	.55	.306	.05	10	Forgiving	.66	.435	.06

Note. The standardized regression coefficients (βs) are those obtained at the end of 10 steps in the analysis. An underlined R^2 indicates that there were no statistically significant ($p < .05$) step-by-step increases in R^2 after that step in the analysis.
* $p < .05$. ** $p < .01$.*** $p < .001$.

at the 10th step; the set of instrumental values combined with these four background and demographic variables accounted for 47% of this variance in Sample 1 and 56% in Sample 2 at the 10th step (see Table 4).

The combination of terminal values and instrumental values accounted for 48% of the variance in conservatism scores in Sample 1 and 60% of this variance in Sample 2 at the 10th step (see Table 5). In both samples step-by-step changes in R^2 continued to be statistically significant ($p < .05$) up to the 14th step. For Sample 1, $R = .71$ and $R^2 = .504$ at the 14th step following the addition of *cheerful, true friendship, imaginative,* and *freedom* to the regression equation in that order. For Sample 2, $R = .78$ and $R^2 = .615$ at the 14th step following the addition of *capable, responsible, a sense of accomplishment,* and *polite* to the regression equation, in that order.

Finally, the combination of all 36 values (terminal plus instrumental) with age, sex, education, and income accounted for 55% of the variance in conservatism scores in Sample 1 and 67% of this variance in Sample 2 at the 10th step (see Table 5). In Sample 1, step-by-step changes in R^2 continued to be statistically significant ($p < .05$) up to the 14th step, where $R = .76$ and $R^2 = .570$, following the addition of *a world of beauty, responsible, independent,* and education to the regression equation, in that order. In Sample 2, statistically significant step-by-step increases in R^2 stopped at the 9th step (see Table 5).

The predictor variables that emerged in the early steps of the analyses reported in Tables 3, 4, and 5 were very similar across the two samples, and they tended to occur in much the same order in both samples. For example, *salvation, family security, national security,* and *equality* emerged in that order for both Sample 1 and Sample 2 in the first four steps of the solution for the terminal values presented in Table 3; age, *broad-minded,* and *obedient* emerged in the first three steps of the second set of analyses presented in Table 4 for both Sample 1 and Sample 2; and age, *salvation, broad-minded,* and *obedient* occupied the first four steps of the second set of analyses reported in Table 5 for both samples. The solutions presented in Tables 3, 4, and 5 therefore appear quite robust—there was a high level of replication from sample to sample despite the fact that different populations were involved and the samples were taken 5 years apart.

Finally, the terminal values were slightly better predictors of conservatism than the instrumental values (see Table 3) and prediction throughout was slightly better for Sample 2 than for Sample 1 (see Tables 3, 4, and 5).

Value Importance and Age

In view of the fact that age emerged from the multiple regression analyses as an important predictor of conservatism, it is also of interest to inquire about the degree to which value importance was a function of

Table 4
Multiple Correlations (R*s*), *Proportion of Variance Explained* (R^2), *and Standardized Regression Coefficients* (βs) *From Stepwise Multiple Regression Analyses Predicting Total Conservatism From Values, Age, Sex, Education, and Income*

Sample 1					Sample 2				
Step	Variable	*R*	R^2	β	Step	Variable	*R*	R^2	β
	Terminal values + age + sex + education + income								
1	Age	.48	.230	.30***	1	Salvation	.56	.315	.56***
2	Salvation	.60	.357	.32***	2	Age	.70	.490	.31***
3	National security	.63	.401	.15***	3	National security	.73	.527	.18***
4	Equality	.66	.429	−.18***	4	Equality	.74	.544	−.10*
5	Mature love	.67	.453	−.17***	5	Mature love	.75	.565	−.09*
6	Family security	.69	.473	.11**	6	Family security	.76	.574	.12**
7	Sex	.70	.486	.12***	7	Social recognition	.76	.580	.10*
8	Inner harmony	.70	.492	−.09*	8	Happiness	.76	.584	.05
9	A world of beauty	.71	.497	−.08*	9	A comfortable life	.77	.586	.08
10	Freedom	.71	.502	−.08*	10	Sex	.77	.591	.07
	Instrumental values + age + sex + education + income								
1	Age	.48	.230	.40***	1	Age	.47	.218	.38***
2	Broad-minded	.59	.343	−.21***	2	Obedient	.65	.423	.25***
3	Obedient	.63	.402	.21***	3	Broad-minded	.70	.489	−.27***
4	Sex	.65	.421	.17***	4	Independent	.71	.511	−.19***
5	Ambitious	.66	.440	.14***	5	Imaginative	.73	.531	−.17***
6	Clean	.67	.451	.11**	6	Income	.73	.539	.09*
7	Independent	.68	.458	−.09*	7	Sex	.74	.544	.10*
8	Education	.68	.464	−.09*	8	Loving	.74	.550	−.08
9	Responsible	.68	.469	.09*	9	Cheerful	.74	.553	−.08
10	Courageous	.69	.473	.07	10	Responsible	.75	.557	−.06

Note. The standardized regression coefficients (βs) are those obtained at the end of 10 steps in the analysis. An underlined R^2 indicates that there were no statistically significant ($p < .05$) step-by-step increases in R^2 after that step in the analysis.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 5
Multiple Correlations (Rs), *Proportion of Variance Explained* (R^2), *and Standardized Regression Coefficients* (βs) *From Stepwise Multiple Regression Analyses Predicting Total Conservatism From All 36 Values and From All 36 Values, Age, Sex, Education, and Income*

Sample 1					Sample 2				
Step	Variable	R	R^2	β	Step	Variable	R	R^2	β
	Terminal values + instrumental values								
1	Salvation	.41	.172	.32***	1	Salvation	.56	.315	.36***
2	Family security	.53	.283	.18***	2	Family security	.65	.419	.13***
3	National security	.58	.341	.17***	3	Obedient	.69	.474	.14**
4	Broad-minded	.62	.384	−.19***	4	National security	.72	.517	.13***
5	Equality	.64	.413	−.17***	5	Equality	.73	.540	−.17***
6	Mature love	.66	.442	−.16***	6	Broad-minded	.75	.556	−.16***
7	Clean	.68	.458	.12***	7	Mature love	.76	.571	−.13***
8	Self-respect	.68	.467	.09**	8	Imaginative	.76	.581	−.14**
9	Independent	.69	.476	−.11**	9	True friendship	.77	.588	−.10*
10	A world of beauty	.70	.484	−.10**	10	Independent	.77	.595	−.10*
	Terminal values + instrumental values + age + sex + income + education								
1	Age	.48	.230	.35***	1	Salvation	.56	.315	.35***
2	Salvation	.60	.357	.26***	2	Age	.70	.490	.34***
3	Broad-minded	.65	.419	−.15***	3	Obedient	.76	.574	.18***
4	Obedient	.67	.454	.14***	4	Broad-minded	.77	.600	−.18***
5	National security	.69	.479	.12**	5	Mature love	.79	.625	−.16***
6	Equality	.71	.502	−.16***	6	Equality	.80	.644	−.12***
7	Mature love	.72	.518	−.11**	7	National security	.81	.654	.09*
8	Sex	.73	.529	.13***	8	Independent	.81	.661	−.12**
9	Ambitious	.74	.541	.11**	9	Imaginative	.82	.666	−.08*
10	Clean	.74	.550	.10**	10	Helpful	.82	.670	−.06

Note. The standardized regression coefficients (βs) are those obtained at the end of 10 steps in the analysis. An underlined R^2 indicates that there were no statistically significant ($p < .05$) increases in R^2 after that step in the analysis.
* $p < .05$. ** $p < .01$. *** $p < .001$.

age in both samples. Table 6 reports the product–moment correlations of value importance with age for each of the terminal and instrumental values. (Correlations showing the positive relationships between age and conservatism are contained in Table 1.)

Table 6 indicates that in both samples there were statistically significant positive relationships between age and the relative importance assigned by respondents to the following values: family security, national security, self-respect, being clean, and being polite. These values tended to increase in relative importance with increasing age. Also in both samples there were significant negative relationships between age and the relative importance assigned by respondents to the following values: an exciting life, equality, freedom, social recognition, true friendship, being broad-minded, and being imaginative. These values tended to decrease in importance with increasing age.

Thus, Table 6 shows that there were several value/age relationships that were consistent from sample to sample. These changes in value importance with age are of interest in their own right but they will not be discussed in detail—see Feather (1975b, chap. 6) and Feather (in press) for relevant discussion of age differences in values.

There were a small number of significant value/age correlations that were not replicated from sample to sample. Most of these significant relationships occurred in Sample 1, and they may reflect the fact that this sample came from a wider and more hetero-

geneous population than Sample 2. Some of the nonreplicated effects may also be chance events.

Discussion

The results of the present study are clear and consistent. As predicted, conservatism was positively related to the relative importance of values that could be taken to involve ego defense and attachment to authority—values that were concerned with salvation, security, cleanliness, and rule following. The higher the conservatism score, the more likely it was that these values would be emphasized. Also as predicted, conservatism was negatively related to the relative importance of values that referred to equality, freedom, love, and pleasure, as well as open-minded, intellectual, and imaginative ways of thinking. The higher the conservatism score, the less likely it was that these values would be emphasized. Conservatism was also associated with age and sex in the direction predicted. Thus, older people tended to be more conservative than younger people, and females were slightly more conservative than males. Finally, the relative importance of some of the values varied systematically with differences in age. All of these results were replicated from sample to sample and they therefore stand as reliable findings. Moreover the replication indicated that correlations were not only in the same direction in both samples but were also similar in their orders of magnitude.[7]

The results also showed that conservatism was negatively related to the education and income levels of heads of households in Sample 1, though the correlations were relatively small. Higher conservatism tended to

[7] Those with high conservatism scores were also more likely to downgrade values concerned with beauty, excitement, and independence, perhaps again reflecting a preference for convention, tradition, structure, and stability as opposed to novelty, independence, complexity, and change.

Table 6
Product–Moment Correlations (Simple rs) Relating Value Importance to Age for Samples 1 and 2

Value importance	Age		Value importance	Age	
	Sample 1 *r*	Sample 2 *r*		Sample 1 *r*	Sample 2 *r*
Terminal values			Instrumental values		
A comfortable life	.05	.05	Ambitious	−.09*	−.06
An exciting life	−.29***	−.29***	Broad-minded	−.11*	−.11*
A sense of accomplishment	.07	.09	Capable	.09*	.12*
A world at peace	−.08	.01	Cheerful	−.05	−.09
A world of beauty	−.10*	−.10	Clean	.20***	.33***
Equality	−.20***	−.16**	Courageous	.05	.16**
Family security	.42***	.47***	Forgiving	−.01	−.02
Freedom	−.23***	−.13*	Helpful	.07	−.00
Happiness	.07	−.10	Honest	.13**	−.00
Inner harmony	.19***	−.03	Imaginative	−.21***	−.27***
Mature love	−.07	.03	Independent	−.10*	−.05
National security	.16***	.31***	Intellectual	−.04	−.08
Pleasure	−.19***	−.10	Logical	.01	.02
Salvation	.13**	.09	Loving	−.10*	−.06
Self-respect	.23***	.19***	Obedient	.01	.00
Social recognition	−.12**	−.14*	Polite	.22***	.15**
True friendship	−.18***	−.34***	Responsible	.13**	−.01
Wisdom	.09*	−.01	Self-controlled	−.07	−.00

Note. Because of missing data, the *N*s for the correlations in this table for Sample 1 (metropolitan Adelaide) ranged from 562 to 574. For Sample 2 (Flinders University), the *N*s ranged from 357 to 358.
* $p < .05$. ** $p < .01$. *** $p < .001$.

be associated with lower levels of education and income. However, conservatism was not significantly related to these socioeconomic measures in Sample 2, probably for the reasons presented earlier.

While the present study is essentially descriptive rather than based upon a causal model, the results do advance our knowledge about general conservatism. In the first place, they indicate the values that relate to this attitude syndrome and thereby provide a reliable set of findings to be explained. One can agree with Rokeach (1973) that there is a compelling need to assemble descriptive information about attitude–value relationships while at the same time seeking explanations about how these relationships came about. These explanations are likely to be complex rather than simple, having regard not only to the impact of different social institutions and environments but also to the active, constructive way in which information is processed by individuals according to the structure and dynamics of personality.

Second, the results of the regression analyses do have some implications about likely causal mediators. The fact that certain values emerged as strong predictors of conservatism even when the background and demographic variables were included in the regression equation (see Tables 4 and 5) suggests that the obtained relationships between value priorities and conservatism were mediated by more than social learning and probably also involved the operation of internal personality dynamics that helped shape a system of relatively consistent attitudes and values. Thus, these results can be understood in terms of both the cognitive learning and the psychodynamic explanations of attitude syndromes such as authoritarianism and conservatism that have been discussed by several authors (e.g., Elms, 1976; Greenstein, 1969; Sanford, 1973). These explanations relate authoritarianism and conservatism to learned conceptions of reality and ways of adjustment that are shaped by the actual conditions of a person's life and to intrapsychic determinants involving the expression, satisfaction, and control of basic needs and the operation of complex ego mechanisms that are in the service of the self.

That both kinds of explanation are necessary is especially suggested by the results presented in Tables 4 and 5, where the best set of predictors of general conservatism involved a combination of values and age, with sex and socioeconomic background entering much further down the list. Age is a variable that acquires meaning only in terms of its correlates. Among these correlates are the experiences—some unique, some shared with others—that would determine changes in both value priorities and conservatism over time. Even with age included as a variable in the regression equation, however, certain values remained as important predictors of conservatism, and these values, involving a preference for authority, rules, safety, and religion and a downgrading of equality, love, and open-minded modes of thought, are consistent with past discussions of conservative and authoritarian attitudes and ideologies (e.g., Adorno et al., 1950; Elms, 1976; Greenstein, 1969; Kirscht & Dillehay, 1967; Rokeach, 1960; Sanford, 1973; Tomkins, 1963; Wilson, 1973). In view of the fact that these values emerged as important predictors after age and the other background variables were controlled, it is reasonable to assume that personality dynamics relating to impulse satisfaction and control and their effects on cognition also played an important role in determining the nature of the value/conservatism relationships.

Given the relationships between values and conservatism described in the present study, longitudinal investigations that enable one to separate out cause and effect are needed in the future. It is nearly 30 years since *The Authoritarian Personality* was published (Adorno et al., 1950). One can now improve upon the procedures used in that seminal research in many ways, not least by employing methodologies that have emerged from life span developmental psychology and go beyond simple cross-sectional and single-cohort longitudinal studies in an attempt to discover the complex influences involved (Feather, 1975b, pp. 143–146). These new procedures are costly and time consuming,

but they are essential if one is to disentangle the factors determining change and continuity in values and attitudes over time. Using these improved methodologies, one could design developmental studies that explore the basis for relationships between values and conservative attitudes across the life span and enable the formulation of more sophisticated theoretical analyses.[8]

Not enough is known about the conditions under which general values emerge and how they come to be related to specific attitudes. It is likely that the sequence of events is complex and that attitudes tied to specific objects and issues develop first, followed by the emergence of general values as abstractions from the welter of a person's experience as it reflects the attitudes and actions both of self and of significant others (see Adelson, 1975; Feather, 1978, in press). Once formed, the dominant values may then inform a person regarding the correct attitudes toward new objects and issues. The hierarchy of values would itself change over the life span, some values increasing in relative importance and others decreasing. And new beliefs and attitudes would also be acquired, concomitant with these general changes and based upon similar causes. This new learning, along with the modifications in the value hierarchy, would not only reflect the inevitable changes that occur across the life span as people grow older, take on new roles and responsibilities, and develop new needs and priorities as they cope with a changing biology and the altered circumstances of their lives; they would also be influenced by general social trends and unique historical events (Feather, 1975b; 1977a; in press). With change in the hierarchy of a person's values and the formation of new beliefs and attitudes, one would also expect to find alteration in some of the more specific orientations laid down in the past, so that the total set of beliefs, attitudes, and values forms a relatively consistent and interrelated system in general harmony with a person's overall self-concept. All of these matters are of obvious relevance to clarifying the basis of the relationships between values and conservatism and deserve detailed investigation in the future.

The results of the present study indicate which values such studies might focus upon. One can agree with Rokeach (1973) that freedom and equality are two terminal values of special significance in the analysis of differences in *political* conservatism. Indeed, for a number of reasons, Rokeach (1973, pp. 165–170) doubts the usefulness of employing a single continuum of liberalism–conservatism, preferring instead to refer to variations in political ideologies that can be defined by differences in the relative importance assigned to the two terminal values freedom and equality. In terms of his two-value model of political ideology (Rokeach, 1973, p. 170), high scorers on the C Scale could be seen as more likely to demonstrate a "profascist" ideology (freedom low, equality low) and low scorers a "prosocialist" ideology (freedom high, equality high), given the fact that total conservatism scores were negatively correlated with the relative importance assigned to both freedom and equality on the Value Survey (see Table 1). The present results show, however, that several other values were associated with general conservatism as defined by the C Scale, and these deserve future investigation also, along with further clarification of the concept of conservatism and the extent to which it is unitary or multidimensional.

---

[8] For an excellent recent example in which these new methodologies have been applied to the study of political orientations, see Jennings and Niemi (1975).

References

Adelson, J. The development of ideology in adolescence. In S. E. Dragastin & G. H. Elder (Eds.), *Adolescence in the life cycle: Psychological change and social context*. New York: Wiley, 1975.

Adorno, T. W., Frenkel-Brunswik, E., Levinson, D. J., & Sanford, R. N. *The authoritarian personality*. New York: Harper, 1950.

Boshier, R. To rotate or not to rotate: The question of the Conservatism Scale. *British Journal of Social and Clinical Psychology, 1972, 11,* 313–323.

Cohen, J., & Cohen, P. *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, N.J.: Erlbaum, 1975.

Elms, A. C. *Personality in politics.* New York: Harcourt Brace Jovanovich, 1976.

Eysenck, H. J. *The psychology of politics.* London: Routledge & Kegan Paul, 1954.

Eysenck, H. J. The structure of social attitudes. *British Journal of Social and Clinical Psychology,* 1975, *14,* 323–331.

Feather, N. T. Factor structure of the conservatism scale. *Australian Psychologist,* 1975, *10,* 179–184. (a)

Feather, N. T. *Values in education and society.* New York: Free Press, 1975. (b)

Feather, N. T. Generational and sex differences in conservatism. *Australian Psychologist,* 1977, *12,* 76–82. (a)

Feather, N. T. Value importance, conservatism, and age. *European Journal of Social Psychology,* 1977, *7,* 241–245. (b)

Feather, N. T. Family resemblances in conservatism: Are daughters more similar to parents than sons are? *Journal of Personality,* 1978, *46,* 260–278.

Feather, N. T. Values in adolescence. In J. Adelson (Ed.), *Handbook of adolescent psychology.* New York: Wiley, in press.

Greenstein, F. I. *Personality and politics: Problems of evidence, inference, and conceptualization.* Chicago: Markham, 1969.

Hays, W. L. *Quantification in psychology.* Belmont, Calif.: Brooks-Cole, 1967.

Jennings, M. K., & Niemi, R. G. Continuity and change in political orientations: A longitudinal study of two generations. *American Political Science Review,* 1975, *69,* 1316–1335.

Kirscht, J. P., & Dillehay, R. C. *Dimensions of authoritarianism: A review of research and theory.* Lexington: University of Kentucky Press, 1967.

Kirton, M. J. Wilson and Patterson's Conservatism Scale: A shortened alternative form. *British Journal of Social and Clinical Psychology,* 1978, *17,* 319–323.

Lipset, S. M. *Political man.* New York: Doubleday, 1960.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., & Bent, D. H. *SPSS statistical package for the social sciences.* New York: McGraw-Hill, 1975.

Ray, J. J. Conservatism, authoritarianism, and related variables: A review and empirical study. In G. D. Wilson (Ed.), *The psychology of conservatism.* New York: Academic Press, 1973.

Robertson, A., & Cochrane, R. The Wilson-Patterson Conservatism Scale: A reappraisal. *British Journal of Social and Clinical Psychology,* 1973, *12,* 428–430.

Rokeach, M. *The open and closed mind.* New York: Basic Books, 1960.

Rokeach, M. *The nature of human values.* New York: Free Press, 1973.

Sanford, N. Authoritarian personality in contemporary perspective. In J. N. Knutson (Ed.), *Handbook of political psychology.* San Francisco: Jossey-Bass, 1973.

Tomkins, S. Left and right: A basic dimension of ideology and personality. In R. W. White (Ed.), *The study of lives.* Chicago: Aldine Atherton, 1963.

Wilson, G. D. Interpretation of the factor structure of the C Scale: A reply to Boshier. *British Journal of Social and Clinical Psychology,* 1972, *11,* 324–325.

Wilson, G. D. (Ed.). *The psychology of conservatism.* New York: Academic Press, 1973.

Wilson, G. D. *Manual for the Wilson-Patterson attitude inventory (W.P.A.I.).* London: National Foundation for Education Research, 1975.

Wilson, G., & Patterson, J. R. A new measure of conservatism. *British Journal of Social and Clinical Psychology,* 1968, *7,* 264–269.

Received October 18, 1978 ■

Springer Publishing Company

The culminating statement of 50 years research in personality!

PERSONALITY AND LEARNING THEORY (2 Vols.)

Raymond B. Cattell, Ph.D., D. Sc.

Vol. 1 The Structure of Personality in Its Environment

Here is the first substantial synthesis linking the study of personality with the control tenets of learning theory. Through multivariate experimental methods Cattell has devised a system by which personality can be captured with mathematical models that operationalize theories of personality structure. Volume 1 relates personality, ability, dynamic traits and states to the dimensions of stimuli, environmental situation, and observer.

This first volume integrates the work of a hundred different researchers, raising statistical and structural concepts of traits, types, and processes to a new level of precision. Using original conceptual developments on integrated and unintegrated motivation measurement, conflict theory, the dynamic lattice, and emotional decision theory, Cattell reexamines and refines known dynamic traits. He includes discussions on trait and perturbation theory, laws of error, psychological relativity, observer distortion, and instrument factors. Volume 1 treats with unusual thoroughness the recently focussed problem of proper incorporation of situation, observer, and other aspects of the environment in personality prediction.

The unique feature of this treatise on personality, at this juncture in psychology, is the completeness of Cattell's theoretical model which presents an embracing and consistent theory of personality and its actions. The first volume will provide a challenging text in graduate courses in personality and an ideal supplement to any graduate course in psychometry.

Vol. **2 A Systems Theory of Maturation and Structured Learning** will be available from Springer in 1980. This two volume work is a landmark of scholarship in psychology.

Vol. 1: 448 pp. 1979 $29.00 hardcover

Cognitive Orientation and Behavior

Hans Kreitler, Ph.D. and Shulamith Kreitler, Ph. D.

Presents a theory of cognitive orientation that describes the chain of cognitive processes from input identification to intent and behavior. This theory provides a model for reliable prediction and modification of behavior. The theory and methods of cognitive orientation are equally pertinent for the practitioner and the researcher in clinical, educational, and social psychology.

"...a great service in showing...that a cognitive psychology is really plausible, but also... leads to a lot of interesting research, which they report." *– David C. McClelland*

447 pp. 1976 $27.50 hard

Family Constellation

Its Effect on Personality and Social Behavior, 3rd. Ed.

Walter Toman, Ph.D.

Completely rewritten, and with expanded research evidence, this well-known book describes how sibling position influences personality and shapes the individual's relationships within and outside the family.

344 pp. 1976 $15.50 hard $8.95 soft

Personality Changes in Aging

A Longitudinal Study of Community Residents

Joseph H. Britton, Ph.D. and Jean O. Britton, Ph.D.

Based on nine years intensive research, the book sheds light on the different rates at which individuals decline with advancing years.

236 pp. 1972 $9.50 hard

Who Will Marry Whom?

Theories and Research in Marital Choice

Bernard I. Murstein, Ph.D.

"Significant insights into the innumerable variables that could potentially influence marital choice."

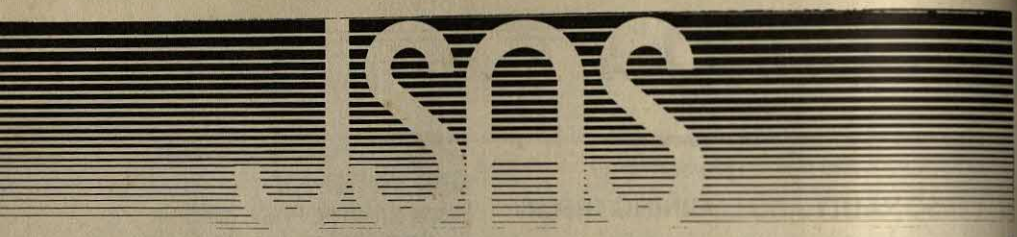
– J. of Personality Assessment

392 pp. 1976 $17.95 hard

For information, or To Order, write:

Springer Publishing Company
200 Park Ave. South, Dept. JPSP99
New York, N.Y. 10003

P19



The Journal Supplement Abstract Service is an information service providing ready access to psychology-related materials not available through conventional communication channels.

JSAS Selects and Disseminates:

- Educational materials
- Massive data collections
- Methodological techniques and procedures
- Major projects in progress
- Descriptions of effective techniques or programs
- Technical reports
- Fresh looks at controversial issues
- Invited lectures
- Management of psychological resources
- APA task force reports
- Bibliographies
- Literature reviews

Manuscripts are accepted by a panel of editors. They meet the same high standards set for American Psychological Association's journals, but are not bound by conventional restrictions of format, length, or editorial policy.

Available titles are listed in the quarterly JSAS *Catalog of Selected Documents in Psychology*. The catalog includes abstracts of newly accepted manuscripts along with manuscript length, price, and ordering information. Each issue of the catalog contains about 50 abstracts. Full-text manuscripts may be purchased individually in paper or microfiche.

Subscription rates for the quarterly catalog are $6 for APA members and $14 for nonmembers (foreign, $16). Please include full remittance for orders of $25 or less. To order, make check payable to APA and mail to:

American Psychological Association
Subscription Dept.
1200 17th Street, NW
Washington, DC 20036

(Continued from front cover)

The Effects of Behavioral Intentions and Consequences on Judgments of the Actor and Other: An S-V-O Analysis 1561
David A. Kravitz and Robert S. Wyer, Jr.

Self-Directed Attention, Awareness of Bodily States, and Suggestibility 1576
Michael F. Scheier, Charles S. Carver, and Frederick X. Gibbons

Attribution of Responsibility to the Self and Other in Children and Adults 1589
F. Fincham and J. Jaspars

Multiple Discovery and Invention: Zeitgeist, Genius, or Chance? 1603
Dean Keith Simonton

Value Correlates of Conservatism 1617
N. T. Feather

Instructions for Submission of Manuscripts 1486

New JPSP Format Delayed Until April 1446

SOCIAL EXCHANGE IN DEVELOPING RELATIONSHIPS

Edited by ROBERT L. BURGESS and TED L. HUSTON

FROM THE PREFACE: Our intent in preparing this book was to encourage the systematic study of the development of relationships. Throughout, several theoretical perspectives are presented, including evolutionary theory, cognitive developmental theory, personality theory, role theory, equity theory, and attribution theory. In each case, however, the authors address the issue of exchange in developing relations. Their views of the exchange process differ as a function of their own theoretical perspectives. Such is the state of the art. . . .

SECTION HEADINGS: Introduction. The Developmental Course of Close Relationships. Beyond the Dyad: Approaches to Explaining Exchange in Developing Relationships. Epilogue. References are included at the end of each chapter.

1979, 432 pp., $24.00 ISBN: 0-12-143550-4

ROLE THEORY: Expectations, Identities, and Behaviors

By BRUCE J. BIDDLE

FROM THE PREFACE: This book constitutes an integrative exposition of role theory. As is well known, the role perspective arose coincidentally in several disciplines in the late 1920's and early 1930's. Its central concern has been with patterns of human conduct—*roles;* with expectations, identities, and social positions; and with context and social structure as well as with individual response. As such, role theory offers a perspective that bridges social psychology, sociology, and anthropology. Thus, it has generated interest among social scientists from many backgrounds.

This work provides a long-needed synthesis for role theory. A single overarching conceptual system is set for the role field which embraces other uses. Research involving role concepts is reviewed. Applications of role concepts are discussed for education, social work, and clinical practice. Advantages and shortcomings of the role stance are examined.

1979, 380 pp., $19.50 ISBN: 0-12-095950-X

FOUNDATIONS OF DIALECTICAL PSYCHOLOGY

By KLAUS RIEGEL

CHAPTER HEADINGS: The Dialectics of Human Development. Historical Introduction. Dialectical Operations: The First and Final Period of Cognitive Development. The Relational Basis of Language. The Temporal Organization of Dialogues. The Recollection of the Individual and Collective Past. Adult Life Crises. The Dialectics of Time. References. Author Index. Subject Index.

1979, 216 pp., $15.00 ISBN: 0-12-588080-4

Send payment with order and save postage and handling charge.
Prices are subject to change without notice.

U.S. customers please note: On prepaid orders—payment will be refunded for titles on which shipment is not possible within 120 days.

Academic Press, Inc.

A Subsidiary of Harcourt Brace Jovanovich, Publishers

111 FIFTH AVENUE, NEW YORK, N.Y. 10003
24-28 OVAL ROAD, LONDON NW1 7DX

Please send me the following:

___copies, *Burgess/Huston*, Social Exchange in Developing Relationships, *$24.00*
___copies, *Biddle*, Role Theory: Expectations, Identities, and Behaviors, *$19.50*
___copies, *Riegel*, Foundations of Dialectical Psychology, *$15.00*

Check enclosed___ Bill me___

Please charge this order to my
___American Express ___Visa
___Master Charge ___Diner's Club

Card No. _____
Expiration Date _____
Signature _____
(*All orders subject to credit approval*)

NAME_____

ADDRESS_____

CITY/STATE/ZIP_____

CA and NY State residents please add sales tax. Direct all orders to Georgia Phillips, Media Dept.

JPerSocPsy/9/79

JOURNAL OF

Personality AND Social Psychology

VOLUME 37 NUMBER 10 · OCTOBER 1979

Psychological Androgyny and Sex Role Flexibility: A Test of Two Hypotheses 1631
Robert L. Helmreich, Janet T. Spence, and Carole K. Holahan

Nonverbal Trait Inference 1645
Sampo V. Paunonen and Douglas N. Jackson

The Role of Category Accessibility in the Interpretation of Information About Persons: Some Determinants and Implications 1660
Thomas K. Srull and Robert S. Wyer, Jr.

Negative and Positive Components of Psychological Masculinity and Femininity and Their Relationships to Self-Reports of Neurotic and Acting Out Behaviors 1673
Janet T. Spence, Robert L. Helmreich, and Carole K. Holahan

Motives for Social Comparison: The Construction–Validation Distinction 1683
Russell H. Fazio

The Social Learning Theory of Sex Differences: Imitation Is Alive and Well 1699
David G. Perry and Kay Bussey

Telling Lies 1713
Bella M. DePaulo and Robert Rosenthal

Attributional Information Processing: A Response Time Model of Causal Subtraction 1723
Eliot R. Smith and Frederick D. Miller

The Effects of Self-Esteem, Success–Failure, and Self-Consciousness on Task Performance 1732
Joel Brockner

A New Round Robin Analysis of Variance for Social Interaction Data 1742
Rebecca M. Warner, David A. Kenny, and Michael Stoto

The Independence of Evaluative and Item Information: Impression and Recall Order Effects in Behavior-Based Impression Formation 1758
Elizabeth K. Dreben, Susan T. Fiske, and Reid Hastie

Self-Focus, Felt Responsibility, and Helping Behavior 1769
Shelley Duval, Virginia Hensley Duval, and Robert Neely

(Continued on inside back cover)

Published monthly by the

AMERICAN PSYCHOLOGICAL ASSOCIATION

ISSN 0022-3514

Acting Editor

Clyde Hendrick, *University of Miami*

Action Editors

Irwin Altman, *University of Utah*
Charles S. Carver, *University of Miami*
Richard J. Harris, *University of New Mexico*
John H. Harvey, *Vanderbilt University*
William Ickes, *University of Missouri–St. Louis*
George Levinger, *University of Massachusetts—Amherst*
Brendan Gail Rule, *University of Alberta, Edmonton, Canada*
David R. Shaffer, *University of Georgia*
Cookie White Stephan, *New Mexico State University*
Barbara Strudler Wallston, *George Peabody College for Teachers*

Consulting Editors

Robert Steven Baron, *University of Iowa*
Paul A. Bell, *Colorado State University*
Ellen Berscheid, *University of Minnesota (Minneapolis)*
Michael H. Birnbaum, *University of Illinois at Urbana-Champaign*
Jack Brehm, *University of Kansas*
Philip Brickman, *Northwestern University*
J. Merrill Carlsmith, *Stanford University*
Paul Christopher Cozby, *California State University, Fullerton*
Rick Crandall, *Texas Christian University*
James M. Dabbs, Jr., *Georgia State University*
John M. Darley, *Princeton University*
Andrew R. Davidson, *University of Washington*
James H. Davis, *University of Illinois at Urbana-Champaign*
Kay Deaux, *Purdue University*
Marshall Dermer, *University of Wisconsin—Milwaukee*
Edward Donnerstein, *Iowa State University*
Ebbe B. Ebbesen, *University of California, San Diego*
Yakov M. Epstein, *Rutgers—The State University*
Irene Hanson Frieze, *University of Pittsburgh*
J. Martin Giesen, *Mississippi State University*
Lewis R. Goldberg, *University of Oregon*
William G. Graziano, *University of Georgia*
Alan E. Gross, *University of Maryland (College Park)*
Nancy Hirschberg, *University of Illinois at Chicago Circle*
Chester A. Insko, *University of North Carolina at Chapel Hill*
Alice M. Isen, *University of Maryland, Baltimore County*
Martin F. Kaplan, *Northern Illinois University*
Samuel S. Komorita, *University of Illinois at Urbana-Champaign*
Vladimir J. Konečni, *University of California, San Diego*
Ellen Langer, *Harvard University*
Mark R. Lepper, *Stanford University*
Darwyn E. Linder, *Arizona State University*
Dale T. Miller, *University of British Columbia, Vancouver, Canada*
Judson Mills, *University of Maryland (College Park)*
Warren T. Norman, *University of Michigan (Ann Arbor)*
Michael Ross, *University of Waterloo, Waterloo, Canada*
Michael F. Scheier, *Carnegie-Mellon University*
Barry R. Schlenker, *University of Florida*
Kelly G. Shaver, *College of William & Mary*
Phillip Shaver, *New York University*
Marvin E. Shaw, *University of Florida*
Steven J. Sherman, *Indiana University*
Harold Sigall, *University of Maryland (College Park)*
Mark Snyder, *University of Minnesota (Minneapolis)*
Melvin L. Snyder, *University of Texas at Austin*
Shelley E. Taylor, *Harvard University*
Stephen G. West, *Florida State University*
Steve Worchel, *University of Virginia*
Mark P. Zanna, *University of Waterloo, Waterloo, Canada*

Assistant to the Editor

Sally Miller

APA Journal Staff

Anita DeVivo, *Executive Editor*
Ann I. Mahoney, Manager, *Journal Production*
Marjorie M. Henry, *Production Supervisor*
Robert J. Hayward, *Advertising Representative*
Juanita Brodie, *Subscription Manager*

The Journal of Personality and Social Psychology publishes original reports of research, methodology, theory, criticism, and literature review in the fields of personality and social psychology of normal processes and populations. Submitted manuscripts should delineate an important question and provide a clear answer to it. In the case of research reports, this will often require demonstration that the reported results are generalizable beyond a single research setting. Consult the editorial in the January 1976 issue (Vol. 33, pp. 1–7) for a detailed statement of policies and instructions to authors.

Manuscripts: Submit manuscripts in quadruplicate, prepared according to the *Publication Manual of the American Psychological Association* (2nd ed.). Copies should be highly legible and will not be returned to authors unless they are original typewritten copies. Mail manuscripts according to instructions on p. 1439 of the August issue.

Required supplementary information: Manuscripts cannot be reviewed for publication unless the author provides supplementary information in the following categories: (a) a cover letter that indicates the relation (or the lack of relation) of the submitted manuscript to other papers by the same author(s) that are currently submitted, in press, or recently published; (b) one copy of each such related paper; (c) one copy of a statement of the procedures of the research reported in the submitted manuscript in sufficient detail to serve as a guide for conduct of an exact replication; and (d) one copy of the summary tables (e.g., correlation tables, analysis of variance summaries) for the *major* analyses reported. If it appears that the supplementary materials in categories (c) or (d) might be bulky, the author should exercise judgment and restaint in selecting what to send. It is not necessary to send supplementary material that repeats verbatim manuscript content.

After a manuscript is accepted, authors will be expected to provide assurance that their data are available in unanalyzed form for reanalyses by other investigators and that their research procedures were consistent with the American Psychological Association's published principles of research ethics, except insofar as may be detailed otherwise in the manuscript.

Copyright and Permission: All rights reserved. Written permission must be obtained from the American Psychological Association for copying or reprinting text of more than 500 words, tables, or figures. Permission is normally granted contingent upon like permission of the author, inclusion of the APA copyright notice on the first page of reproduced material, and payment of a fee of $10 per page, table, or figure. Abstracting is permitted with credit to the source. Permissions and fees are waived for authors who wish to reproduce their own material for personal use; fees only are waived for authors who wish to use their own material commercially. Permission and fees are waived for the photocopying of isolated articles for nonprofit classroom or library reserve use by instructors and educational institutions. Libraries are permitted to photocopy beyond the limits of U.S. copyright law: (1) those post-1977 articles with a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through the Copyright Clearance Center, P.O. Box 765, Schenectady, NY 12301; (2) pre-1978 articles, provided that the per-copy fee stated in the Publishers' Fee List is paid through the Copyright Clearance Center, P.O. Box 765, Schenectady, NY 12301. Address requests for reprint permission to the Permissions Office, APA, 1200 Seventeenth Street, N.W., Washington, D.C. 20036.

Subscriptions: Subscriptions are available on a calendar year basis only (January through December). Nonmember rates for 1979: $60 domestic, $64 foreign, $6 single issue. APA member rate: $20. Write to Subscription Section, APA.

Change of Address: Send change of address notice and a recent mailing label to the attention of the Subscription Section, APA, 30 days prior to the actual change of address. APA will not replace undelivered copies resulting from address changes; journals will be forwarded only if subscribers notify the local post office in writing that they will guarantee second-class forwarding postage.

Reprints: Authors may order reprints of their articles from the printer when they receive proofs.

Back Issues and Back Volumes: For information regarding back issues or back volumes write to Order Department, APA.

Microform Editions: For information regarding microform editions write to any of the following: Johnson Associates, Inc., P.O. Box 1017, Greenwich, Connecticut 06830; University Microfilms, Ann Arbor, Michigan 48106; or Princeton Microfilms, Princeton, New Jersey 08540.

Published monthly in one volume per year by the American Psychological Association, Inc., 1400 North Uhle Street, Arlington, Virginia 22201. Printed in the U.S.A. Second-class postage paid at Arlington, Va., and at additional mailing offices.

JOURNAL OF
Personality
AND
Social Psychology

Volume 37 October 1979 Number 10

Psychological Androgyny and Sex Role Flexibility: A Test of Two Hypotheses

Robert L. Helmreich, Janet T. Spence, and Carole K. Holahan
University of Texas at Austin

In a conceptual replication and extension of a study by Bem and Lenney, male ($n = 90$) and female ($n = 118$) college students rated their comfort in and preference for performing several series of masculine, feminine, and neutral activities. Correlations between these ratings and scores on the masculinity (instrumentality) and femininity (expressiveness) scales of the Personal Attributes Questionnaire (PAQ) of Spence and Helmreich tended to be theoretically reasonable in sign but in each sex were low in magnitude and only occasionally significant. Classification of subjects into four PAQ groups (androgynous, masculine, feminine, and undifferentiated) on their joint masculinity and femininity scores revealed that androgynous and masculine subjects of both sexes had higher comfort ratings, independent of type of task, than did feminine and undifferentiated subjects, suggesting the importance of instrumentality and expressiveness per se. For forced-choice preference ratings, significant differences were found only in males, masculine subjects having a stronger preference for sex-typed tasks than those in other categorical groups. The PAQ variable, however, accounted for only a small percentage of the variance. The data support the Spence–Helmreich hypothesis that the PAQ and other similar instruments are largely measures of instrumental and expressive personality traits rather than sex roles and that these personality dimensions are only minimally related to many sex role behaviors.

Interest in the concept of psychological androgyny has generated an increasing volume of empirical research, much of it using the Bem Sex Role Inventory (BSRI: Bem, 1974, 1977) or the Personal Attributes Questionnaire (PAQ: Spence & Helmreich, 1978; Spence, Helmreich, & Stapp, 1974, 1975) as the operational measure of the construct. Each of these instruments contains separate masculinity and femininity scales that are essentially orthogonal. The androgynous individual, according to current usage, is defined as one who scores relatively high on both the masculinity (M) and femininity (F) scales of these self-report instruments.

While the developers of the BSRI and the PAQ share certain theoretical assumptions that have guided both their empirical inves-

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3710-1631$00.75

Draw No. 134
Date 7.3.80
File No. librarian
Research

tigations and their interpretation of the results, their views differ in other critical respects. These differences, in turn, lead to predictions about the behavior of "androgynous" individuals that do not always coincide. The present study, using the PAQ, involved conditions in which the implications of the two theories led to different expectations. In order to explain its rationale, it is first necessary to describe the empirical properties of both the PAQ and the BSRI.

The PAQ M scale consists of a series of descriptions of traits, each stereotypically more characteristic of men than women but socially desirable to some degree in both sexes. Similarly, the F scale consists of descriptions of socially desirable traits more characteristic of women. In content, the trait clusters reflect what Parsons and Bales (1955) have described as instrumental (masculine) and expressive (feminine) characteristics. Factor analyses of the PAQ reproduce the two empirically derived scales, thus adding further justification to their separation (Spence & Helmreich, 1979; Helmreich, Spence, & Wilhelm, in press). The scales have been demonstrated to discriminate between the sexes in diverse populations, varying widely in age, ethnicity, and social class (e.g., Spence & Helmreich, 1978), thus justifying use of the labels *masculinity* and *femininity*.

In sum, the PAQ is a quite conventional self-report measure of socioaffective instrumental and expressive traits, the items making no mention of instrumental or expressive behaviors or the situations that elicit them. These trait dimensions, according to our theoretical conception, are internally located response predispositions that combine with situational variables and other person variables to determine behavior but are not themselves identical to behavior.

Of the evidence that has been collected to support the construct validity of the PAQ scales as measures of instrumentality and expressiveness, a recent experiment by Klein (Note 1) is particularly instructive. Klein observed the dominance behavior of female college students in three-member problem-solving groups, two members of which were the experimenter's confederates. Subjects' scores on the PAQ M scale were positively correlated with degree of dominance behavior, whereas their scores on the F scale showed a slight negative correlation with this behavior. As might be expected from these correlations, masculine women (high in M, low in F) were most dominant, followed in order by women who were androgynous (high M, high F), feminine (low M, high F), and undifferentiated (low M, low F). This order was found both when the two confederates were female and when they were male. In all PAQ groups, however, level of dominance was markedly (and equally) reduced when the two confederates were male, a condition containing implicit role-related constraints on women's undisguised expression of this type of behavior. The results thus show that situational factors (presence or absence of sex role expectations) and personality traits (M and F) independently contributed to dominance behavior.

Like the PAQ, the BSRI consists of general trait descriptions, many of the M items being socially desirable instrumental traits and many of the F items being socially desirable expressive traits that are similar or identical to PAQ items on the parallel scales. However, not all items on the BSRI can be described as instrumental or expressive (e.g., "masculine" and "feminine") or (on the F scale) as socially desirable, and unlike the PAQ, recent factor analytic studies (e.g., Pedhazur & Tetenbaum, 1979) indicate a four- rather than a two-factor structure. Nonetheless, empirical evidence supporting the construct validity of the BSRI as (primarily) a measure of instrumentality and expressiveness has been found in behavioral studies by Bem and

Support for this research was provided by National Science Foundation Grant BNS 78-08911 (Janet T. Spence and Robert L. Helmreich, Principal Investigators). The paper was prepared while the second author was a Fellow at the Center for Advanced Study in the Behavioral Sciences with support from the Spencer Foundation, National Institute of Mental Health Grant 5T32MH14581-03, and The University of Texas Research Institute.

Requests for reprints should be sent to Robert L. Helmreich, Department of Psychology, The University of Texas at Austin, Austin, Texas 78712.

her colleagues (Bem, 1975, 1977; Bem, Martyna, & Watson, 1976).

The somewhat mixed content of the BSRI, the identification of the instrument as a sex role inventory, and Bem's theoretical discussions combine to suggest that she posits the existence of global constructs of masculinity and femininity that are manifested in a variety of gender-related behaviors, role attitudes, and personal qualities collectively identified as *sex roles*. She further appears to assume that, despite their limited empirical content, the BSRI and PAQ M and F scales are (in principle) valid measures of these global constructs. Thus, according to her conceptualization (e.g., Bem, 1977; Bem & Lenney, 1976), androgynous individuals, high in both M and F scores, are "behaviorally flexible" with respect to all manner of gender-related phenomena. As such, they are willing or able to exhibit masculine behaviors, feminine behaviors, or both, as situationally appropriate. Individuals with sex-typed personalities, on the other hand, will tend to avoid or exhibit lower levels of cross-sex-typed behaviors.

This theoretical position rests on the supposition that the empirically diverse indicators of masculinity and of femininity are all highly correlated, so that an individual exhibiting one set of attributes or behaviors in the class can reasonably be assumed to exhibit approximately the same degree of all other attributes and behaviors. If the subsets of masculine and feminine behaviors, attitudes, and personality characteristics that differentiate the sexes do covary within individuals, then it follows that it is useful to postulate global unidimensional concepts, and further, that it is justifiable to use instruments such as the PAQ or BSRI that tap a limited range of gender-related traits to make inferences about other kinds of masculine and feminine qualities.

Spence and Helmreich (1978) have not denied the existence of associations between various classes of masculine and feminine attributes and behaviors but have proposed that they often are weak and/or complexly determined. More particularly, they hypothesize that socially desirable masculine and feminine (i.e., instrumental and expressive) personality traits are not likely to be strongly related to sex role preferences or behaviors unless these behaviors call quite directly for instrumental or expressive skills. They thus suggest that personality instruments such as the PAQ and the BSRI cannot legitimately be substituted for more face valid measures of sex role attitudes and preferences, and that findings obtained with these personality instruments will not necessarily parallel those with sex role measures. For example, theoretically reasonable but very weak relationships have been found between the PAQ and BSRI M and F scales and the Attitudes Toward Women Scale (AWS: Spence & Helmreich, 1973), a measure of beliefs regarding appropriate roles of women vis-à-vis men (Orlofsky, Astin, & Ginsburg, 1977; Spence & Helmreich, in press). The values, however, are low and often nonsignificant even in very large samples. Similarly, the study by Klein (Note 1) cited earlier suggested that sex-typed (feminine) women, as measured by the PAQ, were no more responsive to the introduction of sex role demands into group interactions than were women showing other personality patterns.

In an experiment that went beyond the investigation of instrumental and expressive behaviors, Bem and Lenney (1976) tested the behavioral flexibility hypothesis by examining the relationship between masculine and feminine traits, as measured by the BSRI, and reactions to gender-related activities. Male and female subjects were asked to demonstrate a series of simple tasks (e.g., pounding a nail in a board) that were stereotypically masculine or feminine and then to indicate their degree of comfort in performing them. Prior to this portion of the study, they were given a description of a series of pairs of tasks, each pair consisting of either two neutral tasks or a masculine and a feminine task, and were asked to indicate which they preferred to demonstrate. Three groups of male and female subjects, designated as *masculine*, *feminine*, and *androgynous*, were preselected to serve in the experiment, assignment of subject to group being based on Bem's (1974) original subtractive (F

minus M) method of scoring the BSRI. Subjects were chosen from the center and the two extremes of the difference score distribution. The results partially supported the behavioral flexibility hypothesis, heavily sex-typed individuals (masculine men, feminine women) tending to show more avoidance of cross-typed tasks than did the other groups. Androgynous (balanced) subjects, however, did not differ significantly from cross-typed subjects.

The present study is a conceptual replication and extension of the Bem and Lenney study, using the PAQ rather than the BSRI. Male and female students, pretested on the PAQ and later given the Attitudes Toward Women Scale (Spence & Helmreich, 1973), were told that they would be asked to demonstrate several activities from a larger group and could indicate their preference beforehand. They were given two lists, each containing stereotypically masculine, feminine, and sex-neutral activities. The first list consisted of simple, everyday tasks similar to those used by Bem and Lenney (1976). The second list contained descriptions of TV commercials or public service announcements, half of which were affectively neutral and half negatively toned. In including a number of tasks that were mildly embarrassing or distasteful, we were exploring the possibility that stronger relationships might be found between reactions to negative items and the PAQ scales than those involving affectively neutral items. Two types of ratings were obtained. For each item, subjects were asked to indicate the degree of anticipated comfort in performing the named activity. They were also asked to give preference ratings within the groups of everyday tasks, affectively neutral announcements, and negative announcements. Subjects were unselected with respect to PAQ and AWS scores, thus permitting assessment of both the independent and joint effects of M, F, and AWS scores on the criterion measures as well as the magnitude of effects.

In deriving hypotheses about the results from Bem's work, several aspects of her theory and the Bem and Lenney study should first be noted. First, in response to data (Spence et al., 1975) indicating that a scoring scheme taking into account absolute level of M and F scores has greater utility than does a subtractive method, Bem (1977) subsequently revised her definition of androgyny to refer only to those relatively *high* in M and F. She also adopted the Spence et al. designation of undifferentiated to refer to those relatively *low* in M and F. Second, in discussing evidence bearing on the hypothesis that androgynous (high M, high F) individuals are more "behaviorally flexible" than others, Bem and others have seemed to treat the type of sex role behavior studied by Bem and Lenney as having equal standing with behaviors demanding the expression of instrumental or expressive qualities. That is, responses to the types of gender-related tasks employed by Bem and Lenney are implicitly regarded as being in the same class as instrumental or expressive behaviors and are therefore implied to be related to the BSRI in the same manner and with the same strength as instrumental and expressive behaviors. There is both an internal contradiction and incompleteness in her theory, however. Although the experimental evidence (e.g., Bem et al., 1976, as reanalyzed in Bem, 1977) indicates that androgynous and undifferentiated individuals are not identical in behaviors designed to reflect instrumental or expressive capacities, Bem (1977) has stated that she would expect *no* differences between those groups on the types of tasks employed by Bem and Lenney. (Presumably undifferentiated individuals, being neither strongly masculine nor strongly feminine in their "sex-role identification," would show little differential reaction to sex-linked tasks.) Thus, on *this type of task*, undifferentiated and androgynous groups are *both* expected to be more "behaviorally flexible" than sex-typed individuals. More specifically, her theory predicts that androgynous and undifferentiated groups, male and female, will not differ from each other but that both categorical groups will show less marked preference for same-sex tasks and will express more comfort about performing opposite-sex tasks than will sex-typed individuals (masculine men, feminine women). Bem is unclear about what to expect in cross-typed individ-

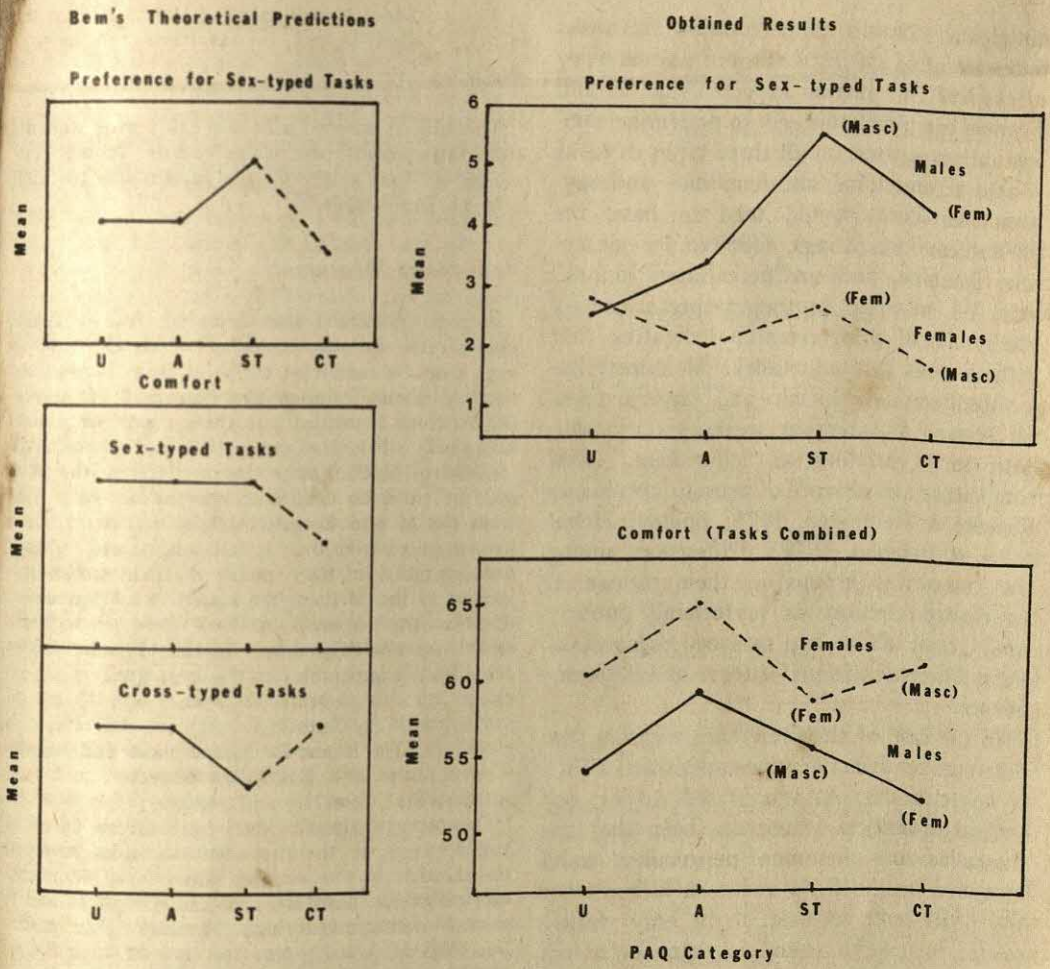


Figure 1. Predicted mean sex-typed preference and comfort ratings (Bem's hypotheses) in undifferentiated, androgynous, sex-typed, and cross-typed groups and obtained means for males and females in the groups. (U = undifferentiated; A = androgynous; ST = sex-typed [masculine male, feminine female]; CT = cross-typed [feminine male, masculine female]. The predicted preference and comfort ratings show the expected order of means. They do not reflect predicted magnitudes of effects. PAQ = Personal Attributes Questionnaire. Masc. = masculine. Fem. = feminine.)

uals (feminine men, masculine women). It would seem to follow from an extension of her logic, however, that these groups would also be less flexible than undifferentiated and androgynous groups, showing the mirror image in their preferences and comfort ratings of same- and opposite-sex tasks as masculine men and feminine women. The implications of these hypotheses for both comfort and preference ratings are shown graphically at the left of Figure 1.

These hypotheses, which also imply fairly substantial relationships between the PAQ scales and preferences for and comfort in performing sex-typed tasks, are to be contrasted with the expectations of the present investigators. We anticipated that while subjects' gender would be a major determinant of their reactions to sex-typed tasks (a point on which Bem is silent), the relationships within each sex between M and F and the criterion measures would be minimal and/or of a different form than predicted by Bem. More specifically, we expected that M scores, and to a lesser degree F scores, would be weakly but positively related to subjects' ratings of their

anticipated comfort in performing the tasks *independent of subjects' sex and the category of the task*. We further expected that M and F would combine additively to determine comfort ratings so that on all three types of tasks —neutral, masculine, and feminine—androgynous individuals would tend to have the highest comfort ratings, followed by masculine, feminine, and undifferentiated individuals. We were led to these expectations by the results of prior research indicating that in both sexes instrumentality (M scores) has a substantial relationship and expressiveness (F scores) a somewhat weaker relationship with such variables as self-esteem, social competence and absence of neurotic tendencies (Spence & Helmreich, 1978; Spence, Helmreich, & Holahan, 1979). Differences among the categorical groups in their ratings of anticipated comfort in performing publicly would come about, we reasoned, because of their differences in self-esteem and self-confidence.

In the case of the preference measure, the difference between our expectations and Bem's is less clear-cut. As was stated earlier, our general theoretical stance has been that the masculine and feminine personality traits measured by the PAQ and to a large extent the BSRI tend, at best, to be only weakly related to masculine and feminine behaviors and characteristics that do not directly implicate instrumentality and expressiveness. On a forced-choice preference measure, it seemed possible that *some* correlation might be found between M and F scores and preference scores, M being related positively and F negatively to preference for stereotypically masculine over stereotypically feminine tasks. These relationships were expected to be somewhat stronger in males than in females because of the greater pressures males experience for sex role conformity. Even in the case of males, however, we anticipated that the relationships would be weak, PAQ scores accounting for a small portion of the variance within each sex. We expected a stronger relationship, however, between preference rating and scores on our sex role attitudes measure (AWS), again, especially in males.

Method

Subjects

The subjects were 90 male and 118 female students from introductory psychology courses at the University of Texas at Austin who participated to fulfill a course requirement.

Assessment Measures

Personal Attributes Questionnaire. At the beginning of the semester, classes of which the subjects were members completed the short form of the Personal Attributes Questionnaire (Spence & Helmreich, 1978), which is divided into three eight-item scales: masculinity (M), femininity (F), and masculinity-femininity (M–F). (Since the results from the M–F scale provided no additional information, only data from the M and F scales will be reported.) Each item is scored from 0 to 4, and total scores for each scale are obtained by summing the item scores. Responses to the M items are scored in a "masculine" direction and responses to the F items in a "feminine" direction. The means on the M and F scales are highly comparable to the normative data for college students presented in Spence and Helmreich (1978), with significant ($p < .01$) sex differences on each scale. The means for M for males and females were 21.6 and 20.0. For F, the respective male and female means were 21.9 and 24.4.

One of the analyses of the joint contribution of M and F scores to the dependent variables involved classification of the subjects into one of four categorical groups, based on a median split of M and F scores: androgynous (high M, high F), masculine (high M, low F), feminine (low M, high F), or undifferentiated (low M, low F). The median scores, taken from norms for male and female college students (Spence & Helmreich, 1978), were 21 for the M scale and 23 for the F scale. The percentage of males and females in each of the four categories again closely paralleled the normative data in Spence and Helmreich (1978, p. 53). The breakdowns for males and females, respectively, were: androgynous, 30% and 32%; masculine, 33% and 13%; feminine, 13% and 31%; and undifferentiated, 21% and 24%.

Attitudes Toward Women Scale. At the conclusion of the testing session, subjects completed the 15-item form of the Attitudes Toward Women Scale (Spence & Helmreich, 1978), a questionnaire inquiring about the rights, roles, and privileges that women ought to have or be permitted. Respondents are asked to indicate their agreement with each statement on a 4-point scale ranging from *strongly agree* to *strongly disagree*. Items are scored 0–3, with high scores indicating more profeminist attitudes. A total score is obtained by summing the 15 item scores. Consistent with previous research, women subjects scored as significantly more liberal on the AWS ($p < .001$) and the means of 25.0 for males and 30.1 for females were highly similar to those found in

other college samples (Spence & Helmreich, 1978).

Activities questionnaire. An activities questionnaire was constructed to measure subjects' reactions to performing various masculine, feminine, and neutral activities. Two lists of activities were included. The first, labeled *everyday behaviors,* described a series of simple tasks comparable to those used by Bem and Lenney (1976). These included, in random order, 4 masculine behaviors, 4 feminine behaviors, and 4 neutral behaviors. The 12 tasks were rated by 10 male and 10 female student judges on 5-point scales labeled 1 = very feminine, 3 = non-sex-typed, 5 = very masculine. The results confirmed the adequacy of the classification. Items and mean ratings are shown in Table 1.

Subjects were asked to imagine themselves performing each activity and to indicate on a 7-point scale how comfortable they would be. The scale ranged from *very uncomfortable* to *very comfortable,* with high scores indicating greater comfort.

The second list, labeled *announcements,* was composed of 24 items referring to a TV announcement of public interest or a description of some product. Subjects were asked to imagine themselves reading a prepared message on each of the topics and to rate how comfortable they would feel, again on a 7-point scale. In content, the items were equally divided among masculine, feminine, and sex-neutral topics. Four of the announcements in each of the three categories were affectively neutral (e.g., feminine: "Point out the positive features of a particular style of purse while holding up the purse"). The other four of the activities in each category were somewhat embarrassing or negative in nature (e.g., masculine: "Point out on a model the positive features of a pair of jockey shorts"). The ratings of undergraduate judges also confirmed the assignment of the announcements to the three sex-typed categories. (Assignment to affectively negative and neutral categories was confirmed post hoc by the comfort ratings of the experimental subjects themselves.)

After making their comfort ratings, subjects were asked to indicate their preferences for performing the previously rated activities. The activities were presented in three separate lists. The first contained the 12 everyday behaviors, the second the 12 affectively neutral announcements, and the third the 12 negative announcements. For each list separately, subjects were asked to indicate the 4 activities they would most prefer to perform and the 4 activities they would least prefer to perform. A score of 3 was assigned to items marked as most preferred, a score of 2 to the unmarked items, and a score of 1 to the items marked as least preferred. For each list, each subject's scores for the 4 masculine, feminine, and sex-neutral items were summed to yield preference scores.

Embarrassment index. After demonstrating two tasks from the everyday behaviors list, one masculine and one feminine, subjects were asked to indicate on a 7-point scale how embarrassed they felt while performing each task. Since Bem and Lenney (1976) found that predemonstration ratings yielded as satisfactory results as postdemonstration ratings, this portion of the study was abbreviated and was conducted primarily to protect the cover story about the purpose of obtaining the initial comfort and preference ratings. Further, it was felt that the subjects' relief in not having to give any of the negative announcements might obscure any differential reactions to the masculine and feminine tasks that might otherwise have occurred. In support of this suspicion, an analysis of the embarrassment ratings revealed no significant differences in either sex. For this reason, no further reference will be made to these data.

Table 1
Everday Tasks by Category (Masculine, Neutral, Feminine) and Mean Ratings by Judges on Masculinity–Femininity

Category/item	Order of presentation	Mean rating
Masculine		
Oil a squeaky hinge on a metal box	2	4.30
Nail two boards together	4	4.60
Tighten a screw	5	4.50
Attach fishing tackle to a line	12	4.65
Neutral		
Play with a yo-yo	3	3.10
Sharpen a pencil	6	2.85
Peel an orange	7	2.95
Put a jigsaw puzzle together	9	3.00
Feminine		
Iron a cloth napkin	1	1.25
Set a table	8	1.50
Measure out a cup of flour	10	1.35
Sew a button to a piece of fabric	11	1.20

Note. Raters were 10 female and 10 male undergraduates.

Procedure

Subjects were tested in same-sex groups ranging in size from two to eight. Since Bem and Lenney (1976) reported stronger results for subjects tested by an opposite-sex experimenter, a pair of female undergraduate experimenters tested male subjects and a pair of male undergraduate experimenters tested female subjects.

Both experimenters were present at the beginning of the experimental session, one of them giving the initial instructions. Subjects were informed that they were participating in the initial phase of a study on impression formation designed to investigate whether observers' judgments about stimulus persons' personalities are influenced by the particular activity the latter are performing or how comfortable they are performing it. Subjects would therefore be asked to

rate a number of activities, several of which they would later carry out while being videotaped. Their data would be used to select activities for the later impression formation study. After a general description of the several types of activities had been given, subjects were informed that they would be given instructions about what to do or a statement to read while they were being videotaped. It was emphasized that there was no interest in how well they were doing the activity but that they were to try to become involved in what they were doing or saying, so that they would look believable.

The experimenter then told the subjects that they would be shown the whole list of tasks before the videotaping and asked for their reactions. The reason given for this procedure was to provide additional information about the tasks and to give the subjects some choice in what they would later be doing. The activities questionnaire was then administered. As each subject completed the activities questionnaire, the second experimenter took the subject into an adjoining room for videotaping.

Each subject was asked to perform a masculine and a feminine task from the list of everyday behaviors. One task was randomly selected from those the subject marked as most preferred and one was randomly selected from those the subject marked as least preferred. After the videotaping, each subject returned to the first room and completed the embarrassment questionnaire and the AWS.

Results

Everyday Behaviors

Comfort ratings. Summed comfort ratings were found for each subject on the masculine, feminine, and neutral tasks. Highly significant ($p < .001$) differences were found in each sex among the mean ratings for the three types of tasks, subjects rating sex-congruent tasks highest, neutral tasks intermediate, and cross-sex tasks lowest. Subjects' gender, then, was a strong determinant of reactions to the task. Within each type of task, however, correlations between the comfort ratings and PAQ M and F scores were all low positive, ranging in magnitude from .08 to .27. Four of these correlations were significant ($p < .05$): in males, the correlations between M and the masculine tasks, and in females, the correlations between M and F and the masculine tasks and between M and the feminine tasks.

This pattern of positive correlations provides preliminary support for our contention that within each type of task, those high in instrumentality and expressiveness would tend to give higher comfort ratings than others. A more direct confrontation between our hypothesis and Bem's is afforded by classifying subjects into categorical groups, based on their joint M and F scores. As illustrated in Figure 1, Bem's theory implies that on gender-incongruent tasks (masculine tasks for females, feminine tasks for males), individuals who are sex-typed in personality will show the lowest comfort ratings, whereas cross-typed individuals will presumably show the highest ratings. On gender-congruent tasks, however, all groups except the cross-typed are expected to show equally high ratings. Finally, no differences are predicted for the neutral tasks. In contrast, our expectations, drawn from self-esteem data, are that similar relationships occur in all three task categories in both sexes, androgynous subjects tending to have the highest means, followed by masculine and then feminine and undifferentiated individuals. Inspection of the means of the four categorical groups in both sexes on all three tasks showed good correspondence with the latter prediction, androgynous or masculine subjects having the highest ratings in all six comparisons and undifferentiated or feminine subjects the lowest.

The data from the masculine and feminine tasks were each analyzed by a 2 (Sex) × 4 (PAQ Category) analysis of covariance, using the neutral task ratings as the covariate. For both masculine and feminine tasks a highly significant ($p < .0001$) main effect was found for sex. For masculine tasks, the main effect of PAQ category was also significant ($p < .01$), but for feminine tasks, it was not ($p < .25$). Of particular importance, the Sex × Category interaction was nonsignificant ($F < 1$) in both analyses.

The pattern of the comfort data clearly does not support Bem's predictions of an interaction between sex, type of task, and PAQ category but corresponds quite closely to our expectations. Still further support for our hypotheses came from the finding that in both sexes, substantial positive correlations were found between the comfort measures on the three types of tasks (average $r = .67$).

The correlation between measures, in con-

Table 2
Correlations Among Independent and Dependent Measures

Subjects	Measure 1	2	3	4	5
Males					
1. Masculinity					
2. Femininity	.08				
3. AWS	−.17	.13			
4. Everday preference	.28***	−.08	−.49***		
5. Everday comfort	.17*	.14	.03	.10	
Females					
1. Masculinity		.16	.00	−.13	.25***
2. Femininity			−.08	−.04	.14
3. AWS				−.14	.17*
4. Everday preference					−.18**
5. Everday comfort					

Note. High preference scores indicate greater preference for gender-congruent tasks. AWS = Attitudes Toward Women Scale.
*$p < .1$. **$p < .05$. ***$p < .01$.

junction with the analyses showing similar relationships with M and F in all task categories, led us to combine ratings on the three types of tasks to yield a single comfort score. Subsequent analyses, undertaken to explicate further the relationships among comfort ratings, PAQ scores, and sex role attitude (AWS) scores were based on the combined comfort measure. Presentation of the results of these analyses will be delayed until after the preference measure has been described.

Preference ratings. It will be recalled that each of the four tasks rated as most preferred was given a score of 3, those rated as least preferred a score of 1, and those unrated a score of 2. The sum of the ratings on the masculine, feminine, and neutral tasks was then obtained for each subject. Subjects exhibiting perfect sex-related preferences would thus have scores of 12 for congruent sex-typed tasks (masculine tasks for males; feminine tasks for females) and of 4 for noncongruent cross-typed tasks; parallel scores for those showing perfect cross-typed preferences would be 4 and 12. In initial analyses, means on the three types of tasks were obtained for males and females, the results indicating that in both sexes, the sex-congruent tasks were most preferred, followed by the neutral tasks, and then the cross-typed tasks. Within each sex, these differences were highly significant ($p < .001$).

A single score, called *preference,* was next obtained for each subject by subtracting the summed preference ratings for cross-typed tasks from the summed ratings for sex-congruent tasks. Possible scores thus ranged from 8 (12 − 4) for those with perfect sex-congruent preferences to −8 (4 − 12) for those with perfect cross-typed preferences. These difference (preference) scores were used as the dependent variable in subsequent analyses.

Relationships among variables. Correlations among the independent and dependent variables are shown in Table 2. As found in previous studies, M and F show low positive relationships in each sex. The relationships of M and F with the AWS also confirm previous findings, the signs of the correlations being theoretically coherent but the magnitudes low (and in this instance, nonsignificant).

Correlations for the combined comfort scores parallel previous reports for the ratings of each type of task: both M and F are positively related to the comfort measure in both sexes. As expected, M (instrumentality) is more strongly related to comfort ratings than F (expressiveness).

To examine the joint influence of trait and role attitude variables on the comfort mea-

sure, separate stepwise multiple regressions were performed for males and females.[1] The interaction term (M × F) was also included with the constraint recommended by Cohen (1978) that it be entered after the main effect terms. The regressions are summarized at the bottom of Table 3. Instrumentality (M) was most strongly related to comfort ratings, followed by expressiveness (F), the effects being larger in females than males. The AWS made no significant contribution. (Separate regressions on comfort ratings for the three types of tasks produced identical results.)

As another approach to the joint contribution of M and F, the mean comfort ratings have been plotted for the four PAQ groups of each sex in Figure 1. Analysis of variance of these data shows a significant main effect for sex, $F(1, 200) = 15.11$, $p < .001$, with females reporting greater comfort. The main effect for PAQ category is also significant, $F(3, 200) = 4.01$, $p < .01$, whereas the interaction F is less than 1. In both sexes, androgynous subjects gave the highest comfort ratings, followed by the masculine and then the undifferentiated and feminine subjects. The contrast of androgynous with all others is significant ($p < .001$), and androgynous and masculine versus undifferentiated and feminine approaches significance ($p < .06$).

Turning to the preference data, inspection of the correlations in Table 2 shows that in males the signs of the correlations were in theoretically coherent directions, stronger preferences for sex-congruent (masculine) tasks being positively associated with M scores and negatively associated with F scores. As we expected, the relationships were low, reaching significance only for M ($r = .28$; $p < .05$). Also as predicted, the relationship between AWS and preference scores was stronger than for M or F ($r = -.49$), with males who were traditional in sex role attitudes showing the greatest preference for sex-congruent tasks. Turning to females, even smaller correlations with M and F are found, neither significant. The AWS shows the largest correlation ($-.14$) with preference and is in the predicted direction, but even this relationship is nonsignificant.

Stepwise multiple regressions were also performed on the preference data and are shown at the top of Table 3. As we predicted, attitude scores provided the best prediction of sex-typed preference in males, followed by M. No significant effects were found in females.

Finally, the mean preference scores were examined for those in each of the four PAQ categorical groups. These means are plotted at the top right of Figure 1, while the results predicted by Bem are shown at the left. The most striking finding to be observed is the markedly greater preference of males than females for sex-typed tasks in all categorical groups except the undifferentiated. Analysis of these data, using a 2 (Sex) × 4 (PAQ Category) analysis of variance, revealed that the main effect for sex is highly significant, $F(1, 200) = 20.60$, $p < .0001$, accounting for 74% of the between-groups variance. The main effect for PAQ category is significant, $F(3, 200) = 2.18$, $p < .01$, as is the interaction between sex and PAQ category, $F(3, 200) = 4.47$, $p < .01$, reflecting a different ordering of means in the two sexes.

Simple analyses of variance by PAQ category within each sex were then undertaken. For females, the F was less than 1, but for males it was highly significant ($p < .001$). Contrasts of the male data indicated that the masculine subjects were significantly ($p < .01$) more sex-typed in their preferences than the androgynous and undifferentiated subjects but did *not* differ significantly from cross-typed, feminine subjects. Since Bem predicts no difference between androgynous and undifferentiated individuals, these groups were also compared. A difference of borderline significance, $t(45) = 1.7$, $p = .15$, was found, reflecting the higher preference of androgynous males for sex-congruent tasks.[2]

---

[1] If regressions are computed including sex as a predictor, highly significant overall Rs are obtained for each variable reflecting the strong sex effect. However, since the signs of correlations with criteria differ in the two sexes, the separate regressions give a clearer picture of relationships.

[2] Bem and Lenney classified their subjects as masculine, feminine, or androgynous using Bem's (1974) subtractive technique based on the concept of balance between masculine and feminine attributes. To max-

Table 3
Regressions of Masculinity, Femininity, and AWS on Sex-Typed Preference and on Comfort Ratings

Step	Variable	Entry F	p	Multiple R	R^2 change	Overall F	p
			Sex-typed preference				
Males							
1	AWS	28.9	.00	.49	—	28.9	.000
2	M	4.7	.03	.53	.04	17.4	.000
3	F	<1	*ns*	.53	.00	11.5	.000
4	M × F	9.8	.002	.60	.07	12.0	.000
Females							
1	AWS	2.1	.15	.13	—	2.1	.15
2	M	1.9	.16	.18	.02	2.0	.14
3	F	<1	*ns*	.19	.01	1.4	.26
4	M × F[a]	—	—	—	—	—	—
			Comfort				
Males							
1	M	2.8	.09	.17	—	2.8	.09
2	F	1.5	.22	.22	.02	2.1	.12
3	AWS	<1	*ns*	.22	.00	1.5	.22
4	M × F	<1	*ns*	.22	.00	1.1	.36
Females							
1	M	7.7	.01	.25	—	7.7	.006
2	F	3.7	.05	.31	.03	5.8	.004
3	AWS	1.6	.20	.32	.01	4.4	.006
4	M × F	<1	*ns*	.33	.00	3.3	.010

Note. M = masculine. F = feminine. AWS = Attitudes Toward Women Scale.
[a] Data showed tolerance insufficient for inclusion.

An analysis of covariance was also performed on the male preference data, using AWS scores as the covariate. The term for AWS was highly significant ($F = 31.0$, $p < .001$), accounting for 65% of the explained variance and 23% of the total variance. The term for PAQ category was also significant ($F = 5.18$, $p < .01$), this variable accounting for 35% of the explained variance and 13% of the total variance.

Announcements. The data on announcements were collected primarily to determine whether the affective quality of sex-linked activities interacted with the PAQ measures in determining preference and comfort ratings. Although significant main effects were found for the positive and negative tasks within the masculine, feminine, and neutral categories in both preference and comfort ratings, no interactions involving category and the PAQ measures occurred. Accordingly, results for positive and negative tasks were combined, and analyses paralleling those performed with the everyday task data were conducted. Sex differences related to type of task were highly significant. The contribution of the PAQ scales to the criterion measures was somewhat weaker than was found with the everyday tasks data, but the overall pattern of the results for the comfort and preference ratings was similar. For this reason, they will not be discussed further.

imize comparability of analyses, subjects were also classified into three groups by applying Bem's procedure to PAQ M and F scores. The preference and comfort scores were then analyzed in 2 (Sex) × 3 (Masculine, Feminine, Androgynous) analyses of variance. The pattern of means for preference paralleled Bem and Lenney's results, whereas those for comfort did not. In both analyses of variance, however, only the sex main effect was significant.

Discussion

Mounting evidence suggests that the PAQ and the BSRI have both construct and predictive validity as measures of instrumental (masculine) and expressive (feminine) attributes. The question to which the present study is fundamentally addressed is whether the masculine and feminine attributes tapped by these psychometric instruments are both strongly and directly related to other masculine and feminine behaviors and attributes, as implied by Bem's theory, or whether these personality attributes tend to be only weakly and indirectly related to sex-related phenomena that do not quite immediately engage instrumental and expressive capacities, as suggested by the present investigators.

Overall, the data from the present study support the latter point of view. Analyses of the ratings of anticipated comfort in task performance showed, as we predicted, weak but positive correlations with M and F in *both* sexes on all three types of tasks, masculine, feminine, and neutral. Assessment of the joint contribution of M and F scores, conducted by assigning subjects to categorical group, indicated that androgynous individuals of both sexes expressed the greatest comfort (on all tasks combined), followed by masculine, undifferentiated, and feminine subjects. Except for the inversion of the undifferentiated and feminine groups (who were, in any event, very similar), this ordering corresponds to that consistently found with measures of self-esteem (e.g., Spence & Helmreich, 1978), bolstering our contention that the relationship between the trait factors and comfort ratings may have been mediated through self-esteem. On the other hand, Bem's theory, which predicts an interaction between PAQ category, sex of subject, and type of task, received no support.

The forced-choice preference ratings were more likely than the comfort ratings to compel the appearance of any association that exists between instrumental and expressive traits and preferences for the kinds of mundane, sex-related tasks we investigated. The results from females were essentially negative, the PAQ (and the AWS) variables being nonsignificantly related to preference for gender-congruent tasks. Positive results, however, were found in males, who as a group, it is interesting to note, showed strikingly higher preference ratings than women. Masculine men exhibited significantly higher preference ratings than did androgynous and undifferentiated men but, curiously enough, nonsignificantly higher ratings than did cross-typed feminine men. One other aspect of these findings from males should also be mentioned. As we predicted, our measure of general sex role attitudes (AWS) was more strongly related to the preference measure than were PAQ scores and, in analyses by PAQ category, accounted for almost twice as much of the variance as the latter. (Had a general measure of *personal* sex role choices been included, even stronger relationships with preference ratings might have been obtained, not only in males but in females.)

The preference data from the male subjects replicate the findings of Bem and Lenney that sex-typed subjects (in this instance, masculine men) show the greatest preference for gender-congruent tasks on a forced-choice measure. This finding supports Bem's theory, but the relative performances of the remaining groups did not provide confirmation of other predictions derived from her theory. Finally, the PAQ variable, even in males, contributed only weakly to the preference ratings. When these data are viewed in the context of all the available evidence, they appear to us to be more supportive of our contention that the trait dimensions measured by the scales are not only *conceptually* distinguishable from various classes of sex role behaviors and attitudes but are often only *minimally* associated with these other sex-related phenomena, if at all.

Finally, even the male preference data do not support the broad hypothesis that androgynous individuals, high in instrumental and expressive qualities, are more behaviorally flexible in their *sex role behaviors* than are those in other categorical groups. The evidence to date suggests that the androgynous are more flexible than others in the limited sense of manifesting a higher degree of both instrumentality and expressiveness in their behavior but that the extension of the

flexibility notion to include sex-related behavior in general is unwarranted.

A question that remains to be answered is why the present results were seemingly weaker than those reported by Bem and Lenney (1976). One explanation may be that the population effects were weak but a greater number of significant findings occurred in that study because subjects were selected from extremes of the distribution of combined M and F scores. The possibility cannot be discounted that procedural discrepancies may have brought about differences in the magnitude of effects in the two studies. However, it should be acknowledged that if method factors are responsible, the generalized *sex role* rigidity purportedly exhibited by individuals who are sex-typed in personality characteristics is an extremely fragile phenomenon rather than the robust one suggested by Bem's theory.

Finally, differences in the BSRI and the PAQ may be cited. The PAQ M and F scales, as we have demonstrated, form unidimensional clusters of socially desirable instrumental and expressive traits. Many of the BSRI M and F scale items can be similarly characterized. Others cannot, however, and the inventory is factorially more complex than the PAQ. These "extraneous" elements may bring about a stronger relationship with sex role behaviors. A specific hypothesis can be derived from a recent analysis of the psychometric properties of the BSRI by Pedhazur and Tetenbaum (1979). These investigators point out that the inclusion of the terms "masculine" and "feminine" on the M and F scales, respectively, greatly influences the discriminant ability of the instrument and also its factor structure. For example, in a discriminant function analysis between male and female respondents, 96.7% of the sample were correctly classified as to gender by the full function (all 40 items), but 93.5% were correctly classified by using the function based on the adjectives *masculine* and *feminine* alone. Factor analyses in each sex yielded four interpretable factors, "masculine" and "feminine" forming a separate, bipolar factor, with "feminine" having a positive loading and "masculine" a negative one. Given the demonstrated import of these two adjectives, it may be that their inclusion to some degree evokes schemata (Markus, 1977) of appropriate masculine and feminine roles and that these schemata may influence endorsement of the trait adjectives that follow their appearance. Admittedly, this is putting a heavy explanatory burden on two items. More generally, interpretation of data obtained with the BSRI will be facilitated if there is better understanding of the influence that the presence of the subset of items that do not tap socially desirable instrumental and expressive traits has, both on responses to the items that do describe such traits and on the relationships of each subset of items to other behavioral and self-report measures.[3]

[3] Bem (1979) has recently developed a Short BSRI dropping, notably, the items *masculine* and *feminine* as well as those with low social desirability. It seems reasonable to hypothesize that results from subjects classified by either the median split or difference method on the Short BSRI would be closer to those obtained in the present study.

Reference Note

1. Klein, H. M. *Psychological masculinity and femininity, self-consciousness, and typical and maximal dominance expression in women.* Unpublished doctoral dissertation, University of Texas at Austin, 1978.

References

Bem, S. L. The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology,* 1974, *42,* 155–162.

Bem, S. L. Sex role adaptability: One consequence of psychological androgyny. *Journal of Personality and Social Psychology,* 1975, *31,* 634–643.

Bem, S. L. On the utility of alternate procedures for assessing psychological androgyny. *Journal of Consulting and Clinical Psychology,* 1977, *45,* 196–205.

Bem, S. L. The theory and measurement of androgyny: A reply to the Pedhazur–Tetenbaum and Locksley–Colten critiques. *Journal of Personality and Social Psychology,* 1979, *37,* 1047–1054.

Bem, S. L., & Lenney, E. Sex-typing and the avoidance of cross-sex behavior. *Journal of Personality and Social Psychology,* 1976, *33,* 48–54.

Bem, S. L., Martyna, W., & Watson, C. Sex typing and androgyny: Further explorations of the expressive domain. *Journal of Personality and Social Psychology,* 1976, *34,* 1016–1023.

Cohen, J. Partialed products *are* interactions; par-

tialed powers *are* curve components. *Psychological Bulletin,* 1978, *85,* 858–866.

Helmreich, R. L., Spence, J. T., & Wilhelm, J. A. A psychometric analysis of the Personal Attributes Questionnaire. *Sex Roles,* in press.

Markus, H. Self-schemata and processing information about the self. *Journal of Personality and Social Psychology,* 1977, *35,* 63–78.

Orlofsky, J. S., Astin, A. L., & Ginsburg, S. O. Differential effectiveness of two classification procedures on the Bem Sex-Role Inventory. *Journal of Personality Assessment,* 1977, *41,* 414–416.

Parsons, T., & Bales, R. F. Family socialization and interaction process. Glencoe, Ill.: Free Press, 1955.

Pedhazur, E. J., & Tetenbaum, T. J. Bem Sex Role Inventory: A theoretical and methodological critique. *Journal of Personality and Social Psychology,* 1979, *37,* 996–1016.

Spence, J. T., & Helmreich, R. L. A short version of the Attitudes Toward Women Scale (AWS). *Bulletin of the Psychonomic Society,* 1973, *2,* 66.

Spence, J. T., & Helmreich, R. L. *Masculinity and femininity: Their psychological dimensions, correlates and antecedents.* Austin, Tex.: University of Texas Press, 1978.

Spence, J. T., & Helmreich, R. L. The many faces of androgyny: A reply to Locksley and Colten. *Journal of Personality and Social Psychology,* 1979, *37,* 1032–1046.

Spence, J. T., & Helmreich, R. L. On assessing "androgyny." *Sex Roles,* in press.

Spence, J. T., Helmreich, R. L., & Holahan, C. K. Negative and positive components of psychological masculinity and femininity and their relationships to neurotic and acting out behaviors. *Journal of Personality and Social Psychology,* 1979, *37,* 1673–1682.

Spence, J. T., Helmreich, R., & Stapp, J. The Personal Attributes Questionnaire: A measure of sex role stereotypes and masculinity–femininity. JSAS *Catalog of Selected Documents in Psychology,* 1974, *4,* 127. (Ms. No. 617)

Spence, J. T., Helmreich, R., & Stapp, J. Ratings of self and peers on sex role attributes and their relation to self-esteem and conceptions of masculinity and femininity. *Journal of Personality and Social Psychology,* 1975, *32,* 29–39.

Received July 1, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 10, 1645-1659

Nonverbal Trait Inference

Sampo V. Paunonen and Douglas N. Jackson
The University of Western Ontario, London, Canada

The purpose of this study was to evaluate empirically the view that personality trait inferences merely reflect consistency in linguistic associations among verbal trait descriptors and have little correspondence with the actual organization of behavior. Trait inferences were obtained in both verbal and nonverbal rating contexts to evaluate their relative reliability, accuracy, and structure. Eighty male and 80 female undergraduates were supplied with either a verbal or nonverbal (cartoonlike) description of one of two target persons and were required to estimate the probability that the target would endorse both verbal and nonverbal self-descriptive personality items. Results indicated very similar reliabilities, accuracies, and structures for the different conditions. The results are interpreted as not supporting the view that explicit semantic cues are necessary in accounting for observed consistencies in ratings of personality. Alternative hypotheses of desirability and the mediation of nonverbal inferences by some putative covert verbal process are evaluated and are found not to be supported by the data.

One of the dogmas in some current theories about personality and the perception of personality is that there are few consistencies, at least of the sort that can be attributed to broad behavioral dimensions or dispositions. Evidence from a variety of sources regarding consistencies in behavior has not been lacking, but its interpretation as being due to broad personality dispositions has been disputed. For example, ratings made about the personality of others in a variety of situations have consistently yielded a stable structure (Wiggins, 1973, pp. 328–349). But the fact that such studies have typically employed semantic trait names has raised doubts regarding the extent to which these consistencies are attributable to underlying behavioral consistencies in the person being rated. D'Andrade (1965) obtained a similar factor structure for 20 traits judged for "similarity of meaning" that had been identified previously by Norman (1963) from data based on ratings of peers. From these data D'Andrade (1965) attempted to draw a general conclusion: "Some of the classifications used by psychologists can be derived solely from similarities in the meanings of words without considering any sample of actual behavior" (p. 215). Similarly, Mulaik (1964) suggested that parallel findings derived from people ratings and word ratings "serve to focus attention upon the role of conventional linguistic usage in determining the correlations between trait rating scales" (p. 507).

The findings of D'Andrade and of Mulaik, advancing the hypothesis that observed consistencies in rating others might be due to linguistic conventions, have been highlighted by others as evidence against the existence of traits or other broadly based regularities in behavior. For example, Mischel (1968) indicates that "so-called traits at least in part exist as components of the verbal terms used to describe the external world; they do not

The authors especially thank Deirdre Flavin for drawing nonverbal stimulus materials. This study was supported by the Social Sciences and Humanities Research Council of Canada Research Grant 410-77-0320-R1, Douglas N. Jackson, Principal Investigator, and Fellowship 453-784712 to Sampo V. Paunonen.

Requests for reprints should be sent to Douglas N. Jackson, Department of Psychology, The University of Western Ontario, London, Ontario, Canada N6A 5C2.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3710-1645$00.75

necessarily mirror the external world itself (D'Andrade, 1965)" (p. 46). As time has passed, earlier tentative statements of this "semantic similarity" hypothesis have tended to be treated as established law. Schneider (1973), for example, states, "The claim is made and indeed rarely questioned that various kinds of linguistic structures produce the implicit personality theory" (p. 302). Bandura (1969), D'Andrade (1974), Mirels (1976), Mischel (1976), and Shweder (1975), among others, have argued forcefully for the primacy of "linguistic characteristics," "concept meanings," "preexisting conceptual schemes," or some similar term in influencing the perception of generalized response dispositions. But rarely are these terms clearly and explicitly defined.

The concept of "similarity of meaning," which is central to the issue, since judgments of the similarity of word meanings have been used to assess the "cognitive structure" of raters, is somewhat obscure in that it has been the subject of a variety of interpretations, not all of which have been made explicit by researchers and not all of which are mutually consistent (D'Andrade, 1974, pp. 177–180). D'Andrade (1974) recognizes this as a "critical problem" and states that, in his treatise, "measures of conceptual similarity have been phrased primarily in terms of semantic similarity" (p. 177). Moreover, he discusses several varieties of semantic similarity and concludes, "it is somewhat ironic that the problem of not knowing what is happening when people are used as measuring instruments, raised at the beginning of this chapter and given as the reason why certain procedures in the social sciences produce invalid results, returns to plague a later formulation about exactly how these invalid results come about" (p. 180). In view of the confusion surrounding the nature of the process purportedly invalidating ratings of personality, it is necessary to state clearly which version of the semantic similarity hypothesis is under investigation in the present study.

Most studies involving ratings of the personalities of others have used verbal trait labels as reference points for rating scales. It has been conjectured that the "similarity in meaning" of these verbal trait descriptors rather than "real world" similarity may account for the observed consistencies in these rating studies. To our knowledge, this hypothesis has not been sufficiently delineated to allow for predictions of potential effects due to varying degrees of verbal and nonverbal target information, target-relevant versus target-irrelevant information, type of rating scale used, or other parameters conceivably biasing trait inferences. D'Andrade (1965), however, underscores the fact that the classification of traits by rating scales has been *confounded* with the use of verbal trait descriptors. Consequently, he questions the significance of trait attribution research, since it is difficult to ascertain whether "studies that have reported in various forms correlations between two or more linguistic labels have been reporting about relations between real world events, or about how similar in meaning the labels are, or some complex interaction of meanings and external events" (p. 228). It is our intention in this study to unconfound the use of verbal materials in designating traits from the trait descriptions by employing nonverbal as well as verbal stimulus and response materials in a trait inference paradigm.

If the observed consistencies in the ratings of personality are simply a product of the words used by experimenters, it is possible that these ratings, as merely a reflection of the perceivers' "cognitive structures," will not correspond to the actual organization of behavior as measured more objectively. An alternative position is the *isomorphism* hypothesis, that relations between semantic cues are highly associated with the actual probable co-occurrence of the behaviors reflected by the verbal cues. Certain authors (D'Andrade, 1974; Shweder, 1975) entertain this possibility, but favor the interpretation that linguistic associations somehow determine the observed, invariant structures and that these structures have little correspondence with the actual organization of behavior. Of course, if the isomorphism hypothesis were confirmed, if semantic meaning were shown to mirror behavioral consistencies validly, such a finding would only serve to reinforce confidence in

the validity of observed behavioral consistencies or traits. Results apparently at variance with the isomorphism hypothesis (D'Andrade, 1974; Shweder, 1975) have been challenged by Block, Weiss, and Thorne (1979).

The issue is an important one theoretically and one about which there is conflicting evidence. Lay and Jackson (1969), for example, found that the structure observed for multidimensional scaling of trait names was similar to that obtained when actual Personality Research Form (PRF) (Jackson, 1974) items were used, which referred to more concrete behavior than that implied by abstract trait names. Furthermore, these structures were quite similar to the factor structure derived from factor analytic findings based on responses to the entire PRF by different subjects. Chan and Jackson (1979) have obtained similar results in the domain of psychopathology.

Recently, Mirels (1976) reported a lack of any statistically significant relationship between judgments of conditional probability of responses to pairs of the Lay and Jackson PRF items and observed conditional probabilities of respondents, claiming that such presumed relationships were "illusory." But Jackson, Chan, and Stricker (1979) reanalyzed Mirels' data and reported new data indicating that when the more appropriate product–moment correlation is used as an index of behavioral co-occurrence rather than conditional probability as used by Mirels, the association between judged and actual co-occurrence of behavior is highly statistically significant and substantial. The evidence seems to suggest that "subjects make judgments of item coendorsement on some basis other than conditional probability" (Jackson et al., 1979, p. 9). Reed and Jackson (Note 1) report a number of studies indicating that when target persons are judged on the basis of broad descriptive paragraphs, the behavior of actual persons like the targets can be predicted with a high degree of reliability and accuracy. When base rate and desirability were also taken into account, predictive validities reached as high as .95.

Further evidence of the relevance of trait ratings comes from a study by Norman and Goldberg (1966). These authors demonstrated greater interrater reliability and criterion validity for peer ratings made on close as opposed to distant acquaintances. Kusyszyn (1968) and Jackson, Neill, and Bevan (1973) also reported similar findings. This observation suggests that raters' perceptions are veridical when they have relevant information upon which to base their ratings. In addition, since Norman and Goldberg analyzed ratings made on the same variables used by D'Andrade (1965), their results lend credence to the notion of an isomorphism between meaning factors and person factors. From the present vantage point, however, all of these studies confounded the use of verbal rating scales and the question of veridicality.

The question arises: Is the frequently observed organization of behavior in ratings of others solely a function of explicit verbal informational cues and verbal rating scales, or might this same structure arise in the absence of such semantic cues? In this study we propose the use of a method not depending upon explicit linguistic elements in the forming or reporting of inferences about the behavior of others. Judges will be asked to make inferences about the behavior of a target person who will be characterized with pictorial representations of some concrete behaviors in which linguistic elements are lacking. Similar pictorial information, again based on the figural representaton of concrete behaviors in a diversity of situations, will be used for obtaining ratings of the probability that a particular target individual will engage in particular behaviors. Also, we propose to include more traditional verbal personality statements or items for comparison purposes, both for providing information about a target person and for obtaining judgments regarding the behavior of particular target individuals. The proposed design separates verbal and nonverbal cues, both in terms of the information used to form an impression of the individual and in terms of the vehicle used in communicating that impression. The study is thus designed to investigate (a) the degree to which consistencies in rated behavior depend upon explicit linguistic or semantic characteristics,

(b) the degree to which verbal and nonverbal information yields inferences about behavior that conform to personality types observed empirically, and (c) the structure of perceptions of personality elicited by verbal and nonverbal materials. We hypothesize that there will be a high degree of consensus among judges regardless of whether verbal or figural cue information or rating scales are used and that, on the whole, judges' characterizations of the behavior of target persons will conform closely to the structuring of the behavior as observed empirically. We are not hypothesizing that individuals necessarily avoid all use of implicit semantic reasoning processes in drawing inferences about personality, but rather that explicit verbal cues, that is, those supplied by the experimenter in transmitting and receiving information, are not a necessary prerequisite for the identification of stable predictions regarding the behavior of others.

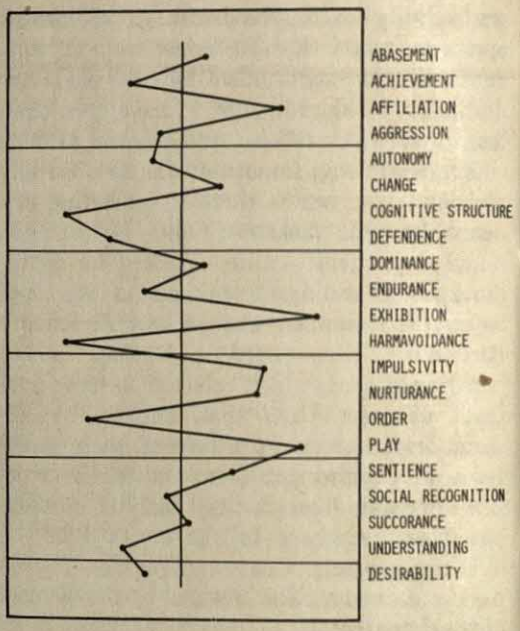


Figure 1. Criterion Profile 1, based on 796 Personality Research Form E male respondents (basis for Ed Nolan target).

Method

Experimental Design

Psychology undergraduates were used in this study as raters who were instructed to make estimates concerning the probability that a target would endorse certain self-descriptive personality items. Each rater was presented with a limited amount of information describing one of two possible targets depicted as having distinct behavioral characteristics. Each of the target characterizations consisted either of a verbal personality sketch or of a set of nonverbal cartoonlike drawings showing a target individual engaging in certain behaviors. Predictions of the likelihood of behavior in particular situations were of one of two types: (a) ratings of the probability that the target individual would endorse each item of a particular set of verbal personality items and (b) a prediction that the target person would engage in the particular behavior depicted by a central figure in each of a number of nonverbal drawings. Predictions of the likelihood of behavior were completed in one of two orders: (a) predictions based on nonverbal drawings first, followed by predictions of responses to verbal personality items, designated Order 1, and (b) predictions based on verbal personality items first, followed by predictions of the behavior depicted by the nonverbal drawings, designated Order 2.

Construction of the Target Descriptions

Subjects were instructed to make behavioral predictions for one of two target characters, who were described either verbally or nonverbally. The attributes of the targets were based on distinct personality types derived empirically from a multivariate clustering procedure (Jackson & Williams, 1975; Skinner, 1977; Skinner, Jackson, & Hoffmann, 1974). Given a matrix of profiles representing the scores of a number of persons on a number of personality scales, this procedure identifies dimensions of persons based on profile shape using an Eckart–Young (1936) decomposition of the data matrix. Persons having similar profiles occupy similar positions in the orthogonal n-dimensional space. Each cluster of individuals is represented by a set of standardized component scores typifying the profile of persons receiving high scores on the dimension.

Figures 1 and 2 show two of five bipolar "modal" profiles extracted from a sample of 796 PRF respondents. These two orthogonal configurations were used as the basis for forming the target descriptions for the present study and were used as the criterion profiles in the assessment of judgmental accuracy. Because the scale means are removed as a first step in the derivation of the modal profiles from the respondent by scale matrix, predictions of these standardized scores cannot be based on stereotype accuracy, as defined by Cronbach (1955; Wiggins, 1973). Judgments seeking to predict these profiles of scores cannot be accurate as a function of the raters' "knowledge of the relative frequency or popularity of the possible responses" (Cronbach, 1955, p. 179) since the "average item self-description summed over all others" (Wiggins, 1973, p.

127) represents a null vector and as such contributes no variance to the individual target profiles. Hence, judgments based on "accuracy in predicting the generalized other" (Cronbach, 1955, p. 179) would, in general, not contribute valid variance to the predictions of these criterion profiles.

An alternative procedure for identifying criterion personality trait profiles associated with a certain type would have been simply to obtain the personality scores of one or more persons highly associated with one of the typal dimensions (Reed & Jackson, Note 1). Although such a procedure has the advantage of simplicity and concreteness, and indeed would yield very similar results, scores so obtained would not be completely free of the stereotype component. But in either case we are dealing with a set of scores derived from the responses of real people, scores that have been shown to possess a degree of convergent and discriminant validity (Jackson, 1974).

Four of the most characteristic traits of Criterion Profile 1 (Figure 1) were used to describe the first target (Ed Nolan). This target was depicted as high on PRF scales for affiliation, exhibition, nurturance, and play. The second target (John Newman) was based on Criterion Profile 2 (Figure 2), denoted by the attributes of aggression, autonomy, dominance, and thrill-seeking (low harmavoidance).

Verbal target descriptions consisted of a name and five sentences written in third-person narrative form and presented as a personality sketch. For example, Ed Nolan was described to raters as follows:

> Ed Nolan is a sociable person who enjoys being with friends and people in general; in fact, he makes an effort to win friendships and maintain close acquaintances. Ed spends a lot of time participating in games, sports, and various social activities and often likes to play practical jokes on his fellow workers. His friends describe him as being very entertaining and delighting in behaviors that attract the attention of an audience. Ed also readily performs favors for others and gives sympathy and comfort to those in need. When people have a problem they often go to Ed for encouragement and consolation.

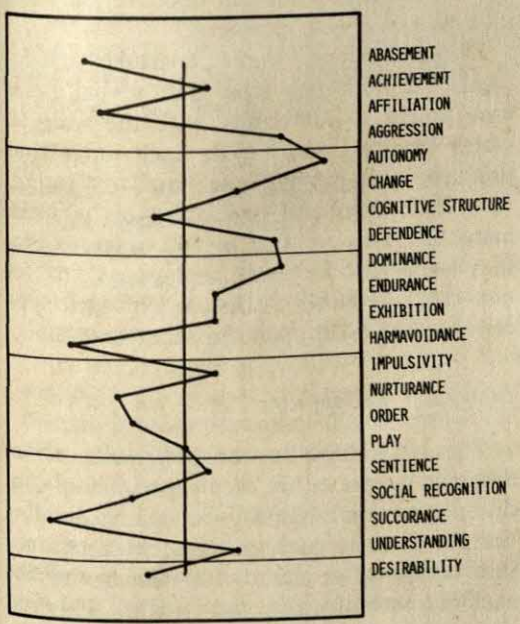


Figure 2. Criterion Profile 2, based on 796 Personality Research Form E male respondents (basis for John Newman target).

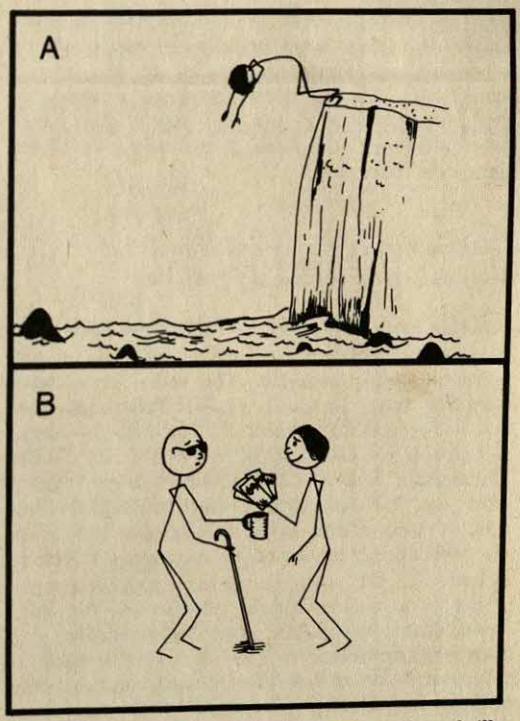


Figure 3. Sample nonverbal items depicting thrill-seeking behavior (A) and nurturant behavior (B).

The items for the nonverbal target descriptions were prepared by commissioning an artist to illustrate a series of scenarios supplied by the authors. Ideas for these scenarios originated from a careful perusal of PRF scale definitions and items. Each illustration consisted of a stick drawing of a central figure performing a behavior in a specific context. An attempt was made to convey pictorially a behavioral manifestation that would be characteristic of a person defining an extreme position on the substantive dimension of interest. For example, the upper panel of Figure 3 shows a nonverbal stimulus designed to reflect high thrill-seeking (low harmavoidance) in the central character, while the lower panel reflects high nurturance. Four items were conceptualized and illustrated for each of the four behavioral dimensions typifying each target, with an emphasis on nonredundancy of item content. In

composing the nonverbal trait descriptions, an attempt was made to imply behavioral consistency and to convey an impression parallel to the verbal counterparts. The 16 items describing a character were randomly ordered in a four-page booklet bearing a picture and the name of the central or target figure on a cover page.

Construction of Verbal and Nonverbal Materials for Behavioral Prediction

Judges were instructed to make behavioral predictions separately on the basis of both verbal and nonverbal rating materials. The verbal items were extracted from Jackson's (1974) Personality Research Form (PRF), Form E. The first five items for each of 17 content scales and two validity scales (infrequency and social desirability) were selected from the PRF test booklet. Each scale had either two or three items keyed in one direction, with the remainder being keyed in the reverse direction.

Items for the nonverbal rating scales were composed in a manner similar to that of the nonverbal target description scenarios. A total of five rationally constructed nonverbal items for each of 17 content scales was created with unidirectional keying. In addition, a five-item validity scale was compiled, reflecting undesirable behaviors but heterogeneous with regard to content or the possible personality scale on which they might be keyed. Three of the content scales of the PRF were not depicted because they represent particularly difficult constructs for nonverbal definition—cognitive structure, change, and defendence. These constructs were also absent in the verbal rating scales.

Each of the 90 nonverbal items was rated for desirability by 12 psychology graduate students using a 9-point rating scale, ranging from *extremely undesirable* to *extremely desirable*.

Subjects and Procedure

Subjects consisted of 80 male and 80 female introductory psychology students completing a course requirement. Run in groups of about 20, each subject was assigned to one of the cells in the design by randomly allocating different booklets to subjects as they appeared. This resulted in 10 males and 10 females in each rating situation.

Each participant was given a folder containing (a) some preliminary information common to all; (b) instructions specific to the rating situation; (c) a target description corresponding to one cell of the design; and (d) the verbal and nonverbal rating scales (order predetermined). The common preliminary information was read aloud by the experimenter while the subjects simultaneously read their text. The specific instructions were read silently by each participant; these instructions were equivalent across raters in form and wording except for necessary modifications such as interchanging the words *verbal description* and *nonverbal description*.

The directions for completing the verbal rating scales instructed the subjects to make a judgment, using a 9-point scale, on the probability that the described target would endorse each of the verbal statements. Instructions for the nonverbal rating scales stated that judges were to estimate the likelihood that the described target would engage in the behaviors depicted by the central figure in each item. Judgments were again recorded on 9-point scales.

Results

Overview of Analyses

The analyses may be divided into three stages corresponding to the three major questions addressed. First, interjudge reliabilities were computed to evaluate the degree to which raters share a consensus regarding their interpretation of nonverbal and verbal items or behavioral exemplars. Second, an assessment was undertaken to appraise the degree to which judgments based upon behavioral exemplars are valid in the sense that they show a pattern of inferences consistent with the structure of behaviors observed empirically. Third, the structures of inferences based upon nonverbal and upon verbal stimulus materials were compared.

Before analyses were undertaken, judgments for both nonverbal and verbal items were scored in conformity with the scale on which the item was keyed. Each judge thus produced two profiles, one based on ratings of verbal items and one based on pictorial materials. This resulted in 160 separate profiles for verbal judgment scales and 160 for nonverbal judgment scales, one of each type for each of the 20 judges in each condition.

Interjudge Reliabilities

The generalizability of judgments across different judges within each experimental condition was appraised by randomly dividing the 20 raters in each of the eight conditions into two equal groups of 10, each having the same number of males and females. This was done separately for nonverbally and verbally based profiles. The 10 profiles were then averaged separately for each of the two randomly

divided groups. The two mean profiles, consisting of 18 scales each (17 content scales and social desirability), were correlated. Table 1 reports interjudge reliabilities corrected by the Spearman–Brown formula to yield an estimate of within-cell group reliability (Gulliksen, 1950, p. 66). These values range from .93 to .99 and reveal a consistently high degree of consensus across treatment groups, targets, verbal and nonverbal descriptions, and verbal and nonverbal judgmental items. Of particular interest is the absence of a significant difference in group reliabilities between verbal and nonverbal rating scales. Average values of .96 for nonverbal and .97 for verbal conditions indicate that high levels of judgmental consistencies are not predicated solely on the presence of an explicit semantic structure in rating scales.

The Spearman–Brown formula was also used to derive estimates of single-subject reliabilities for each of the conditions. These range from .40 to .83, with a mean of .61 over all 16 conditions. More properly, these coefficients reflect the generalizabilities of judgmental sensitivities over judges expressed in terms of single judges. Were we seeking to develop a reliable test of judgmental sensitivity for single judges over targets, many more targets than one per judge, used here, would be appropriate, because targets are analogous to items on a test.

Table 1
Interjudge Reliabilities for Eight Experimental Conditions for Nonverbal and Verbal Ratings

Target description	Rating scale: Nonverbal	Rating scale: Verbal
Ed Nolan		
Order 1		
Nonverbal	.96	.98
Verbal	.93	.94
Order 2		
Nonverbal	.98	.96
Verbal	.98	.97
John Newman		
Order 1		
Nonverbal	.94	.97
Verbal	.96	.97
Order 2		
Nonverbal	.97	.99
Verbal	.98	.98

Note. Order 1 = predictions based on nonverbal drawings first, followed by predictions of responses to verbal personality items. Order 2 = predictions based on verbal personality items first, followed by predictions of the behavior depicted by the nonverbal drawings.

Evaluation of Judgments in Terms of Modal Types

In the present study, experimentally derived profiles based on the raters' judgments were expected to emulate empirically obtained criterion personality profiles (Reed & Jackson, Note 1). A close correspondence between the two is an indication of the accuracy of the inferential judgments observed in the present study. Such evidence would support the interpretation that the inferred relationships between personality characteristics accurately reflect these relationships as revealed in the covariation of responses to structured personality items.

Individual accuracies or *sensitivities* (Jackson, 1972; Reed & Jackson, 1975), reflecting the degree to which a judge manifested judgments congruent with observed behavioral consistencies, were computed by simply correlating both of his/her judgmental profiles of 17 content scales with the relevant criterion profile. Group sensitivities were obtained by first averaging the 20 individual profiles in a treatment condition and then correlating this group mean profile with the relevant criterion profile. These sensitivity indices were computed separately for both verbal and nonverbal rating scales. This computational technique (i.e., the product–moment correlation between two profiles) measures the degree of similarity between each judge's rating of the target profile shape and the profile shape as identified empirically in groups of respondents.

Individual sensitivities ranged from −.40 to .93. Sensitivities falling below the .30 level (of which there were only about 13%), like high PRF infrequency scores, can be attributed not only to a subject's lack of awareness of trait relationships, but possibly to some

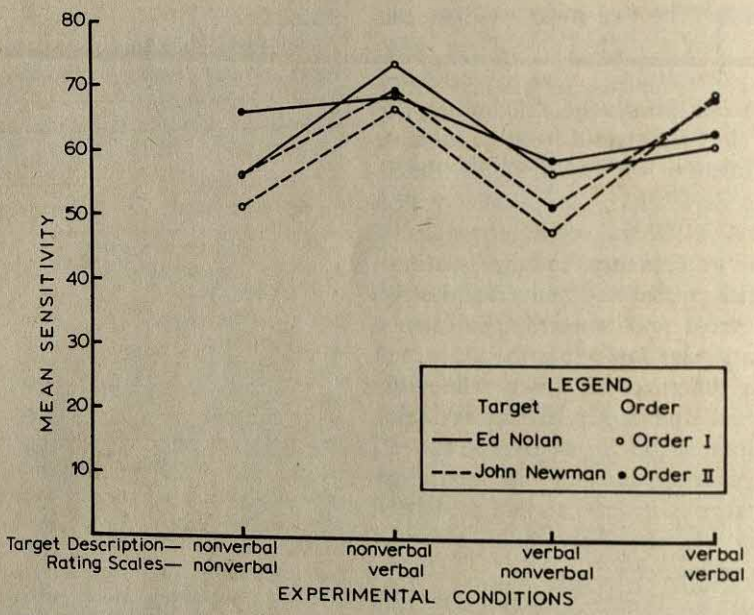


Figure 4. Mean sensitivities for nonverbally and verbally based judgments of behavior, by condition.

combination of lack of motivation, misconstrued instructions, random responding, or deliberate noncompliance. Figure 4 shows the means for the individual sensitivities for each experimental group. The most salient feature of Figure 4 is the substantial level of average sensitivity to trait covariation across targets and experimental verbal and nonverbal conditions. One notes also in Figure 4 a modest decrease in mean sensitivities for nonverbal ratings as opposed to verbal ratings, in each condition. However, verbal and nonverbal informational cues appear to result in similar levels of group sensitivity. To test these effects for significance and to examine any other effects on sensitivity due to the factorial manipulations, a four-way fixed-effects analysis of variance was run using the nonverbal and verbal rating scales as two levels of the fourth factor. Subjects were nested in the other three factors (description type, target person, order of dependent variables), and the data used as the dependent variable was the individual sensitivity index (this is referred to by Kirk, 1968, Section 8.13, as a split-plot factorial 222.2 design). This analysis revealed a significant main effect of rating scale type; sensitivities were attenuated when judgments were made on the nonverbal scales as opposed to the verbal items, $F(1, 152) = 51.58$, $p < .001$. Only one other effect was observed, and that was a Rating Scale × Target interaction, $F(1, 152) = 7.47$, $p < .01$; the drop in sensitivities as one examined the nonverbal judgmental profiles compared to the verbal profiles was more noticeable with the target John Newman (based on Criterion Profile 2). It should be emphasized that although they were significant, these effects account for a relatively small proportion of the variance. It should also be emphasized that the type of information used to form an impression of the person (verbal or nonverbal) and the order factor had no effects on rater accuracy. A separate analysis investigating whether or not gender of rater affected ratings also revealed no significant differences.

Table 2 shows the sensitivity of the group as a whole for both the verbal and nonverval personality rating scales for each of eight conditions. The accuracy of the group consensus in estimating the personality characteristics of the targets ranges from values of .58 to .95. This accuracy seems to occur regardless of the target being rated, the na-

ture of the target description, the order of dependent measures, or the nature of these rating scales.

The mean correlation of the group mean profiles based on the Ed Nolan target with the John Newman criterion profile is −.22. When the group mean profiles based on John Newman are correlated with the Ed Nolan criterion profile, the average correlation is .18.

Evaluating the Role of Desirability

The argument that the group sensitivities were simply based on the connotative social desirability components common to target descriptions and rating scale items was put to test by partialing this portion of the variance out of the ratings and recomputing the group sensitivities using standard partial regression techniques to control statistically the effects of desirability in both the predictor and the criterion.

Because the predictor profiles were based on either verbal or nonverbal items, two different desirability vectors were used for this analysis; the social desirability values of the scales as given in the PRF manual (Jackson, 1974, p. 12) were partialed from the verbally based profiles and the criterion profiles. The corresponding values for the nonverbal scales were obtained from the social desirability ratings of the nonverbal items. This vector was partialed from the judgmental profiles based on those items and the criterion profiles. The correlation between the two social desirability vectors is .82. Group sensitivities corrected for desirability are presented as parenthesized values in Table 2. A comparison with the uncorrected values reveals that there is little if any evidence that any of these values decline as a function of eliminating desirability. In fact, there is some evidence that desirability is acting as a suppressor variable (Conger & Jackson, 1972); when its effects are removed, accuracy in predicting the criterion profile is slightly increased.

After the effects of desirability had been removed from the correlations between the predictor and criterion profiles, estimated population base rates were partialed from these correlations. The vector of PRF Form E mean scale scores for males (Jackson, 1974, p. 42) was partialed from the correlations with the verbal rating scales. The mean change in group sensitivity is .01. A vector of mean scale scores derived from the self-reports of 128 male respondents to a set of 202 nonverbal items, of which the nonverbal rating scale items used in this study formed a subset, was partialed from the correlations with the nonverbal rating scales. A mean change in group sensitivity of .01 is also observed in this analysis, indicating that the prediction of population endorsement frequencies was not a large component of the subjects' judgments of either target.

Table 2
Correlations Between Group Mean Profiles and Criterion Profiles, by Condition

Target description	Rating scale: Nonverbal	Rating scale: Verbal
Ed Nolan		
Order 1		
Nonverbal	.73 (.80)	.93 (.94)
Verbal	.73 (.79)	.84 (.85)
Order 2		
Nonverbal	.85 (.91)	.95 (.95)
Verbal	.76 (.77)	.83 (.83)
John Newman		
Order 1		
Nonverbal	.77 (.78)	.84 (.84)
Verbal	.58 (.61)	.81 (.81)
Order 2		
Nonverbal	.75 (.77)	.84 (.84)
Verbal	.63 (.67)	.83 (.82)

Note. Correlations with social desirability partialed out are in parenthesis.

Evaluating the Structure of Verbal and Nonverbal Cues

A multivariate analysis of variance (MANOVA) was used to evaluate the effect of the factorial manipulations on the 18 component profiles of trait ratings. The analysis (a multivariate split-plot factorial with rating scale type as the within-block treatment and description type and order of ratings as the between-block treatments) was run separately for the ratings of the target John Newman and the target Ed Nolan. The mean profiles

Table 3
Group Mean Profiles, by Condition and Rating Scale Type, for Two Targets

Target description	Order	Rating scale	Abasement	Achievement	Affiliation	Aggression	Autonomy	Dominance	Endurance	Exhibition	Harmavoidance	Impulsivity	Nurturance	Order	Play	Sentience	Social recognition	Succorance	Understanding	Social desirability
									Target: Ed Nolan											
Nonverbal	1	nonverbal	4.37	4.60	7.18	2.83	5.07	3.78	4.65	6.85	4.63	5.39	7.09	3.88	7.41	5.50	5.30	4.62	4.64	7.60
Verbal	1	nonverbal	4.13	4.15	7.87	2.46	4.52	3.67	4.40	7.50	4.84	5.29	7.55	3.62	7.59	5.23	4.49	4.72	4.53	7.61
Nonverbal	2	nonverbal	4.88	3.90	7.42	2.82	5.24	4.25	4.59	7.09	3.50	5.65	6.88	3.88	7.23	5.72	5.05	4.94	4.63	7.35
Verbal	2	nonverbal	4.64	4.14	7.59	3.05	4.70	4.30	4.46	7.27	5.38	5.84	6.44	3.96	7.56	5.25	5.02	5.30	4.30	7.41
Nonverbal	1	verbal	5.56	4.81	7.10	4.10	4.54	5.37	4.81	7.97	2.99	6.54	6.88	4.24	7.47	4.88	5.04	4.58	3.95	6.47
Verbal	1	verbal	6.27	4.39	7.87	3.48	4.26	4.73	5.07	7.92	4.13	5.74	7.35	4.22	7.13	4.78	5.40	4.74	3.26	6.48
Nonverbal	2	verbal	5.85	5.13	7.02	4.56	4.65	5.52	4.87	8.05	3.48	6.79	6.77	4.24	7.07	5.37	4.48	4.95	3.87	6.22
Verbal	2	verbal	6.28	4.91	7.59	4.22	4.05	5.54	5.03	7.41	4.70	5.78	6.40	3.79	7.13	4.83	5.99	5.05	3.34	6.42
									Target: John Newman											
Nonverbal	1	nonverbal	3.20	4.80	5.14	6.93	6.70	6.54	4.86	5.16	2.89	5.83	4.02	5.93	5.99	5.46	6.05	3.85	5.26	5.31
Verbal	1	nonverbal	2.54	4.10	5.91	6.88	5.88	7.10	4.67	5.74	2.29	6.49	3.59	5.11	6.44	4.13	6.97	3.52	4.27	4.41
Nonverbal	2	nonverbal	3.19	5.07	5.03	6.86	6.71	7.12	5.34	5.42	2.45	5.80	3.22	5.80	6.19	4.99	6.67	3.70	5.10	4.75
Verbal	2	nonverbal	2.96	4.47	5.73	7.08	6.17	7.13	4.48	5.58	2.36	6.15	3.09	5.13	6.40	3.68	6.73	3.04	4.27	3.94
Nonverbal	1	verbal	2.56	5.80	3.42	7.84	6.53	7.82	5.08	6.51	2.12	6.28	3.32	5.41	5.95	5.10	4.34	3.74	4.30	5.37
Verbal	1	verbal	2.06	5.31	4.15	7.47	7.45	7.77	4.88	7.35	2.38	6.17	3.46	4.86	6.82	4.03	4.14	3.01	3.67	5.82
Nonverbal	2	verbal	2.34	6.09	3.52	7.55	6.89	7.84	5.32	6.82	2.41	5.79	3.21	5.51	5.75	4.96	3.89	3.74	3.95	5.57
Verbal	2	verbal	2.77	5.69	3.89	7.19	6.89	7.48	5.33	7.12	2.65	6.10	3.05	5.00	6.06	3.94	3.86	3.06	3.72	6.09

Table 4
Unrotated Principal Components Factor Analysis Solutions of Matrices of Group Mean Profile Intercorrelations

	Target description	Order	Rating scale	Target: Ed Nolan Factor 1	Factor 2	Factor 3	Target: John Newman Factor 1	Factor 2	Factor 3
1	nonverbal	1	nonverbal	.94	−.33	.03	.94	−.23	.24
2	verbal	1	nonverbal	.95	−.30	−.01	.91	−.37	−.15
3	nonverbal	2	nonverbal	.96	−.17	.16	.95	−.24	.15
4	verbal	2	nonverbal	.95	−.24	−.12	.93	−.32	−.15
5	nonverbal	1	verbal	.94	.28	.17	.95	.26	.10
6	verbal	1	verbal	.96	.16	−.14	.96	.23	−.14
7	nonverbal	2	verbal	.91	.34	.19	.93	.35	.08
8	verbal	2	verbal	.91	.28	−.30	.94	.31	−.12
Total variance (%)				88.1	7.3	2.7	88.1	8.6	2.2

for the various conditions are presented in Table 3.

The MANOVA revealed the following effects for the ratings of John Newman (using Wilks' lambda criterion). Description type and rating scale type were both significant beyond the .001 level, approximate $F(18, 59)$ was 3.25 and 29.9, respectively, and there was a significant Description Type × Rating Scale interaction, $F(18, 59) = 2.7$, $p < .003$. The same effects were replicated with the Ed Nolan target, $F(18, 59) = 2.7$, $p < .003$; $F(18, 59) = 19.0$, $p < .001$; $F(18, 59) = 4.8$, $p < .001$, respectively, in addition to a significant order effect, $F(18, 59) = 1.9$, $p < .04$.

Despite the above differences among the mean vectors, an examination of the intercorrelations of these mean profiles of Table 3 shows a substantial degree of consensus for all experimental groups judging the same target. The within-target values all exceed .70, with the mean for the target John Newman being .87 and that for the Ed Nolan target being .86. The mean heterotarget correlation, however, was only −.02.

A principal components factor analysis was undertaken to seek to identify the relative contributions of the descriptive attributes of the targets and the experimental manipulations on the judgments. Components were extracted from the intercorrelation matrix of group mean profiles for each target. In this analysis, the experimental conditions were treated as variables and the trait dimensions as replications. For both targets, the largest components in Table 4 show a pattern of high positive loadings, accounting for more than 88% of the total variance. Since the primary common element running through all eight sets of ratings was the common target, this factor may be interpreted as reflecting a consistent conception of the attributes of the target person, cutting across the different combinations of verbal and nonverbal sets of cues and response formats. The second factor, accounting for somewhat less than 9% of the variance for each target, is a bipolar dimension, clearly separating ratings made on the nonverbal personality rating scales from ratings made on the parallel verbal form. The third factor, although very small, consistently separates ratings based on verbal and nonverbal target descriptions.

Discussion

The authors recognize certain limitations of the present study. Only one type each of verbal and nonverbal material was used, making it difficult to substantiate claims of generalizability to other rating situations. Any differences in the judgments of verbal and nonverbal rating scale items may have been the consequence of slightly different instructional sets for completing these ratings. Also, rating validities might have been different had a more objective criterion measure of

actual target behavior been available rather than that based on self-report data. In spite of these limitations, the results clearly demonstrated that (a) judges showed extremely high degrees of agreement when inferring the behavior of a target from limited information; (b) although profile differences were found across verbal and nonverbal conditions, the profile similarities were more interesting; (c) highly similar factor structures were attained regarding the perceived patterning of target traits across verbal and nonverbal informational and rating conditions; (d) the ratings of two targets were approximately orthogonal, corresponding to the orthogonal patterning of target characteristics as measured empirically; (e) although there was somewhat less accuracy regarding the prediction of behavior using pictorial items as compared with verbal items, both sets of rating formats yielded substantial levels of accuracy. Although it was not assessed in this investigation, the conclusion follows from Norman and Goldberg's (1966) study that even greater levels of interrater reliability and criterion validity would have been expected had more target-relevant information been available to the raters.

What implications do these results have for the "semantic overlap" hypothesis? The answer to this question is unfortunately complicated by the rather vague manner in which this hypothesis has been stated. Perhaps the clearest version of this hypothesis is its extreme form, namely, that the perceived structuring of personality traits is due entirely to some form of linguistic similarity in word meanings employed in trait rating scales and that such relationships would be different or even nonexistent if these words were not used in the rating scales. Clearly, the present results tend to disconfirm this interpretation of the semantic overlap hypothesis. Verbal descriptors are by no means necessary in eliciting consistencies in behavioral predictions. Indeed, this hypothesis provides no basis for interpreting the almost identical patterning of behavioral predictions made with both verbal and nonverbal target information or rating formats. Clearly, predictions similar to those obtained in the verbal rating conditions are made in the total absence of explicit verbal information.

There are alternative weaker forms of the hypothesis that merit some consideration. These may be summarized as follows: (a) linguistic factors do not account for all of the expected covariation of behavior, but contribute substantially to such findings; (b) connotative attributes of trait descriptions, like desirability, account for a major portion of the variance in observations of personality; or (c) even nonverbal cues elicit common linguistic associates in judges that are semantically related in some form of verbal network. Ratings of personality made in a context devoid of any explicit verbal materials may be mediated by the implicit use of cognitively generated verbal labels.

With regard to the first alternative form of the semantic overlap hypothesis, that semantic processes account for substantial independent variance incrementing that obtainable from nonverbal processes, the data are generally nonsupportive. There were no significant differences in mean levels of correlation with the criterion between verbal and nonverbal target descriptions. The differences in accuracy between verbal and nonverbal rating scales, though statistically significant, were not substantial, a finding particularly noteworthy in view of the fact that the comparison was between unselected nonverbal items and highly selected verbal items. It is reasonable to assume that the psychometric characteristics of the PRF items are superior to those of the nonverbal illustrations. These scenarios were amassed without the consideration of important item properties such as content saturation, frequency of endorsement, and components of desirability, whereas the verbal items used have undergone stringent item selection procedures. In addition, one might expect language to provide a richer and more precise description of the behavior to be rated. This latter point is important in appraising the earlier criticisms of trait-oriented rating scales. Whereas earlier rating scales usually employed simple trait names or general descriptive phrases, the PRF items judged in the verbal rating conditions were largely descriptions of concrete behaviors. No one has explicated precisely the manner in which semantic or linguistic associative rules might conceivably apply to ratings of this type.

The second weaker semantic overlap hypothesis is that personality ratings are due to common connotative meanings such as desirability (evaluation), activity, and potency (Mischel, 1968). Two points might be raised regarding this interpretation. First, although these connotative components of meaning have often been studied in the context of verbal meaning, the Osgoodian formulation (Osgood, 1952) is general and does not depend on verbal meaning per se. In our study, it would be a simple matter for judges implicity to rate nonverbal stimuli for desirability. Whether verbal or nonverbal, no formulation is available for linking these connotative item attributes with the type of profiles comprising the empirical criterion for accuracy in this study. Second, the connotative attribute most prominently linked to ratings of personality, that is, desirability, when held constant statistically in our data, demonstrated no incremental effect on accuracy. Judgmental accuracy could not be attributed to a desirability stereotype, nor could it be accounted for as the prediction of normative item endorsement frequencies, since (a) the removal of base rates of endorsement did not alter the degree of association between the predictions and their respective criteria, and (b) the mean ratings for the two targets were approximately orthogonal.

To what extent are the results derived from nonverbal cues and nonverbal rating scales due to verbal mediation? This is a difficult question to answer, simply because such an hypothesis refers to covert, unobserved processes in accounting for empirical results. There are three relevant points here. First, this study addressed itself initially to semantic overlap between *explicit* verbal cues in ratings of personality. To our knowledge, no one has attempted to explicate the manner in which such supposed covert complex processes might operate. Second, if one might assume that these verbally mediated processes did occur, for example, between the processing of nonverbal personality descriptions and nonverbal ratings, one would expect some loss in fidelity in comparison with the situation in which verbal descriptions of personality would be rated nonverbally. (Verbal mediation would assume that the nonverbal cues would have to be coded verbally and entered into a semantic space, with a verbal personality inference made and this information transformed into a nonverbal rating. From this perspective, going from verbal cues to nonverbal ratings would be less subject to error, since access to the semantic space would be more direct.) Contrary to expectations from a verbal mediation hypothesis, accuracy is not attenuated going from nonverbal cues to nonverbal ratings. Similarly, going from nonverbal cues to verbal ratings is not less accurate than going from verbal cues to verbal ratings, contrary to expectations based on a verbal mediation hypothesis.

The third point to consider with regard to this question of verbal mediation concerns the lack of an effect for the order manipulation. It might be argued that if verbal mediation were to occur, the intervention of verbal material subsequent to the study of the nonverbal target description but prior to making the nonverbal judgments would expedite this cognitive process, thereby intensifying accuracy and consistency in the predictions of behavior. There are two reasons to expect greater fidelity on the nonverbal rating scales after, rather than before, the verbal rating scales have been filled out: (a) the verbal items would elicit covert verbal labels not evoked by the preceding nonverbal target description nor evident in the subsequent nonverbal illustrations; and (b) the process above would happen to a certain degree in all the subjects receiving common verbal material prior to their nonverbal judgments on the nonverbal target description (i.e., subjects in the second order conditions). Since the explicit labels and terms contained in the verbal rating scales would be constant for all subjects (whereas cognitively generated labels might not be), this would mean that the intervention of verbal material would serve to unify, across raters, the covert labels upon which the trait attributions may be predicated. In short, one would expect the introduction of explicit verbal material in a nonverbal rating context to have an incremental effect on inferential sensitivity under the verbal mediation hypothesis. However, there was no effect of order of ratings on rating accuracy or consistency.

Although it is clear that a verbal mediation hypothesis cannot readily account for our

major findings, we do not wish to imply that verbal processes are nonexistent in judgments of personality. Most natural languages, including English, are rich with trait descriptive terms. It would be silly to suppose that people did not sometimes use them. Clearly, research is needed on the conditions under which verbal processes contribute to judgments of others. Although such research may well demonstrate their presence, the present findings support the view that verbal information is neither necessary nor sufficient to account for major findings regarding the structure of perceived personality relationships.

In reconciling the data from this study with previous findings of the structure of personality ratings, one might extend the notion of a parallelism between verbal and nonverbal judgmental processes, which was originally developed to account for findings in the area of cognitive psychology (Paivio, 1975), to the area of person perception. Such a notion, which obviates the concept of verbal mediation, requires the acceptance of two networks of associatively linked entities, one being a network of verbal descriptors, and the other, a network of imaginal entities. Presumably, personality inferences would be based on either or both cognitive systems depending on the input modality, contextual cues, and requirements of the task. This dual coding approach "distinguishes between imaginal and verbal cognitive processes . . . [which are] independent but partly interconnected systems for encoding, storage, organization, transformation (i.e., reorganization), and retrieval of stimulus information" (Paivio, 1975, p. 148). Empirical studies have demonstrated that nonverbal, or imaginal, thinking "can be characterized by remarkable speed, accuracy, and flexibility of information processing" (p. 161).

Conclusions

From an experimental comparison of the use of nonverbal and verbal cues in a personality judgment task, and a reinterpretation of extant data, we may conclude that

1. Substantially reliable and accurate behavioral inferences can be obtained in a judgmental context devoid of explicit verbal materials.

2. The structure of judgments based on nonverbal cues corresponds closely to the structure of ratings obtained in a situation where verbal cues preponderate.

3. The data were not consistent with the interpretation that a process of verbal mediation accounted for the similarities between nonverbal and verbal personality inference.

4. Neither could the major findings of this study be attributed to the use of a common desirability stereotype or "generalized other" in the use of nonverbal and verbal cues. Statistical control of desirability and response base rates showed that they had no incremental effect on accuracy.

5. The data support the view that linguistic informational cues, relationships, and rating scales are neither necessary nor sufficient to account for observed consistencies in personality.

Reference Note

1. Reed, P. L., & Jackson, D. N. *Personality measurement and inferential accuracy* (Research Bulletin No. 419). Unpublished manuscript, The University of Western Ontario, London, Canada, 1977.

References

Bandura, A. *Principles of behavior modification.* New York: Holt, Rinehart & Winston, 1969.

Block, J., Weiss, D. S., & Thorne, A. How relevant is a semantic similarity interpretation of personality ratings? *Journal of Personality and Social Psychology,* 1979, *37,* 1055–1074.

Chan, D. W., & Jackson, D. N. Implicit theory of psychopathology. *Multivariate Behavioral Research,* 1979, *14,* 3–19.

Conger, A. J., & Jackson, D. N. Suppressor variables, prediction, and the interpretation of psychological relationships. *Educational and Psychological Measurement,* 1972, *32,* 579–599.

Cronbach, L. J. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychological Bulletin,* 1955, *52,* 177–193.

D'Andrade, R. G. Trait psychology and componential analysis. *American Anthropologist,* 1965, *67,* 215–228.

D'Andrade, R. G. Memory and the assessment of behavior. In H. Blalock (Ed.), *Measurement in the social sciences.* Chicago: Aldine-Atherton, 1974.

Eckart, C., & Young, G. The approximation of one matrix by another of lower rank. *Psychometrika,* 1936, *1,* 211–218.

Gulliksen, H. *Theory of mental tests.* New York: Wiley, 1950.

Jackson, D. N. A model for inferential accuracy. *The Canadian Psychologist,* 1972, *13,* 185–194.

Jackson, D. N. *Personality Research Form Manual.* Port Huron, Mich.: Research Psychologists Press, 1974.

Jackson, D. N., Chan, D. W., & Stricker, L. J. Implicit personality theory: Is it illusory? *Journal of Personality,* 1979, *47,* 1–10.

Jackson, D. N., Neill, J. A., & Bevan, A. R. An evaluation of forced-choice and true-false item formats in personality assessment. *Journal of Research in Personality,* 1973, *7,* 21–30.

Jackson, D. N., & Williams, D. R. Occupational classification in terms of interest patterns. *Journal of Vocational Behavior,* 1975, *6,* 269–280.

Kirk, R. E. *Experimental design: Procedures for the behavioral sciences.* Monterey, Calif.: Brooks/Cole, 1968.

Kusyszyn, I. A comparison of judgmental methods with endorsements in the assessment of personality traits. *Journal of Applied Psychology,* 1968, *52,* 227–233.

Lay, C. H., & Jackson, D. N. Analysis of the generality of trait inferential relationships. *Journal of Personality and Social Psychology,* 1969, *12,* 12–21.

Mirels, H. L. Implicit personality theory and inferential illusions. *Journal of Personality,* 1976, *44,* 467–487.

Mischel, W. *Personality and assessment.* New York: Wiley, 1968.

Mischel, W. *Introduction to Personality.* New York: Holt, Rinehart & Winston, 1976.

Mulaik, S. A. Are personality factor raters' conceptual factors? *Journal of Consulting Psychology,* 1964, *28,* 506–511.

Norman, W. T. Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology,* 1963, *66,* 574–583.

Norman, W. T., & Goldberg, L. R. Raters, ratees, and randomness in personality structure. *Journal of Personality and Social Psychology,* 1966, *4,* 681–691.

Osgood, C. E. The nature and measurement of meaning. *Psychological Bulletin,* 1952, *49,* 197–237.

Paivio, A. Imagery and synchronic thinking. *Canadian Psychological Review,* 1975, *16,* 147–163.

Reed, P. L., & Jackson, D. N. Clinical judgment of psychopathology: A model for inferential accuracy. *Journal of Abnormal Psychology,* 1975, *84,* 475–482.

Schneider, D. J. Implicit personality theory: A review. *Psychological Bulletin,* 1973, *79,* 294–309.

Shweder, R. A. How relevant is an individual difference theory of personality? *Journal of Personality,* 1975, *43,* 455–484.

Skinner, H. A. The eyes that fix you: A model for classification research. *Canadian Psychological Review,* 1977, *18,* 142–151.

Skinner, H. A., Jackson, D. N., & Hoffmann, H. Alcoholic personality types: Identification and correlates. *Journal of Abnormal Psychology,* 1974, *83,* 658–666.

Wiggins, J. S. *Personality and prediction: Principles of personality assessment.* Reading, Mass.: Addison-Wesley, 1973.

Received August 11, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 10, 1660-1672

The Role of Category Accessibility in the Interpretation of Information About Persons: Some Determinants and Implications

Thomas K. Srull and Robert S. Wyer, Jr.
University of Illinois at Urbana-Champaign

Many personality trait terms can be thought of as summary labels for broad conceptual categories that are used to encode information about an individual's behavior into memory. The likelihood that a behavior is encoded in terms of a particular trait category is postulated to be a function of the relative accessibility of that category in memory. In addition, the trait category used to encode a particular behavior is thought to affect subsequent judgments of the person along dimensions to which it is directly or indirectly related. To test these hypotheses, subjects first performed a sentence construction task that activated concepts associated with either hostility (Experiment 1) or kindness (Experiment 2). As part of an ostensibly unrelated impression formation experiment, subjects later read a description of behaviors that were ambiguous with respect to hostility (kindness) and then rated the target person along a variety of trait dimensions. Ratings of the target along these dimensions increased with the number of times that the test concept had previously been activated in the sentence construction task and decreased with the time interval between these prior activations and presentation of the stimulus information to be encoded. Results suggest that category accessibility is a major determinant of the way in which social information is encoded into memory and subsequently used to make judgments. Implications of this for future research and theory development are discussed.

When individuals are asked to judge themselves or another person, they are unlikely to perform an exhaustive search of memory for all cognitions that have implications for this judgment. Rather, they are likely to base their judgment on some subset of these cognitions that is most readily accessible (Tversky & Kahneman, 1973, 1974). Using quite different research methodologies, Carlston (1977) and Lingle and Ostrom (in press) have both shown that once a judgment of a stimulus person has been made on the basis of new information, this judgment is subsequently used as a basis for later inferences about the person independent of the information upon which the judgment was originally based. Similarly, Ross and his colleagues (Ross, Lepper, & Hubbard, 1975; Ross, Lepper, Strack, & Steinmetz, 1977) have found that once a person has constructed an explanation of an event involving himself or another person, this construction, rather than the information that stimulated it, is used to predict the likelihood of future events. Each body of research therefore suggests that the most easily accessible cognitions about an object or event (i.e., those that have been acquired and used most recently) have a major influence on future judgments.

Similar considerations arise when a person

The present research was supported by U.S. Public Health Service Training Grants MH-14257 and MH-15140 to the first author and National Science Foundation Grant BNS 76-24001 to the second author. We thank Frank Costin for generously providing us with many of his stimulus materials. We would also like to extend our sincere appreciation to two anonymous reviewers, each of whom provided us with trenchant but very constructive criticism of an earlier draft of the manuscript.

Requests for reprints should be sent to Thomas K. Srull or Robert S. Wyer, Department of Psychology, University of Illinois, Champaign, Illinois 61820.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3710-1660$00.75

is asked to interpret new information about a social stimulus. In many instances, the information one receives is ambiguous; that is, it can be interpreted in more than one way. For example, information that someone told his girl friend that her new hair style is unattractive could be interpreted, or *encoded,* as either "honest" or "unkind." Which encoding is actually made may depend upon which of the two relevant concepts (honest or unkind) is most easily accessible at the time the information is received (cf. Bruner, 1957). Once the behavior is encoded as an instance of one of these trait concepts, the implications of this encoding, rather than those of the original behavioral information, may be used as a basis for subsequent judgments about the person (Carlston, 1977; Higgins, Rholes, & Jones, 1977). If this is true, judgments of a person may often be affected substantially by rather fortuitous events that lead one or another concept to be more accessible to the judge at the time information about the person is initially received.

This paper reports two in a series of studies designed to investigate the possibility raised above and to explore its implications. In so doing, they supplement and extend an earlier study by Higgins, Rholes, and Jones (1977). These authors reasoned that if subjects were required to use trait terms in the course of performing one task, these terms would become more accessible and therefore more likely to be used to encode subsequent behavioral information about a person in an unrelated context. To test this hypothesis, subjects first performed a color-naming task that required them to remember four trait terms. In experimental conditions, the traits were all potentially applicable for encoding a normatively based set of behavioral descriptions (e.g., thinking about crossing the Atlantic in a sailboat), whereas in control conditions they were all inapplicable. Moreover, in some cases the four terms all had positive evaluative implications (e.g., "adventurous"), while in other cases they all had negative evaluative implications (e.g., "reckless"). As part of a second, ostensibly unrelated experiment, subjects read a story about a stimulus person that contained these behavioral descriptions. Subjects then both wrote a description of the target person in their own words and estimated how much they liked this person. Experimental subjects tended to describe the target person with trait terms that were the same as or synonymous with those to which they had been exposed in the color-naming task. In addition, their subsequent ratings of the target person were biased in the direction of the evaluative implications of these terms. In contrast, control subjects who were exposed to trait terms that were inapplicable for encoding the target's behavior did not vary systematically in either the evaluative implications of their free descriptions of the target or their ratings of him.

The findings of Higgins et al. provide intriguing support for the general hypothesis that once a concept is activated or "primed" as a result of its use for one purpose, its relative accessibility is enhanced, and its likelihood of being used to encode subsequent information increases. Moreover, the implications of the encoding, rather than those of the original stimulus material, are used as a basis for later judgments. Several additional questions are raised by these results, however. First, in the Higgins et al. study, subjects were primed with specific trait labels, each of which presumably represented a particular cognitive category or concept. To the extent that trait and behavioral concepts are interrelated in memory, however, the accessibility of trait concepts may also be increased by priming specific behaviors that exemplify these trait concepts. In this regard, it may be useful to conceptualize the representation of person information in memory in terms of schemata, or configurations of interconnected traits and behaviors at different levels of generality (cf. Cantor & Mischel, 1977). Such schemata may be hierarchical, with trait terms centrally located and behavioral instances of the traits more peripheral. Once a schema is activated, it may then be used to interpret and organize subsequent information (Bransford & Johnson, 1972; Lingle & Ostrom, in press). Thus, exposure to behavioral instances of a trait in one context may activate a schema associated with this trait, and the schema may then serve as a basis for inter-

preting subsequent behavioral information that is received in other contexts (Wyer & Srull, in press).

This reasoning implies that it is not necessary to prime the name of a trait in order to increase the likelihood that the schema associated with it will be used to interpret subsequent information. Rather, one needs only prime behavioral instances of the trait that are represented in the schema. It is conceivable, however, that several such instances must be primed in order to activate the schema itself. One reason for this is that behaviors are often ambiguous; that is, they can be considered instances of several different traits. To this extent, they are less likely to be interpreted as representative of any given trait when considered in isolation. Thus, it is reasonable to expect that the likelihood of activating a trait schema will increase with the number of behavioral instances used to prime it.

Once a trait concept or schema is activated as a result of exposure to representative behaviors, its accessibility, and thus its effect on the interpretation of subsequent information, is likely to decrease over time. This prediction is both intuitively reasonable and formally derived on the basis of several existing theoretical formulations. For example, Wyer and Srull (in press) have developed a model of social information processing in which the accessibility of a primed concept is postulated to decrease as a function of the number of other potentially relevant concepts used during the subsequent time interval. Since the likelihood of using at least one other potentially relevant concept becomes greater as the time interval increases, the likelihood of using the initially primed concept to interpret new information becomes correspondingly less.

Wyer and Carlston (1979) have formulated a model of person memory that also predicts this decrease. According to this formulation, which in many respects is similar to the semantic memory model of Collins and Loftus (1975), the residual "excitation" that remains at the location of a previously activated trait concept decreases over time once the concept is no longer used. As a result, increasingly greater amounts of excitation are required to reactivate the concept as time goes on, and the trait category is less likely to be invoked.[1] The two models therefore differ in that the first attributes the decrease in accessibility to be a function of interference effects (the likelihood of which increases over time), while the second attributes the decrease to be a function of time per se, independent of any interference produced by other concepts.

To summarize, then, the above reasoning implies that the accessibility of a trait schema, and therefore the likehood that it will be used to interpret new information, will increase with the number of schema-related behavioral concepts that have been activated prior to the receipt of this information but will decrease with the length of time between the activation of these concepts and acquisition of the new information to be interpreted. The present set of experiments tests these hypotheses.

The experiments reported extend upon the findings of Higgins et al. in one other important way. Specifically, Higgins et al. found that the accessibility of trait terms that were inapplicable for encoding the behavioral information did not affect subsequent evaluations of the target person. This suggests that priming does not have a direct influence on subjects' judgments of the target but affects these judgments only through its mediating influence on the interpretation of the target's behavior. Once the behavior has been encoded in terms of a given trait (e.g., "hostile"), however, and an overall impression of the target has been formed on the basis of this encoding, the target may then be ascribed other, evaluatively similar traits (e.g., "unintelligent") that are related only on the basis of subjects' implicit personality theories (Rosenberg & Sedlak, 1972). This possibility

---

[1] The effects of priming appear to dissipate very rapidly with the semantic tasks that are typically used to test the Collins and Loftus model. Wyer and Carlston's extension of the model does not postulate this rapid a decay of excitation, however. More important, the stimulus domain and judgment tasks to which the Wyer and Carlston model is theoretically applied are quite different from those considered by Collins and Loftus.

was also explored in the present set of experiments.

Method

Overview

Two experiments were run. The first investigated the effects of priming concepts related to hostility. The second was a conceptual replication in which concepts related to kindness were primed. The procedure used was the same in each experiment. Subjects first performed a "word comprehension" task in which they constructed sentences from sets of words. These sets were constructed so that each sentence completed would describe a behavior either related or unrelated to hostility (kindness). The total number of questionnaire items and the proportion of items related to the concept being primed were both systematically varied. Then, as part of a separate experiment on impression formation, subjects read a paragraph about a hypothetical target person who manifested a series of behaviors that were ambiguous with respect to the trait being primed. Judgments of both the target person and of individual behaviors were then analyzed as a function of the length of the priming questionnaire (30 or 60 items), the proportion of hostile (kind) priming items in the questionnaire (20% or 80%), and the time interval between completion of the priming task and the presentation of the target information (no delay, 1 hour, or 24 hours).

Ninety-six introductory psychology students (8 in each experimental condition) participated in Experiment 1 for course credit, and a different group of 96 students participated in Experiment 2.

Selection of Behavioral Descriptions

Experiment 1. To select behavioral descriptions that varied both in terms of the hostility they conveyed and in terms of the ambiguity of their implications, 43 subjects who did not participate in the main experiment were asked to rate a large pool of individual behaviors along a scale from 0 ("not at all hostile") to 10 ("extremely hostile"). From this pool were selected 5 behaviors that were judged to convey high hostility ($M = 8.08$) and 5 that were judged to convey low hostility ($M = .58$). In addition, 10 "ambiguous" behaviors were selected on the basis of two criteria: first, the mean hostility rating of each ambiguous behavior ($M = 3.99$) was lower than the mean rating of any behavior identified as hostile and higher than that of any behavior identified as nonhostile; second, the standard deviation of ratings for each ambiguous item was greater than the largest standard deviation of any item in either of the other two groups ($SD = 2.76$). The 10 ambiguous items were randomly divided into two groups of 5, and each group was then used to construct a vignette describing a hypothetical target person. In addition, all 20 behaviors were used as test items to be rated individually in a manner to be described.

Experiment 2. A procedure similar to that described above was used to select behavioral descriptions related to kindness. In this case, the mean ratings ($n = 40$) of the items selected as kind, ambiguous, and unkind were 8.71, 4.24, and .81, respectively. Again, the standard deviation of each ambiguous item was greater than the largest standard deviation of any item in the other two groups ($SD = 2.92$).

Procedure

Except where noted, the procedures described were identical in both experiments.

Administration of the priming task. Subjects, who were run in groups of four to eight, were greeted by a male experimenter. The experimenter introduced himself as a graduate student and stated that subjects had not actually been assigned to him but that the "real" experimenter had agreed to let him pretest a word comprehension test he was trying to develop. The exercise was described as a test of how people perceive word relationships based on their first immediate impressions. The task consisted of a number of items adapted from materials developed by Costin (1969, 1975). Each item consisted of a set of four words, and the subject's task was to underline three of the words that would make a complete sentence. The subject was told to complete each item as quickly as possible. Each item listed the words in random order and was constructed in such a way that the subject could form at least two possible sentences. In Experiment 1, however, each possible sentence formed from the hostile priming items (e.g., "leg break arm his") necessarily conveyed hostility while each item formed from other (filler) items (e.g., "her found knew I") did not. In Experiment 2, the filler items were identical, but each possible sentence formed from the kind priming items (e.g., "the hug boy kiss") conveyed kindness.

Although the effects of priming were expected to increase with the number of times an instance of the trait concept was primed, a more diagnostic test of these effects was constructed by varying both the total number of items in the questionnaire and the proportion of these items that was relevant to the trait concept. Specifically, subjects completed a total of 30 or 60 items, either 20% (6 or 12 items) or 80% (24 or 48 items) of which were related to the concept being primed. (Items of each type were distributed randomly throughout the questionnaire.) The influence of the number of priming items per se would then be indicated by significant effects for both of these manipulations. Moreover, the proportion variable should have a greater effect when the total number of questionnaire items is large than when it is small.

To avoid suspicion, all subjects within any given experimental session received a questionnaire of the

same length, but the proportion of priming items was varied. In addition, the particular priming items used in constructing the questionnaire were varied, so that pooled over subjects within each condition, each item occurred the same number of times.

Presentation of stimulus materials. After completing the priming task, the graduate student thanked the subjects for helping him. They were then turned over immediately to the "real" experimenter, told to return 1 hour later, or told to return 24 hours later. (In the latter two cases, subjects had previously been notified through the mail that they were scheduled for those times as well. The early dismissal from the first session was attributed to a small mix-up that prevented one of the planned experiments from being ready on time.)

The second experimenter (a female) then led subjects to believe that the scheduled experiment consisted of three separate and unrelated tasks. The first task (the only one relevant to the present article) was described as a task of impression formation. Subjects were asked to read a short vignette about a stimulus person (Donald) that described a series of events occurring during the course of one afternoon. Two vignettes, serving as stimulus replications, were constructed for use in each experiment. Each vignette contained a different set of five behaviors that were ambiguous with respect to their implications for the primed trait. These behaviors were embedded within other information that was irrelevant to the trait. For example, one of the vignettes used in Experiment 1, which described behaviors that were ambiguous with respect to hostility, was the following.

> I ran into my old acquaintance Donald the other day, and I decided to go over and visit him, since by coincidence we took our vacations at the same time. Soon after I arrived, a salesman knocked at the door, but Donald refused to let him enter. He also told me that he was refusing to pay his rent until the landlord repaints his apartment. We talked for a while, had lunch, and then went out for a ride. We used my car, since Donald's car had broken down that morning, and he told the garage mechanic that he would have to go somewhere else if he couldn't fix his car that same day. We went to the park for about an hour and then stopped at a hardware store. I was sort of preoccupied, but Donald bought some small gadget, and then I heard him demand his money back from the sales clerk. I couldn't find what I was looking for, so we left and walked a few blocks to another store. The Red Cross had set up a stand by the door and asked us to donate blood. Donald lied by saying he had diabetes and therefore could not give blood. It's funny that I hadn't noticed it before, but when we got to the store, we found that it had gone out of business. It was getting kind of late, so I took Donald to pick up his car and we agreed to meet again as soon as possible.

Similar vignettes that were ambiguous with respect to kindness, were used in Experiment 2.

In each case, subjects were asked after reading the vignettes to form an impression of the person described and then rate him along a series of trait dimensions, six of which (hostile, unfriendly, dislikable, kind, considerate, and thoughtful) were assumed to imply either a high or a low degree of hostility/kindness, and the others of which (boring, selfish, narrow-minded, dependable, interesting, and intelligent) were expected to be evaluatively loaded but descriptively unrelated to either hostility or kindness. These ratings were made along a scale from 0 ("not at all") to 10 ("extremely") and in each case included reverse scoring on half the items.

Ratings of individual behaviors. After rating the target person, subjects rated the hostility (in Experiment 1) or kindness (in Experiment 2) conveyed by each of the 20 individual behaviors selected on the basis of the pretest data described earlier. These ratings were made along a scale from 0 ("not at all hostile/kind") to 10 ("extremely hostile/kind").

Ratings of trait co-occurrence. Finally, subjects were asked to estimate the co-occurrence of hostility (in Experiment 1) and kindness (in Experiment 2) with each of the other 11 traits. Items were of the form "If a person is hostile [kind], how likely is it that he is ———?" and were rated along a scale from 0 ("not at all") to 10 ("extremely").

Postexperimental data. To check on the extent to which subjects might have had insight into the objectives of the experiment despite the several precautions taken to dissociate the priming questionnaire from the impression-formation task, subjects in Experiment 2 were asked to indicate which of the four tasks they performed during the experimental session(s) were most likely to be related to the same hypothesis. Following this question, they were also asked to indicate any other tasks they thought might be related.

Collection of Normative Data

To facilitate the interpretation of the expected priming effects, normative data were collected on the 20 behaviors used in each experiment. To avoid possible context effects associated with the large pool of behaviors originally tested, different groups of subjects rated the amount of hostility conveyed by each behavior used in Experiment 1 ($n = 28$) and the amount of kindness conveyed by each behavior used in Experiment 2 ($n = 34$) on scales ranging from 0 ("not at all hostile/kind") to 10 ("extremely hostile/kind"). These ratings were all made under neutral testing conditions and were later used for comparative purposes.

Results

Experiment 1

Preliminary analysis. The traits *unfriendly, dislikable, kind, considerate,* and

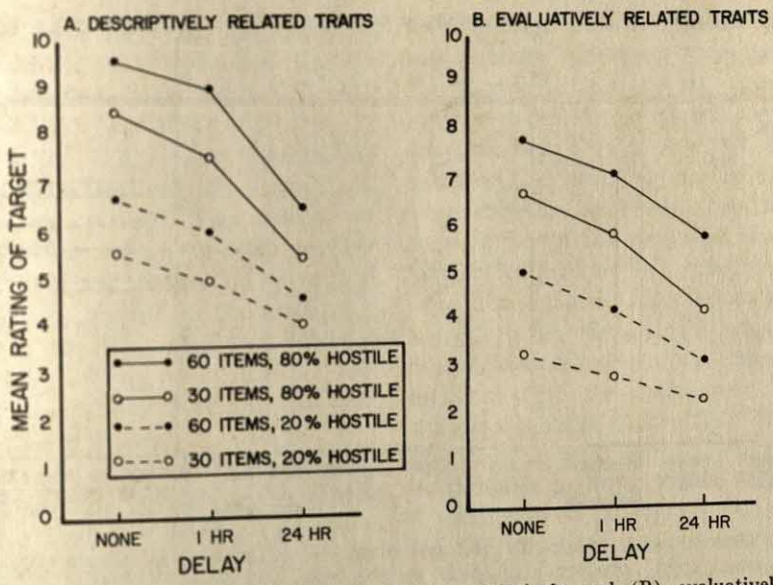


Figure 1. Mean ratings of target person along (A) descriptively and (B) evaluatively related traits to hostility as a function of questionnaire length, proportion of hostile priming items, and delay.

thoughtful were assumed to be denotatively related to the trait *hostile*. Ratings of the estimated co-occurrence of these traits with hostility, which ranged (after reverse scoring on the last three) from 8.79 to 8.98, indicate that they were all thought to covary with hostility to a substantial degree. Thus the six traits were summed across to produce a single index of the perceived hostility of the target. Similarly, ratings of the remaining six traits were summed after appropriate reverse scoring to provide a single index of the unfavorableness of the target along dimensions that are evaluatively loaded but not descriptively related to hostility, as evidenced by co-occurrence ratings ranging from 5.52 to 6.45.

Ratings of target person. Mean ratings of the target along both hostility-related dimensions and evaluative dimensions not directly related to hostility are shown in Figure 1 as a function of the length of the priming questionnaire, the proportion of hostile priming items, and the delay between completion of the priming task and presentation of the stimulus materials. Analyses as a function of these variables and the stimulus replication are relevant to several hypotheses. First, ratings of the target along both sets of dimensions were expected to increase with the number of times hostility-related concepts had previously been activated. Support for this can be seen in Figure 1, which shows that ratings of the target increased monotonically with the number of hostility-related items contained in the questionnaire. The hypothesis is supported statistically by significant main effects of both questionnaire length, $F(1, 72) = 123.74$, $p < .001$, and the proportion of hostile priming items, $F(1, 72) = 590.67$, $p < .001$. If the effect of priming is a linear function of the number of times hostility was previously primed, the effect of proportion should be greater when the questionnaire is long than when it is short. While the interaction of proportion and questionnaire length was not significant ($F < 1$), the pattern of results is consistent with predictions; specifically, when collapsed over delay conditions, the difference in the mean trait ratings of the target between low and high proportion lists was greater for the long questionnaire (2.70) than for the short questionnaire (2.51).

Second, the effect of priming was expected to decrease with the time interval between the priming task and presentation of the stimulus information to be encoded. The data shown in Figure 1 clearly support this hypothesis, which is tested statistically by the main effect

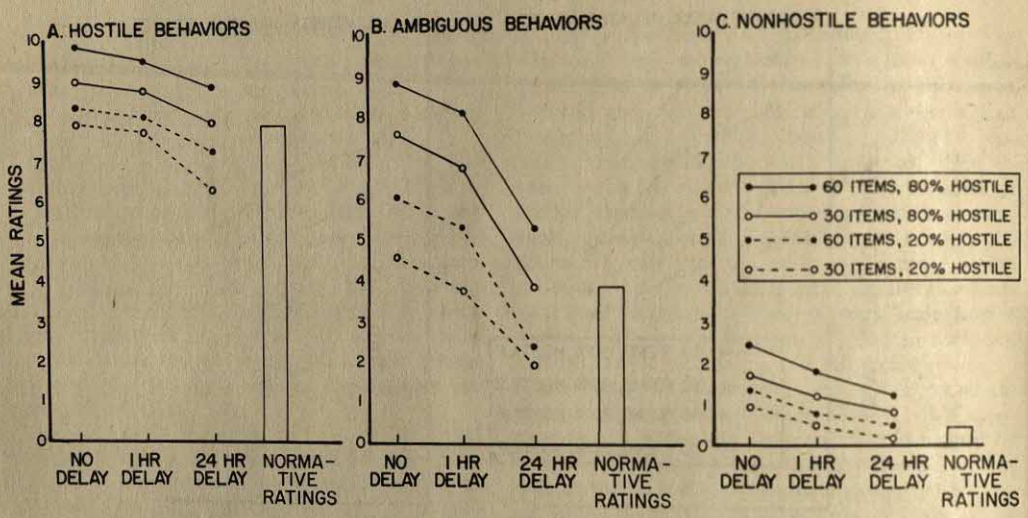


Figure 2. Mean ratings of (A) hostile, (B) ambiguous, and (C) nonhostile behaviors as a function of questionnaire length, proportion of hostile priming items, and delay.

of delay interval, $F(2, 72) = 133.32$, $p < .001$.[2] Figure 1 also suggests that the magnitude of the decrease in priming effects over time is a positive function of the number of times the category was initially activated (i.e., the number of relevant priming items in the questionnaire). Only the interaction of delay and proportion of priming items was statistically significant, $F(2, 72) = 8.30$, $p < .001$, indicating that the effect of delay increased with the proportion of priming items in the initial questionnaire. Since the number of relevant items presented differed more as a function of proportion than as a function of overall questionnaire length, the greater contingency of delay on the former variable is not surprising.

Finally, priming effects on judgments of hostility were expected to generalize to ratings along dimensions that are evaluative but related to hostility only indirectly through subjects' implicit personality theories. Data in the right panel of Figure 1 clearly support this hypothesis. However, these effects were not as pronounced as they were on judgments along dimensions that were descriptively related to the primed concept (left panel of Figure 1). This hypothesis is supported by significant interactions of trait type with both delay, $F(2, 72) = 5.13$, $p < .01$, and the proportion of hostile priming items, $F(1, 72) = 5.70$, $p < .05$. Specifically, the effect of delay interval on judgments of the target along descriptively related dimensions was greater ($M = 7.65$, 6.94, and 5.28 in the immediate, 1-hour, and 24-hour conditions, respectively) than its effect on judgments along evaluative but not descriptively related dimensions ($M = 5.77$, 5.00, and 3.90, respectively). However, the proportion of hostile priming items had less effect on hostility-related judgments ($M = 5.41$ and 7.83 under 20% and 80% conditions) than on evaluative judgments along other dimensions ($M = 3.49$ and 6.29, respectively).

Ratings of individual behaviors. The mean ratings of hostile, ambiguous, and nonhostile behaviors are plotted in Figure 2 as a function of experimental variables. Analyses of these data yielded an obviously significant effect of behavior type, along with the predicted main effects of delay, proportion of hostile priming items, and length of the priming questionnaire (in each case, $p < .001$). By far the greatest effects occurred on ratings of ambiguous behaviors, as evidenced by significant

[2] It should be noted that all of the results reported in this paper are based on analyses of variance that assume homogeneity of treatment–difference variances. The small positive bias that results when this assumption is not completely satisfied (see, e.g., Huynh & Feldt, 1970) would appear insignificant in relation to the general strength of the results.

interactions of behavior type with questionnaire length, $F(2, 144) = 65.64$, $p < .001$; proportion of hostile priming items, $F(2, 144) = 352.95$, $p < .001$; and delay, $F(4, 144) = 223.69$, $p < .001$. The 4-way interaction among these variables was also significant, $F(4, 144) = 3.38$, $p < .02$. This reflects the fact that the interaction among delay, number of priming items, and proportion of critical priming items was greater for the ambiguous behaviors than for the other two behavior types, $F(2, 72) = 4.79$, $p < .02$. In fact, this component accounts for 73.8% of the total interaction sums of squares.

The differences in ratings under the 24-hour delay conditions are sufficient to justify the conclusion that the priming task influenced judgments even after a fairly long time interval had elapsed. However, they do not indicate in an absolute sense whether the delayed ratings were positively affected by priming under all conditions. Evidence bearing on this question is provided by a comparison of these ratings with normative ratings of the behaviors made by subjects who were not exposed to the priming task. These normative ratings are also presented in Figure 2. Unfortunately, the comparisons are not easy to interpret. Subjects who completed priming questionnaires in which only 20% (6 or 12) of the items were hostility-related made *less hostile* ratings of both the hostile and ambiguous behaviors than subjects who received no priming at all. Taken at face value, this suggests that priming under these conditions had a negative effect after a delay of 24 hours. However, since comparable results did not obtain in Experiment 2 (see below) and such negative effects are difficult to account for theoretically, such a conclusion must be treated very cautiously pending replication. (Indeed, it seems more reasonable to attribute the finding to spuriously high normative ratings of the behaviors than to negative effects of priming.)

Experiment 2

Preliminary analysis. The traits *considerate, thoughtful, hostile, unfriendly,* and *dislikable* were assumed to be descriptively related to the trait *kind*. Mean estimates of the co-occurrence of these traits with kindness (after reverse scoring on the last three) ranged from 8.54 to 8.85, indicating that they were all thought to covary with kindness to a substantial degree. Ratings of the six traits were therefore summed across to produce a single index of the perceived kindness of the target. Ratings of the remaining six traits were also summed after appropriate reverse scoring to provide a single index of the perceived favorableness of the target along dimensions that are evaluatively but not descriptively related to kindness. Ratings of the estimated co-occurrence of these traits with kindness ranged from 5.69 to 6.24.

Ratings of target person. Mean ratings of the target person along dimensions both descriptively and evaluatively related to kindness are plotted in Figure 3 as a function of experimental variables. These effects are similar in most respects to those obtained in Experiment 1. The hypothesis that ratings would increase monotonically with the number of times concepts related to kindness were previously activated is supported by main effects for both questionnaire length, $F(1, 72) = 35.62$, $p < .001$, and the proportion of kind priming items in the questionnaire, $F(1, 72) = 158.67$, $p < .001$. Moreover, the effect of proportion was significantly greater when the questionnaire was long ($M = 4.54$ and 6.05 under 20% and 80% conditions) than when it was short ($M = 4.24$ and 5.18, respectively), $F(1, 72) = 8.85$, $p < .01$.

The hypothesis that priming effects would decrease over the time interval between the priming task and stimulus presentations was again strongly supported, $F(2, 72) = 47.79$, $p < .001$. Moreover, the magnitude of this decrease was greater when the proportion of kind priming items in the questionnaire was high than when it was low, $F(2, 72) = 19.66$, $p < .01$, and greater when the questionnaire was long than when it was short, $F(2, 72) = 5.98$, $p < .01$. These findings indicate that the effect of delay is a positive function of the number of times that kindness was initially primed.

Finally, the priming manipulations had very similar effects on ratings of both dimensions that are descriptively related to

kindness and those that were evaluative but descriptively unrelated. This again supports the hypothesis that once the target's behavior is encoded in terms of a trait, it will also be assigned other characteristics that are evaluatively associated with this trait. However, the effect of delay on ratings of descriptively related dimensions ($M = 5.94$, 5.15, and 4.54 in the immediate, 1-hour, and 24-hour conditions, respectively) was greater than for ratings of evaluatively related dimensions ($M = 5.23$, 4.88, and 4.29, respectively), $F(2, 72) = 3.61$, $p < .05$. Moreover, the effect of proportion was greater for descriptively related judgments ($M = 4.55$ and 5.87 under 20% and 80% conditions) than for evaluative judgments along other dimensions ($M = 4.23$ and 5.36, respectively). However, this difference was not reliable, $F(1, 72) = 1.54$, *ns*.

There are two related differences between these data and those obtained in the first experiment. First, the delay interval in Experiment 1 had an appreciable effect at all combinations of questionnaire length and proportion of critical priming items, whereas the effect of delay in the present experiment was negligible when the proportion of priming items was low, $F(2, 72) = 1.68$, *ns*. Second, the effect of the two priming variables after a 24-hour delay was pronounced in the first experiment, but was much less so in this study. In fact, neither questionnaire length ($F < 1$) nor proportion of kind priming items, $F(1, 72) = 3.12$, *ns*, had reliable effects after such a delay. In sum, these data suggest that a greater number of priming items was necessary to increase the accessibility of a schema associated with kindness than was required to increase the accessibility of a schema associated with hostility. Moreover, the accessibility of kindness-related concepts appears to decrease more rapidly over time than does the accessibility of concepts related to hostility.

Ratings of individual behaviors. Mean ratings of individual behaviors designated as kind, ambiguous, and unkind on the basis of normative data are shown in Figure 4 as a function of experimental variables. The effects of these variables are generally similar to their effects on ratings of the target person; that is, the estimated kindness of all three types of behaviors increased with both questionnaire length and the proportion of kind priming items contained in the questionnaire, each $F(1, 72) \geq 43.60$, $p < .001$, while decreasing as a function of the time interval between the priming task and making these estimates, $F(2, 72) = 51.73$, $p < .001$. Moreover, the effects of the time delay increased with both questionnaire length, $F(2, 72) = 6.75$, $p < .01$, and the proportion of kind priming items, $F(2, 72) = 25.75$,

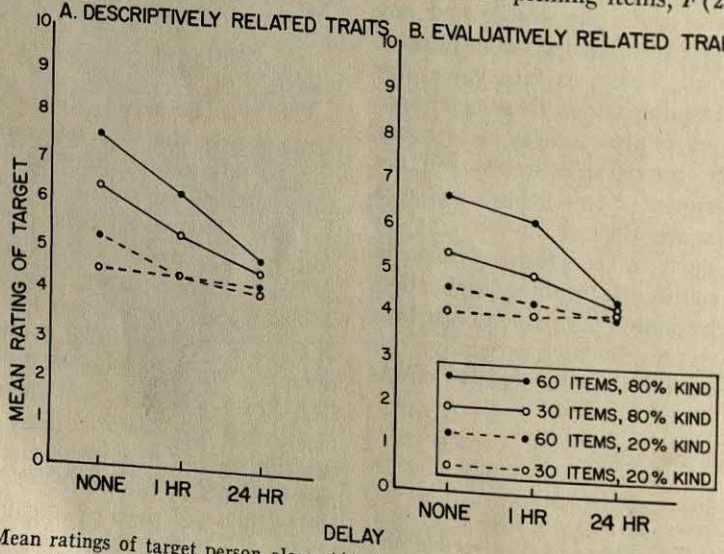


Figure 3. Mean ratings of target person along (A) descriptively and (B) evaluatively related traits to kindness as a function of questionnaire length, proportion of kind priming items, and delay.

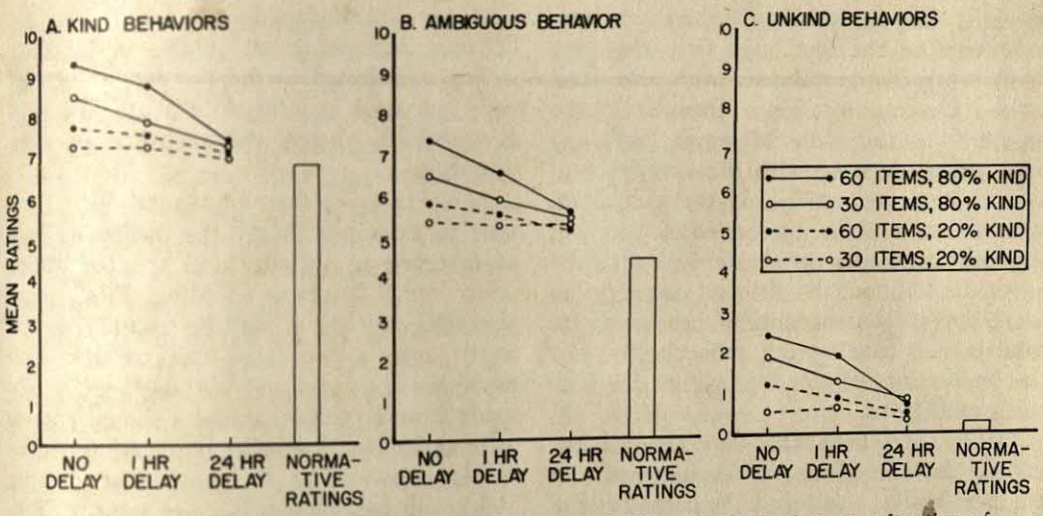


Figure 4. Mean ratings of (A) kind, (B) ambiguous, and (C) unkind behaviors as a function of questionnaire length, proportion of kind priming items, and delay.

$p < .001$. As in the case of target ratings, priming had virtually no effect when only 6 or 12 kindness-related items were involved, and the effect of priming after a 24-hour delay was negligible.

In addition to the above effects, there was an interaction between behavior type and the proportion of kind priming items, $F(2, 144) = 10.53$, $p < .001$. This interaction reflects the fact that the proportion variable had a greater effect on ratings of ambiguous items than on unambiguous ones, $F(1, 72) = 14.42$, $p < .001$. This component accounts for 74.5% of the total interaction sums of squares. Finally, there was also an interaction among behavior type, proportion of kind priming items, and delay, $F(4, 144) = 2.44$, $p < .05$, which is again attributable to the greater effect of experimental variables on ratings of ambiguous behaviors than on ratings of unambiguously kind or unkind behaviors, $F(2, 72) = 3.31$, $p < .05$. This component accounts for 73.9% of the total interaction sums of squares.

The mean normative ratings of each type of behavior under neutral conditions are also shown in Figure 4 for comparison. These ratings, unlike the corresponding ones in Experiment 1 (see Figure 2), are invariably below subjects' ratings in any one of the priming conditions. Thus, each level of priming increased the perceived kindness of all three types of behaviors. However, this effect was much less pronounced after a 24-hour delay than after either of the shorter delays.

Supplementary analysis. Despite the elaborate precautions taken to separate the priming and experimental tasks, a postexperimental questionnaire was administered to determine whether subjects had insight into their actual relatedness. Two findings suggest that subjects were not complying with any implicit demand characteristics. First, only 5 of 96 subjects thought the first (priming) task was related to any of the other three tasks performed during the course of the experimental session(s). Moreover, only 1 of these 5 connected the priming task to the subsequent impression-formation experiment. Thus, subjects were just as likely to relate the priming task to an objectively irrelevant experiment as to the ratings of actual concern. Since each of the four tasks was highly dissimilar, subjects appeared to be guessing randomly and to have no insight at all into the relationship of the two tasks.

Discussion

The two experiments reported in this paper are consistent in their implications for the processing of information about persons. Specifically, once a trait concept or schema is made more accessible by previous cognitive activity, the likelihood that the same schema will be used to encode new information is

increased. The accessibility of these concepts, and therefore the likelihood that they are subsequently used, increases with the number of times that instances of them have been activated in the past. Moreover, although these effects decrease with the time interval between their activation and the acquisition of information to be interpreted or encoded, they are sometimes detectable even after 24 hours. In addition, the effect of category accessibility on the encoding of behavioral information is much more pronounced when the implications of this information are relatively ambiguous.

Finally, once behavioral information is encoded, these encodings affect judgments of the person who manifested the behavior with respect to both the trait originally primed and other traits that are related to it only indirectly through subjects' implicit personality theories. The fact that these effects were typically less on judgments of the latter than of the former traits suggests that this generalizability is not simply due to a halo effect produced by exposure to "good" or "bad" concepts on the priming task. Moreover, this generalization process is very likely to be contingent upon the original encoding of the information about the person being judged. As Higgins et al. (1977) found, increasing the accessibility of a trait concept has little effect on later evaluations of a target person unless the concept is applicable for encoding the information about this person. Once the encoding takes place and an overall impression of the target is formed, however, it may then be used as a basis for inferring traits of the target that are unrelated to either the original behavioral information or the material in the priming task.

In combination, the present experiments and the earlier study reported by Higgins et al. suggest that category accessibility is an important mediator of social judgment processes. However, direct evidence of encoding was not obtained in the present study. Such evidence is suggested by Higgins et al., who found that only the priming of trait terms that were potentially applicable for describing the target's behavior affected subjects' later characterizations of him. Thus, primed traits that were evaluatively toned but inapplicable for describing the target's behavior had no effect. While similar data were not collected in the present study, the paradigm used is sufficiently similar to suggest tentatively that similar processes were operating.

At least two other approaches could be used to examine directly the mediating role of category accessibility and the conditions under which it affects encoding. First, reaction time procedures may be useful. For example, once a particular trait concept such as *hostile* is primed and is thus theoretically made more accessible, subjects should encode information with hostile implications more quickly, as well as be capable of making judgments based upon it more rapidly. Second, our analysis has assumed that encoding effects occur at the time the stimulus information is received. If this is true, priming trait concepts *after* presenting behavioral information about a target (i.e., after the information has already been encoded) should not have any effect on subsequent judgments of the person. Evidence that priming produces no effect under these conditions would provide clear evidence that priming effects are in fact mediated by the encoding of behavioral information.

The difference between the ultimate duration of priming effects obtained in this type of situation and those typically obtained with semantic and lexical decision-making tasks is striking. For example, the effects of priming a familiar noun category (e.g., "bread") on the time required to respond to other concepts that are closely associated with it (e.g., "butter") appear to dissipate in a matter of milliseconds (for a summary, see Schvaneveldt & Meyer, 1973). This suggests that the processes underlying these phenomena are quite different from those that underlie priming effects obtained in the experiments reported here. It also suggests that an "interference" interpretation of the sort postulated by Wyer and Srull (in press) may be more appropriate for describing the processing of complex social stimulus information than the spreading activation approach that is typically used to account for priming effects (cf. Collins & Loftus, 1975). On the other hand, the fact that the present effects

are sometimes still detectable after a period of 24 hours remains intriguing. It is possible that factors such as the experimental laboratory, the presence of the same fellow subjects, and other situational conditions provided a variety of relatively novel contextual cues that were rich enough to "reprime" the trait categories originally made salient. This possibility should also be explored in future research.

The generalizability of the main findings over experiments suggests that these effects are not unique to particular trait concepts or to traits at a given level of favorableness. The amount of priming required to activate different trait concepts appears to differ, however. In the first experiment, as few as six instances of hostile behavior were sufficient to activate the schema related to this trait and thus to affect subsequent judgments. However, many more instances of kindness were apparently required to increase the accessibility of a kindness-related schema. It is possible that this difference is due to unique characteristics of the two traits or the particular set of priming behaviors used. However, to the extent that these materials are representative of favorable and unfavorable traits, it would appear that favorable trait concepts are generally more difficult to activate using the present procedures. Impression formation research (e.g., Birnbaum, 1974; Wyer & Hinkle, 1976) has consistently shown that favorable information has less influence on judgments than unfavorable information does. One reason is that favorable information is typically more ambiguous (for direct evidence of this, see Wyer, 1974). In the present case, since favorable behaviors are socially desirable, instances of these behaviors may be considered less indicative of traits to which they correspond (cf. Jones & Davis, 1965). Therefore, when considered in isolation, these behaviors are less likely to activate a particular trait schema. This possibility may deserve further investigation using a broader sample of trait concepts than those considered in the present studies.

While the effects of manipulations of category accessibility decrease over the interval between the priming task and presentation of the information to be encoded, they appear to increase with the time interval between the presentation of this information and subsequent judgments (Higgins et al., 1977). These two effects should be distinguished. The likelihood that a trait category is accessed and used to interpret information should decrease over the time period since it was most recently activated, for reasons noted earlier. Different considerations arise in explaining the effect of a delay between the encoding of information and judgments. Immediately after exposure to the stimulus information, both the raw information and the encoded representation of it are easily accessible. To the extent that the encoding does not capture all the implications of the original information for the judgment to be made, the judgment may be based on a composite of both. However, the accessibility of the original input material may decrease more rapidly over time than the encoded representation of it, producing the increased effect of encoding reported by Higgins et al. Additional evidence of increased effects of encoding over time has been reported by Carlston (1977), and an investigation of comparable effects within the present paradigm may also be worthwhile.

A final issue to be considered concerns the way in which various target persons will be differentially affected by prior activations of a particular trait schema. The interpretation of information about persons is obviously an overdetermined process, and the accessibility of a particular trait schema is likely to be only one of several determinants. Wyer and Srull (in press) have proposed that schematic representations of specific individuals are built up on the basis of repeated experiences. In this regard, the effects of increased accessibility of a particular trait schema on the encoding of new information may be an inverse function of the amount of information already known about the target person. Thus priming effects may be most pronounced when the target person is previously unknown.

References

Birnbaum, M. H. The nonadditivity of personality impressions. *Journal of Experimental Psychology*, 1974, *102*, 543–561.

Bransford, J. D., & Johnson, M. K. Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, 1972, *11*, 717–726.

Bruner, J. S. On perceptual readiness. *Psychological Review*, 1957, *64*, 123–152.

Cantor, N., & Mischel, W. Traits as prototypes: Effects on recognition memory. *Journal of Personality and Social Psychology*, 1977, *35*, 38–48.

Carlston, D. E. *The recall and use of observed behaviors and inferred traits in social inference processes*. Unpublished PhD dissertation, University of Illinois, 1977.

Collins, A. M., & Loftus, E. F. A spreading-activation theory of semantic processing. *Psychological Review*, 1975, *82*, 407–428.

Costin, F. The scrambled sentence test: A group measure of hostility. *Educational and Psychological Measurement*, 1969, *29*, 461–468.

Costin, F. Measuring hostility with the scrambled sentence test. In L. K. Verma & C. Bagley (Eds.), *Race and education across cultures*. London: Heinemann, 1975.

Higgins, E. T., Rholes, W. J., & Jones, C. R. Category accessibility and impression formation. *Journal of Experimental Social Psychology*, 1977, *13*, 141–154.

Huynh, H., & Feldt, L. S. Conditions under which mean square ratios in repeated measurements designs have exact F-distributions. *Journal of the American Statistical Association*, 1970, *65*, 1582–1589.

Jones, E. E., & Davis, K. E. From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2). New York: Academic Press, 1965.

Lingle, J. H., & Ostrom, T. M. Thematic effects of attitudes on the cognitive processing of attitude relevant information. In R. E. Petty, T. M. Ostrom, & T. C. Brock (Eds.) *Cognitive responses in persuasive communications: A text in attitude change*. Hillsdale, N.J.: Erlbaum, in press.

Rosenberg, S., & Sedlak, A. Structural representations of implicit personality theory. In L. Berkowitz (Ed.) *Advances in experimental social psychology* (Vol. 6). New York: Academic Press, 1972.

Ross, L., Lepper, M., & Hubbard, M. Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm. *Journal of Personality and Social Psychology*, 1975, *32*, 880–892.

Ross, L., Lepper, M. R., Strack, F., & Steinmetz, J. Social explanation and social expectation: Effects of real and hypothetical explanations on subjective likelihood. *Journal of Personality and Social Psychology*, 1977, *35*, 817–829.

Schvaneveldt, R. W., & Meyer, D. E. Retrieval and comparison processes in semantic memory. In S. Kornblum (Ed.) *Attention and performance IV*. New York: Academic Press, 1973.

Tversky, A., & Kahneman, D. Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 1973, *5*, 207–232.

Tversky, A., & Kahneman, D. Judgment under uncertainty: Heuristics and biases. *Science*, 1974, *185*, 1124–1131.

Wyer, R. S. *Cognitive organization and change: An information-processing approach*. Potomac, Md.: Erlbaum, 1974.

Wyer, R. S., & Carlston, D. E. *Social cognition, inference and attribution*. Hillsdale, N.J.: Erlbaum, 1979.

Wyer, R. S., & Hinkle, R. L. Informational factors underlying inferences about hypothetical persons. *Journal of Personality and Social Psychology*, 1976, *34*, 481–495.

Wyer, R. S., & Srull, T. K. The processing of social stimulus information: A conceptual integration. In R. Hastie, E. B. Ebbesen, T. H. Ostrom, R. S. Wyer, D. L. Hamilton, & D. E. Carlston (Eds.), *Person memory: Cognitive bases of social perception*. Hillsdale, N.J.: Erlbaum, in press.

Received August 25, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 10, 1673-1682

Negative and Positive Components of Psychological Masculinity and Femininity and Their Relationships to Self-Reports of Neurotic and Acting Out Behaviors

Janet T. Spence, Robert L. Helmreich, and Carole K. Holahan
The University of Texas at Austin

Negatively valued masculinity (M^-) and femininity (F^-) personality scales were developed to supplement the positively valued Masculinity (M^+) and Femininity (F^+) scales of the Personal Attributes Questionnaire (PAQ; Spence & Helmreich). M^- consisted of traits that had been judged to be (a) more typical of males than females, (b) undesirable in both sexes, and (c) agentic or instrumental in content. Two F^- scales were developed, both containing stereotypically feminine, undesirable traits, one set of traits referring to communionlike characteristics (F_C^-) and the other to verbal passive–aggressive qualities (F_{VA}^-). Significant sex differences in the predicted direction were found on all scales. In both sexes, low and typically nonsignificant correlations were found between parallel positive and negative scales, but highly significant negative correlations were found between positive and negative cross-sex scales. These findings provide additional evidence for the multidimensionality of masculinity and femininity. Scores on a self-esteem measure were positively correlated with M^+ and F^+, uncorrelated with M^-, and negatively correlated with the F^- scales. Different patterns of scores were associated with two types of problem behaviors. In both sexes, neuroticism was most highly correlated (in a negative direction) with M^+, and acting out behavior was most strongly correlated (in a positive direction) with M^-. The next highest correlation in both instances was with F_{VA}^-.

Studies of trait stereotypes (e.g., Rosenkrantz, Vogel, Bee, Broverman, & Broverman, 1968; Spence, Helmreich, & Stapp, 1974, 1975) have consistently demonstrated that the typical male and female are perceived as differing in a number of personality attributes. Males are reported to be higher than females in a cluster of characteristics reflecting personal competencies and goal orientation, whereas women are reported to be higher in a cluster of characteristics reflecting social–emotional sensitivity and an interpersonal orientation. The same differences in the personalities of men and women are frequently mentioned in the discussions of social-psychological theorists. Parsons and Bales (1955), for example, have distinguished between the extradomestic, instrumental role responsibilities assigned to men in most societies and the expressive, domestic role responsibilities assigned to women and have proposed that these differential role assignments are paralleled by underlying differences in the relative strengths of instrumental and expressive characteristics in the two sexes. Similarly, Bakan (1966) has identified two fundamental properties that characterize living organisms: a sense of *agency*, manifested in such characteristics as self-assertion, self-protectiveness, and self-aggrandizement, and a sense of *communion*, manifested in selflessness and a desire to be at one with others. Bakan has further identified agency as the "male principle," stronger in males than in females, and communion as the "fe-

This study was supported by National Science Foundation Grant BNS-78-08911 (Janet T. Spence and Robert L. Helmreich, Principal Investigators).

Requests for reprints should be sent to Janet T. Spence, Department of Psychology, University of Texas at Austin, Austin, Texas 78712.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3710-1673$00.75

male principle," stronger in females than in males.

It has frequently been assumed by psychological theorists that along with other attributes that differentiate the sexes and thus define "masculinity" and "femininity," these clusters of instrumental, agentic traits and expressive, communal traits are essentially incompatible and may thus be treated as endpoints of a single masculinity–femininity continuum. This proposition is joined with still a further assumption—that appropriate sex-typing has beneficial consequences for the individual, the masculine male and the feminine female exhibiting a higher degree of social adjustment and psychological health than those who deviate from the patterns of behaviors and psychological characteristics expected of their sex. More recently, these views have been challenged by a number of investigators (e.g., Bem, 1974; Block, 1973; Carlson, 1971; Constantinople, 1973; Spence & Helmreich, 1978; Spence, Helmreich, & Stapp, 1975) who have proposed that "masculine" instrumental characteristics and "feminine" expressive characteristics form separate dimensions that not only vary independently but also contribute positively to effective functioning in members of *both* sexes.

Data supporting the implications of these revisionist views have come from investigations employing two personality instruments, the Personal Attributes Questionnaire (PAQ; Spence & Helmreich, 1978; Spence, Helmreich, & Stapp, 1975) and the Bem Sex Role Inventory (BSRI; Bem, 1974), both of which contain separate Masculinity (M) and Femininity (F) scales. The PAQ M scale contains a series of socially desirable agentic traits, and the F scale a series of socially desirable communal traits. Although the BSRI is more mixed in content, the bulk of the items on the M and F scales may be similarly described.

Results obtained with these self-report instruments have consistently demonstrated sex differences on the M and F scales in the predicted direction, thus lending support to the common belief that the sexes differ in relative degree of agentic and communal characteristics. Within each sex, however, the correlation between scales has been found to be close to zero, and a number of individuals of both sexes have been found who score relatively high on both scales ("androgynous" individuals) or relatively low on both scales ("undifferentiated" individuals). The empirical data (e.g., Bem, 1977, Spence et al., 1975; Spence & Helmreich, 1978) also suggest that M scores, and to a lesser extent F scores, are positively associated with a number of indices of adjustment and social competence, irrespective of the sex of the individual. Thus "androgynous" individuals tend to be more socially effective than those who are sex typed.

It is critical to note that theoretical discussions of personality differences between the sexes have focused on socially desirable characteristics that purportedly contribute to execution of socially approved sex roles; similarly, studies of trait stereotypes have concentrated on identifying perceived differences between men and women in positively sanctioned attributes. In this same mode, the M and F scales of the PAQ and, to a large extent, of the BSRI,[1] are confined to the measurement of socially desirable traits.

Both observation and psychological theory suggest, however, that even within the category of agentic and communal traits, there are a number of "masculine" and "feminine" characteristics that are socially *undesirable* and have consequences for their possessors or those about them that may be deleterious. Bakan's (1966) theorizing is particularly relevant to this point. He proposes that a strong sense of agency, unmitigated by a sense of communion, is destructive to the individual and to society. Similarly, com-

---

[1] Ratings of the ideal man and woman on each of the BSRI items obtained by Gilbert, Strahan, and Deutsch (Note 1) indicate that on four of the F-scale items (shy, gullible, childlike, and soft-spoken), the means fell toward the nonfeminine pole for both sexes, thus suggesting their social undesirability. In addition, mean ratings of the two sexes fell toward the opposite pole on 6 of the 20 M and two of the 20 F items. As will be noted subsequently, items that corresponded to this pattern on the original PAQ (Spence et al., 1975) were assigned to a third, bipolar scale labeled masculinity–femininity (M–F).

munion must be mitigated by agency if the individual is to function effectively. The developmental task of males is thus to learn to balance masculine agency with some degree of communion, and of females, to balance feminine communion with some degree of agency.

In the present article, we will describe an extension of the PAQ to include scales tapping some of the socially undesirable components of psychological masculinity and femininity and will present data that explore the implications of these scales for self-esteem and two types of problem behaviors. Prior studies (Kelly, Caudill, Hathorn, & O'Brien, 1977; Heilbrun, Note 2) confirm the observation that negatively valued as well as positively valued attributes distinguish the sexes. A number of undesirable traits identified in these studies as being more characteristic of males than of females refer to the *absence* or *opposite* of desirable expressive, communal characteristics, whereas those identified as being more characteristic of females often refer to the *absence* or *opposite* of desirable instrumental, agentic characteristics. Setting up scales of undesirable attributes containing such items would be likely to provide no conceptual or empirical information not already provided by the existing PAQ M and F scales. Still other attributes identified in these studies are not easily classified into agentic versus communal categories and are of unknown theoretical significance. In expanding our instrument to include the negatively toned masculine and feminine attributes, it was our intent to develop scales that conceptually parallel the M and F scales, containing negatively valued masculine characteristics that reflect Bakan's "unmitigated agency" and negatively valued feminine characteristics that reflect "unmitigated communion." The procedures used to develop these scales and their content will be described in a later section.

Correlations between the original (socially desirable) M and F scales, it will be recalled, are close to zero and tend to be positive in sign. While similar relationships might occur between the negatively valued M and F scales, Bakan's (1966) theory suggests that stronger relationships that are *negative* in sign might be found between the original F scale (positive F or F^+) and the new, socially undesirable M scale (negative M or M^-). That is, these M^- characteristics (representing unmitigated agency) would tend to appear in those relatively lacking in socially desirable communal attributes. Similarly, a negative relationship might be expected between the positive M (M^+) and the negative F (F^-) scales, traits reflecting unmitigated communion tending not to appear in those with a healthy sense of self. Predictions about the relationships between parallel positive and negative scales (M^+ and M^-, F^+ and F^-) are more problematic. One might argue that those exhibiting one cluster of masculine or feminine traits would be more likely to exhibit a second cluster of sex-linked traits, or conversely, that those exhibiting a particular cluster of socially *desirable* characteristics would in general be less likely to exhibit socially *undesirable* characteristics. Perhaps, within a group of individuals, these opposing tendencies could be expected to cancel each other, leading to a minimal relationship between positive and negative scales belonging to the same sex-typed category.

In addition to obtaining normative and correlational data on the expanded PAQ to test these possibilities, we administered to the subjects the Texas Social Behavior Inventory (TSBI; Helmreich & Stapp, 1974), a measure of self-esteem and social competence, and a biographical questionnaire that inquired into the incidence of two types of problem behaviors: emotional distress of a neurotic nature, and sociopathic, acting out behaviors. In prior research (e.g., Spence & Helmreich, 1978), the TSBI has been found to be strongly correlated in both sexes with M^+ scores and moderately correlated with F^+ scores. Bakan's theorizing about the self-destructive implications of unmitigated agency and communion led to the expectation that considerably weaker if not negative relationships would be found between our measure of self-esteem and social competence and the negative EPAQ scales. A negative correlation with F^- (which reflects a lack of strong sense of self) seemed particularly likely.

In including the biographical question-

naire, we were exploring the possibility that different patterns of positive and negative M and F scale scores would be related to neurotic and acting out behaviors. Our expectations were clearest for the M scales. We anticipated that of all the scales, M^+ would be most strongly associated (in a negative direction) with emotional problems of a neurotic nature, that is, men and women exhibiting these difficulties would tend to be low in socially desirable agentic qualities. In conformity with prior studies, male students were expected to report more acting out behaviors than females, even with personality variables controlled. Within each sex, however, we predicted that M^- would show the strongest relationship with these behaviors in a positive direction, that is, men and women high in acting out would also tend to be high in the undesirable aspects of agency.

Method

Development of Masculinity and Femininity Scales

The development of the positive PAQ scales is described in detail elsewhere (Spence & Helmreich, 1978). Briefly, the instrument in its present form consists of 24 trait dimensions, drawn from those that had been judged in pilot work (Spence et al., 1974) to differentiate stereotypically between the typical man and woman. Ratings were also obtained of where the ideal man and woman fell on each trait dimension. These ratings were used to assign the items to three 8-item subscales. Masculinity (M^+) items were defined as those in which the mean ratings of the ideal man and woman both fell toward the stereotypically masculine pole; femininity (F^+) items were those in which the mean ideal ratings both fell toward the stereotypically feminine pole. Items were also selected to reflect agentic and communal traits for the M^+ and F^+ scales, respectively. A third group of items was identified, with items that were bipolar with respect to social desirability, that is, the ideal man fell toward the masculine pole and the ideal woman toward the feminine pole, and both were assigned to a separate scale labeled masculinity–femininity ($M–F^+$). The $M–F^+$ scale contains 2 agentic items (aggressive and dominant) and 6 items of more communal, expressive content describing emotional and personal vulnerability (e.g., cries easily, feelings easily hurt, high need for security).

Similar procedures were used in the development of the negative masculinity (M^-) and femininity (F^-) scales. Based primarily on pilot work by one of our students, Kirk Heilbrun (Note 2), with an item pool drawn largely from the Adjective Check List, a number of traits were identified whose presence was (a) judged to be socially undesirable for members of *both* sexes, (b) attributed more frequently to males than to females, and (c) agentic in content. The first two criteria were verified by submitting the items to additional groups of male and female college students, with instructions to rate the typical or the ideal member of each sex. In all instances, significant differences in the mean ratings of the typical male and female were obtained, whereas mean ratings of the ideal member of each sex fell toward the pole that indicated a relative *absence* of the stereotypically masculine trait. Agentic content was determined by ratings of the investigators and their graduate students. Eight items meeting these criteria were selected for the negatively valued masculinity (M^-) scale. (Eight items were chosen to equal the number of each of the socially desirable scales on the 24-item PAQ.) These items are: arrogant; boastful; egotistical; greedy; dictatorial; cynical; looks out only for self; and hostile. Each item is set up on a 5-point scale, with one extreme indicating a high degree of the trait (e.g., very arrogant) and the other a low degree (e.g., not at all arrogant).

The same procedures were used to identify trait items that are attributed more frequently to women than to men but at the same time are considered to be undesirable in both sexes. However, we were less successful in finding stereotypically feminine characteristics that appeared to reflect the excessive selfishness implied by negative communion. We abandoned our original goal as impossible to achieve fully and chose instead to investigate two sets of undesirable, feminine items (four items per set) because of the possibility that they would yield interesting data. The first cluster of four items comes close to unmitigated communion and is identified as F_C^-. The second cluster of four items describes a type of verbal passive–aggressiveness and is identified as F_{VA}^-. The items on the F_C^- scale are: spineless, servile, gullible, and subordinates self to others. Those on the F_{VA}^- scale are: whiny, complaining, fussy, and nagging. Items are set up on a 5-point scale, the extremes indicating a high and low degree of the trait.

Extended Personal Attributes Questionnaire (EPAQ). The negative and positive scales are combined to form a 40-item extended PAQ (EPAQ), each item being accompanied by a 5-point scale and scored from 0 to 4.[2] The M^+, M^-, and $M–F^+$ items are scored in a masculine direction and the F^+, F_{VA}^-, and F_C^- scales in a feminine direction.

[2] Copies of the scales and additional information on their psychometric properties can be obtained from the authors. The results of factor analyses are reported in Helmreich, Spence, & Wilhelm (in press).

Item scores are summed to obtain a total score for each scale.

Other Instruments

Texas Social Behavior Inventory. The 16-item short form of the TSBI (Helmreich & Stapp, 1974) was used as a self-report measure of self-esteem and social competence. Items are scored 0–4 and are summed to yield a total score.

Biographical questionnaire. The biographical questionnaire contained 39 items that asked subjects to indicate the frequency of certain kinds of behaviors or events at one or more periods of their lives (college, high school, grade school) and their current attitudes and feelings about certain aspects of their lives. After factor and other analyses, certain items were dropped, either because they had no predictive value (frequency of illness, level of school performance) or because their occurrence was too infrequent to provide useful data (appearance before juvenile court, involuntary psychological counseling). The remaining items were assigned to 10 scales, 8 of which described acting out or sociopathic behavior and 2 of which described emotional distress of a more neurotic nature. The acting out scales contained items referring to frequency of occurrence of (a) current use of alcohol and other drugs; (b) misdemeanors such as property destruction, shoplifting, and other minor thefts at three age periods (grade school, high school, and college); (c) lying at three age periods; (d) verbal and physical fights in grade school; (e) verbal fights in high school and college; (f) physical fights in high school and college; (g) school misbehavior in high school; (h) school misbehavior in grade school. These scales were positively correlated with each other in both sexes and entered into similar relationships with the EPAQ scales. They were therefore combined for purposes of data reduction into a single acting out scale.

The neurotic scales inquired primarily about current feelings. The first of these scales contained items referring to depression, certainty of life goals, satisfaction with social life, general life satisfaction, and voluntary seeking of professional help for psychological problems. The second neurotic scale contained items referring to frequency of feeling nervous, tense, fearful, and anxious. These scales were also significantly positively correlated and were combined to form a single neuroticism index.

Subjects and Procedures

The test battery, which included, in order, the EPAQ, the TSBI, and the biographical questionnaire, was administered to 220 male and 363 female students in introductory psychology courses at the University of Texas at Austin as part of a course requirement. The battery was given in mixed sex groups and was administered by a male and a female experimenter.

Results

EPAQ Scales

A series of t tests comparing the means of the male and female subjects indicated that the sexes differed significantly ($ps < .001$) in the predicted direction on all six scales. Except for the M–F$^+$ scale, the means of both sexes fell toward the socially desirable pole, which in the case of the negative scales indicated a relatively low degree of the indicated trait. On the M–F$^+$ scale, the mean for males fell slightly toward the masculine pole, and for the females, slightly toward the feminine pole. Since their values were highly similar to those previously reported (Spence & Helmreich, 1978), the means on the positive PAQ scales will not be given here. For the negative scales, the means for males and females, respectively, were 13.63 and 12.21 for M$^-$, 6.39 and 7.28 for F$_C^-$, and 5.28 and 6.48 for F$_{VA}^-$. (The maximum possible score is 32 for M$^-$ and 16 for F$_C^-$ and F$_{VA}^-$.)

Correlations between the EPAQ measures are shown for each sex in Table 1. The rs between the three original PAQ scales replicate those of previous studies (Spence & Helmreich, 1978), M$^+$ and F$^+$ having a slight positive correlation and M–F$^+$ having a substantial positive correlation with M$^+$ and a somewhat lower negative correlation with F$^+$. Turning to the negative scales, it will be noted that as with the positive scales, the values of the coefficients are similar for the two sexes. It will also be observed that the correlations between the two F$^-$ scales are positive but low, suggesting the desirability of scoring these two scales separately.

Particularly informative are the correlations between the parallel positive and negative scales. For M$^+$ versus M$^-$, and F$^+$ versus F$_C^-$, the rs are positive in both sexes, and for F$^+$ versus F$_{VA}^-$, the rs are both negative. However, even when the values are significant, they are small in magnitude, the highest being only .15. (The bipolar M–F$^+$ scale, scored in a masculine direction, has a somewhat higher positive correlation with M$^-$ than does M$^+$, .20 for males and .17 for females, $p < .05$.) For all practical purposes, then, the comparable positive and negative scales

may be treated as independent and may thus enter into very different relationships with other variables. These data also provide further reassuring evidence of a lack of contaminating social desirability response bias in the EPAQ scales; were such a bias systematically distorting subjects' responses, a substantial negative correlation would be expected between counterpart positive and negative scales.

As was expected, more substantial correlations, all significant and negative in direction, were found between the *cross-typed* positive and negative scales. Those high in M^+ thus tended to be low on both F^- scales. These same negative relationships were also found for the $M–F^+$ scale, the correlation between $M–F^+$ and F_C^- in both sexes being particularly marked. Similarly, those high in F^+ tended to be low in M^-. Finally, M^- showed a substantial positive correlation with F_{VA}^- and a negative but small correlation with F_C^-.

In previous studies we have found it heuristically useful to classify subjects jointly on their M^+ and F^+ scores by a median split method to reveal how M^+ and F^+ combine to influence some dependent variable. The present subjects were categorized into four groups by this method, using college student norms (Spence & Helmreich, 1978): Androgynous (above the combined male and female medians on both M and F), Masculine (above on M, below on F), Feminine (below on M, above on F), and Undifferentiated (below on both). Means were next determined for males and females in each PAQ category on the three negative scales. Simple analyses of variance indicated that for each sex and scale, significant differences ($p < .001$) occurred across categorical groups. Inspection of the means, shown in Table 2, suggests that in both sexes, M^+ and F^+ combine additively to determine M^-. That is, with M^+ having a slight positive correlation and F^+ a more substantial negative correlation with M^-, one would anticipate from an additive combination that the Masculine group would have the highest mean on M^-, followed by the Undifferentiated, Androgynous, and Feminine groups. This order obtained in both sexes. The data from the F^- scales also produced the ordering of PAQ groups predicted by an additive combination of M^+ and F^+, the one exception being several slight inversions for females on F_{VA}^-. In view of magnitudes of the underlying correlations, the latter result probably reflects sampling error. What should be noted about these orders is that

Table 1
Correlations of EPAQ Measures in Males and Females

Scale	1	2	3	4	5	6
1. M^+	—	.14	.50	.12	−.28	−.38
2. F^+	.09	—	−.30	−.33	−.09	.07
3. $M–F^+$	.45	−.29	—	.20	−.24	−.45
4. M^-	.10	−.47	.17	—	.40	−.09
5. F_{VA}^-	−.29	−.14	−.27	.45	—	.28
6. F_C^-	−.30	.15	−.40	−.14	.07	—

Note. EPAQ = Extended Personal Attributes Questionnaire. Males are above the diagonal, and females are below. For males, $df = 227$, $r_{.05} = .13$. For females, $df = 361$, $r_{.05} = .10$. M^+ = positively valued Masculinity Scale. F^+ = positively valued Femininity Scale. $M–F^+$ = positively valued Masculinity–Femininity Scale. M^- = negatively valued Masculinity Scale. F_{VA}^- = Feminine Verbal Aggressive Scale. F_C^- = Feminine Communion Scale.

Table 2
Means on the Negative Scales for Males and Females Classified by Category on M^+ and F^+

Negative scale	PAQ category: Undifferentiated	Feminine	Masculine	Androgynous
		Males		
M^-	14.00	10.39	15.01	12.87
M_{VA}^-	6.87	4.75	4.53	5.10
F_C^-	7.15	7.32	5.58	6.22
		Females		
M^-	13.93	10.31	16.59	11.74
F_{VA}^-	7.13	6.89	7.08	5.58
F_C^-	7.51	7.89	5.54	7.09

Note. PAQ = Personal Attributes Questionnaire. M^+ = positively valued Masculinity Scale. F^+ = positively valued Femininity Scale. $M–F^+$ = positively valued Masculinity–Femininity Scale. M^- = negatively valued Masculinity Scale. F_{VA}^- = Feminine Verbal Aggressive Scale. F_C^- = Feminine Communion Scale.

Table 3
Correlations between EPAQ Scales and Self-Esteem, Neuroticism, and Acting Out Measures for Males and Females

Scale	Self-esteem		Neuroticism		Acting out	
	Males	Females	Males	Females	Males	Females
M^+	.66***	.68***	−.53***	−.41***	−.09	−.01
F^+	.23***	.20***	−.06	−.06	−.17**	−.02
$M\text{-}F^+$	.39***	.37***	−.43***	−.27***	.03	−.02
M^-	.02	.00	.13	.19***	.26***	.18***
F_{VA}^-	−.16**	−.30***	.33***	.31***	.22***	.09
F_C^-	−.40***	−.25***	.24***	.04	.13*	.08

Note. EPAQ = Extended Personal Attributes Questionnaire. M^+ = positively valued Masculinity Scale. F^+ = positively valued Femininity Scale. $M\text{-}F^+$ = positively valued Masculinity–Femininity Scale. M^- = negatively valued Masculinity Scale. F_{VA}^- = Feminine Verbal Aggressive Scale. F_C^- = Feminine Communion Scale.
* $p < .05$. ** $p < .01$. *** $p < .001$.

they differ for each negative scale, so that no categorical group is uniformly highest or lowest. For each scale, however, the Undifferentiated group exhibits the highest or second highest amount of the undesirable masculine or feminine qualities, and the Androgynous group the lowest or second lowest amount.

EPAQ and Self-Esteem

In previous studies with the PAQ (Spence et al., 1975; Spence & Helmreich, 1978), it has been demonstrated that in both sexes all three scales are positively related to scores on the TSBI, used as a measure of self-esteem and social competence. Substantial correlations are found with M^+, and lower but highly significant correlations with $M\text{-}F^+$ and F^+. Inspection of the correlations obtained in the present samples of males and females, shown in Table 3, reveals values highly similar to those of prior investigations. As had been anticipated, different relationships were found with the negative scales. In both sexes, the correlations with the self-esteem measure are close to zero for M^- and significantly negative for the F^- scales. Particularly marked is the relationship with F_C^- in males ($r = -.40$).

To examine the joint contributions of the EPAQ scales to self-esteem, males and females were each divided into three self-esteem groups, those falling into the upper quarter, middle half, and lower quarter of their respective score distributions. Multivariate analyses of variance on the scales were then computed for each sex and were found to be highly significant in each sex ($ps < .0001$). Inspection of the pattern of means reveals that in both sexes, systematic trends occurred on all scales except M^-. Scores increased from low to high self-esteem groups on all three positive scales and decreased on the two F^- scales. Univariate Fs were significant ($ps < .01$) except in the case of males and F_{VA}^-.

EPAQ and Neuroticism

Scores on our measure of neuroticism and of self-esteem and social competence turned out to be negatively related (rs of $-.55$ and $-.36$ for males and females, respectively, $ps < .001$). Although it was not surprising to find that those who perceived themselves as being socially effective tended to report experiencing less anxiety, depression, and dissatisfaction than others did, the contents of the two scales (and the underlying psychological phenomena that they were designed to reflect) are not mirror images of each other. It was therefore anticipated that the relationships between the EPAQ scales and the TSBI would not necessarily be highly similar to those between the former and the neuroticism measure.

The correlations between neuroticism and EPAQ scores, reported in Table 3, showed

that this is indeed the case. As expected, the strongest relationship is with M^+ (*r*s of −.53 and −.41 in males and females). Smaller but significantly negative correlations are also found with $M\text{–}F^+$ (*r*s of −.43 and −.27). However, the correlations with F^+ are nonsignificant. Thus, expressiveness and interpersonal skills, as measured by F^+, appear to contribute beneficially to social competence and self-esteem but appear to be unrelated to the kind of emotional disturbances—anxiety, depression, and so forth—tapped by the neuroticism scale.

The three negative scales correlate positively with neuroticism in both sexes, with the largest relationship being found with F_{VA}^- (*r*s of .33 and .31). In females, M^- had a significant (but modest) correlation, whereas F_C^- did not. In males the opposite pattern was found: F_C^- had a significant negative correlation, whereas M^- was nonsignificant.

Next, the males and females were each divided into three neuroticism groups, those falling into the upper quarter, middle half, and lower quarter of their respective score distributions, in order to determine the conjoint contributions of the EPAQ scales. Multivariate analyses of variance on the scales were highly significant in each sex ($ps < .0001$). Inspection of the pattern of means and the univariate *F*s reveals that the high neurotic males are significantly lowest in M^+ and $M\text{–}F^+$ and highest in F_{VA}^- and F_C^-, with a nonsignificant elevation in M^-. High neurotic females are also significantly lowest in M^+ and $M\text{–}F^+$ (and nonsignificantly on F^+) and highest on the three negative scales. For women, however, the effects of M^- are highly significant, whereas those for F_C^- are not. It is possible that the presence of negative cross-sex-typed characteristics is more consistent with emotional disturbance than is the manifestation of sex-congruent negative attributes.

A final analysis was undertaken in which subjects were classified as above or below the median on M^+ and F_{VA}^-, the two scales showing the strongest correlations in each sex. (For this purpose the rounded mean of the medians of males and females on each scale was used as the cutting point.) The percentage of subjects in each neuroticism group showing the most negative constellation (low M^+, high F_{VA}^-) was computed. For males, the low M^+, high F_{VA}^- pattern was found in 4% of the low neuroticism group, 17% of the medium, and 47% of the high. The comparable figures for females were 20% low, 35% medium, and 61% high.

EPAQ and Acting Out Behaviors

Inspection of the acting scores indicated that in both sexes, the distributions were moderately skewed; few of the subjects (particularly females) admitted to having manifested these behaviors to any marked degree. The scores were therefore subjected to a *z* transformation, and statistical analyses performed on the transformed scores.

Pearson correlations between acting out scores and scores on the self-esteem and neuroticism measures were not significant in either sex. Correlations between acting out scores and scores on each of the EPAQ scales for each sex are shown in Table 3. As anticipated, the highest correlations are between acting out and M^-. In both sexes, the values are both positive and highly significant, the relationship being more marked in males than in females. (The mean acting out score for males is also significantly higher than that for females.) For females, no other correlation reached significance, although the signs are in reasonable directions. For males, a significant negative correlation is found with F^+ and a significant positive correlation with F_{VA}^- and F_C^-.

To determine the joint contributions of the EPAQ scales to acting out, males and females were divided into three acting out groups, these falling into the upper quarter, middle half, and lower quarter. Multivariate analyses of variance on the EPAQ scales were again computed for each sex and were found to be significant ($ps < .001$) in both sexes, although the results are stronger in males. Examination of group means and univariate *F*s indicated that scores in the high acting out groups are elevated significantly on M^- and F_{VA}^- in each sex. A similar linear pattern is also found on F_C^- but is significant only in females.

The subjects were also classified as being above or below the median on M^- and F_{VA}^-, the scales most strongly associated with acting out. Those exhibiting the most unfavorable combination, above the median on both M^- and F_{VA}^-, were then identified. In males, 63% of the high acting out group showed this high M^- and F_{VA}^- constellation, whereas only 30% of the moderate and 27% of the low groups were so classified. In females, the figures were 55% for high, 45% for medium, and 35% for low.

Discussion

On both the M^- and the two F^- scales, the mean self-report scores of men and women differed significantly in the direction predicted by the stereotype ratings. Despite the sex differences on the negative as well as the positive scales, the correlations between counterpart positive and negative scales were small and, typically, nonsignificant. Thus in neither sex did socially desirable and undesirable masculine or feminine traits tend to covary. Masculinity and femininity have typically been treated as if they were unidimensional constructs, those exhibiting a given cluster of masculine or feminine behaviors or attributes being assumed to possess other types of masculine or feminine characteristics to the same degree. The present results join a mounting body of evidence suggesting instead that the correlations among the empirically diverse categories of attributes and behaviors distinguishing the sexes are often low. When significant correlations *are* found between two clusters of masculine or feminine characteristics, the relationship may be attributable more to the particular characteristics involved than to the mere fact that they "belong" to the same sex-typed category.

The multifaceted nature of masculinity and femininity is further demonstrated by the pattern of correlations between cross-typed scales. The relationship between M^+ and F^+, as has repeatedly been found in prior studies, is low and positive in both sexes. More substantial negative correlations occur between cross-typed positive and negative scales, as was anticipated. Those high in undesirable agentic qualities, as reflected on the M^- scale, tended to exhibit a relatively low degree of desirable, interpersonally oriented qualities, as reflected on the F^+ scale. Similarly, those high in the socially desirable agentic qualities on the M^+ scale tended not to exhibit the verbal passive–aggressive qualities reflected in the F_{VA}^- scale or the self-subordinating qualities of the F_C^- scale.

Both positive and negative scales contribute to measures of self-esteem and social competence, neuroticism, and acting out behavior, although the patterns of relationships differ in the three types of measures. For self-esteem, all three positive scales, but particularly M^+, were positively related, and both F^- scales were negatively related to the criterion measure in both sexes. The correlation with M^-, however, was essentially zero. Although neuroticism tended to be associated with low self-esteem, F^+ was not significantly related to this type of emotional disturbance, although, as anticipated, substantial negative correlations were found between neuroticism and M^+ and $M–F^+$. All three negative scales were positively related to neuroticism, the highest correlation in both sexes being with F_{VA}^-.

The associations between acting out behaviors and the EPAQ scales tended to be somewhat weaker but of sufficient predictive value to be of potential practical as well as theoretical value. In both sexes, a high degree of acting out was most strongly associated with a high degree of M^- and, secondarily, with a high degree of F_{VA}^-. In both sexes, other scales also made significant contributions, but the pattern was not the same. Whether this discrepancy was due to chance or to genuine sex differences cannot be determined without testing additional samples.

A comment should be made about the relatively high percentage of men and women exhibiting the high M^- and F_{VA}^- constellation even in the low acting out groups (27% of the males and 35% of the females). This finding was to be anticipated. Manifestation of these undesirable masculine and feminine characteristics can take many forms less obviously antisocial than the behaviors tapped by the acting out measure. Additional information

about the individual and his or her life circumstances is required if their avenues of expression are to be predicted. Further, these negatively sanctioned behaviors are relatively infrequent. While our data suggest that individuals who exhibit sociopathic, acting out behaviors have a substantial probability of being high in M^- and F_{VA}^-, the probability that those who are high in M^- and F_{VA}^- also exhibit acting out behaviors should differ little from those with other personality constellations.

Although the results obtained with the neuroticism and acting out measures are promising, the samples were drawn from a nonclinical population, few of whose members exhibited any significant degree of problem behavior. Acting out was particularly mild, especially in women, and some of the behaviors (e.g., excessive use of alcohol and other drugs) may in some cases reflect escape from emotional distress rather than the manifestation of hostile, antisocial tendencies. We are currently extending our investigation of the implications of the positive and negative components of agency and communion to include clients of a student counseling center and adolescent and adult offenders whose activities have brought to them the attention of legal authorities. Assuming the replicability of the findings reported here, comparisons of these individuals with those from unselected populations might be expected to yield even more marked relationships with the EPAQ scales.

Reference Notes

1. Gilbert, L., Strahan, R. F., & Deutsch, C. Personal communication, 1978.
2. Heilbrun, K. *The derivation of negative sex stereotypes from the Adjective Check List.* Unpublished manuscript, 1976. (Available from K. Heilbrun, Department of Psychology, University of Texas at Austin, Austin, Texas 78712.)

References

Bakan, D. *The duality of human existence.* Chicago: Rand McNally, 1966.

Bem, S. L. The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology,* 1974, *42,* 155–162.

Bem, S. L. On the utility of alternate procedures for assessing psychological androgyny. *Journal of Consulting and Clinical Psychology,* 1977, *45,* 196–205.

Block, J. H. Conceptions of sex roles: Some cross-cultural and longitudinal perspectives. *American Psychologist.* 1973, *28,* 512–526.

Carlson, R. Sex differences in ego functioning. *Journal of Consulting and Clinical Psychology,* 1971, *37,* 267–277.

Constantinople, A. Masculinity–femininity: An exception to the famous dictum? *Psychological Bulletin,* 1973, *80,* 389–407.

Helmreich, R. L., Spence, J. T., & Wilhelm, J. A. A psychometric analysis of the PAQ. *Sex Roles,* in press.

Helmreich, R., & Stapp, J. Short forms of the Texas Social Behavior Inventory, an objective of self-esteem. *Bulletin of the Psychonomic Society,* 1974, *4,* 473–475.

Kelly, J. A., Caudill, S., Hathorn, S., & O'Brien, C. G. Socially undesirable sex-correlated characteristics: Implications for androgyny and adjustment. *Journal of Consulting and Clinical Psychology,* 1977, *45,* 1185–1186.

Parsons, T., & Bales, R. F. *Family socialization and interaction process.* Glencoe, Ill.: Free Press, 1955.

Rosenkrantz, P. S., Vogel, S. R., Bee, H., Broverman, I. K., & Broverman, D. M. Sex role stereotypes and self-concepts in college students. *Journal of Consulting and Clinical Psychology,* 1968, *32,* 287–295.

Spence, J. T., & Helmreich, R. L. *Masculinity and femininity: Their psychological dimensions, correlates and antecedents.* Austin: University of Texas Press, 1978.

Spence, J. T., Helmreich, R., & Stapp, J. The Personal Attributes Questionnaire: A measure of sex role stereotypes and masculinity–femininity. JSAS *Catalog of Selected Documents in Psychology,* 1974, *4,* 43. (Ms. No. 617)

Spence, J. T., Helmreich, R., & Stapp, J. Ratings of self and peers on sex role attributes and their relation to self-esteem and conceptions of masculinity and femininity. *Journal of Personality and Social Psychology,* 1975, *32,* 29–39.

Received September 25, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 10, 1683-1698

Motives for Social Comparison: The Construction–Validation Distinction

Russell H. Fazio
Indiana University

It is suggested that informational social comparison can be motivated by two different concerns: (a) an interest in obtaining from others any information about the entity that they have at their disposal and that the individual lacks or (b) an interest in obtaining information about the validity of one's judgment by determining if that judgment derives from one's personal biases or from the qualities of the entity. A comparison act motivated by the former concern is termed *construction* and by the latter is termed *validation*. The results of Experiment 1 suggest that a construction process occurs only when one's perceived level of information is low and others might possess additional information about the object. Two additional experiments examined the impact of agreement and disagreement from sources similar or dissimilar on a relevant attribute. Experiment 2 examined whether the judgments of each of these others are attributed largely to the other's similarity/dissimilarity or to the amount of information about the entity that the other is thought to possess. Experiment 3 examined the implications of the attributional patterns obtained in Experiment 2 for choice of comparison other under motives to construct vs. validate. The two motives were found to lead to differing preferences regarding choice of comparison other.

As individuals attempting to understand, predict, and exercise some control over a rather complex social environment, we are often interested in how our judgment of some object compares with others' judgments of that object. Such a proposition suggests that social comparison is desired and useful when one is in need of information about the adequacy of one's judgment (cf. Festinger, 1954; Goethals & Darley, 1977; Jones & Regan, 1974). The more important the judgment, the more likely that such informational social comparison will occur.

The present research focuses upon informational social comparison and investigates the proposal that such comparison can be motivated by two quite different concerns.

1. Sometimes we are interested in obtaining from others any information about the object that they have at their disposal and we lack. The goal of this process, which shall be termed *construction,* is simply the collection of information about the object—information that may be useful in constructing a judgment of the object.

2. A second concern that can motivate informational social comparison is an interest in testing the validity of one's judgment—what shall be termed a *validation* process. The individual wants to know whether she/he has processed and integrated the available information properly. Would others, given the same information, reach the same judgment? Or has she/he been unduly influenced by some personal idiosyncrasy? Naturally, the individual hopes that the answers to these questions lead to the conclusion that she/he is correct, that is, she/he hopes to achieve confirmation.

This article is based upon a doctoral dissertation submitted to the Department of Psychology of Princeton University. The author thanks Joel Cooper, chairman of the dissertation committee, for his thoughtful supervision of this research project, and the other members of the committee, John Darley and Ron Kinchla, for their helpful suggestions.

Requests for reprints should be sent to Russell H. Fazio, Department of Psychology, Indiana University, Bloomington, Indiana 47401.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3710-1683$00.75

The construction–validation distinction is based upon the sort of information that the individual is seeking. Construction involves information about the object; validation concerns information about one's judgment of that object. (The distinction is similar to one drawn by Mettee & Smith, 1977, who note a difference between the discovery and construction of reality, i.e., construction, and the confirmation of already existing reality concepts, i.e., validation.) An analogy to two types of research may serve to clarify the distinction further. Often, research is conducted for exploratory purposes and in order to generate hypotheses. A few intuitions may guide such research, but the main purpose is not to test those intuitions but to collect data that can help one to form hypotheses. This type of investigation is analogous to the construction process. The validation process corresponds to research in which an empirical test of a hypothesis is conducted. The researcher seeks to test the validity of his or her hypothesis.

Just as a researcher might first conduct an exploratory investigation and then an empirical test of a hypothesis, an individual might first be concerned with construction and then with validation. That is, the two processes are likely to occur in sequence. A person may first desire information about the entity so that he or she can construct a tentative judgment of the entity. He or she may then seek to test the newly formed hypothesis empirically via a validation process. If validation is achieved, then the informational social comparison process will have come to a successful conclusion. If it is not achieved, the individual is likely to decide that she/he is in need of more information about the entity. The person may return to a construction process in order to discover what "facts" he or she has yet to consider. The person may evaluate this newly acquired information, construct a more accurate opinion, and then seek to validate that opinion.

It is our contention that the construction–validation distinction is critical to an understanding of informational social comparison. It shall be argued that construction and validation processes are very different, and three experiments suggesting this will be reported. Not only do concerns with construction versus validation diverge as a function of relevant variables, but choice of comparison other differs, given motives to construct vs. validate.

Experiment 1

Experiment 1 examined what variables might determine the extent to which an individual is concerned with construction versus validation. It was hypothesized that an individual would be most interested in construction and least interested in validation when the individual believed that she/he lacked information and believed that the other people available for comparison might possess a greater amount of information.

In order to examine social comparison processes, an experimental task of sufficient difficulty to prompt a need for information was required. Furthermore, the task needed to be one that permitted a simple and plausible manipulation of perceived level of information. For these reasons, the autokinetic effect was chosen as the experimental task. From the work of Sherif and his associates (e.g., MacNeil & Sherif, 1976; Sherif, 1935), it is clear that estimating the distance that the dot of light "moves" is a difficult and ambiguous task. In addition, perceived level of information could be easily manipulated by leading subjects to believe that they had been assigned to a condition in which they would or would not be provided with additional information about the distance that the light moves. Half of the subjects were provided with bogus information about the light's movement, whereas the other subjects were not.

Although subjects participated in the experiment individually, they were each led to believe that three other subjects were also present. The amount of information about the light that these supposed others might possess was manipulated as the second factor in the design. Half of the subjects were told that they and each of the supposed other subjects would be *individually* assigned to an information condition. The other half were told that each *group* would be assigned to a condition.

If the possibility that the available others possess more information does exist, as in the individual assignment conditions, then a construction motive may become operable. The lower a person's perceived level of information, the more that person may attempt to gather information from the available others (i.e., to construct). A subject in the group assignment conditions was led to believe that everyone possessed the same level of information that he or she had. This equivalence is predicted to undermine the construction motive, since it is not possible to collect additional information about the light. However, assignment to this condition should lead to a corresponding increase in a desire to validate one's judgment. Given whatever information upon which the subject has based his/her judgment, can she/he find confirmation of that judgment? Thus, desires for construction versus validation are predicted to be jointly dependent upon perceived level of information and individual versus group assignment. Subjects in the no information – individual assignment condition should exhibit construction versus validation interests that differ from the interests displayed in each of the other conditions.

Once the two independent variables, perceived level of information and assignment, had been manipulated, observations of the autokinetic effect began. During one particular series of trials, the subject was informed of the judgments of each of the supposed others. These bogus responses were constructed so as to manipulate agreement. One of the supposed others was shown to consistently agree with the actual subject, whereas the other two consistently disagreed. For another series, subjects were ostensibly to be formed into pairs, and the members of a pair were to discuss their judgments prior to reaching a decision about the distance that the light moved. Ratings of the extent to which each of the supposed others was desired as a partner served as the main dependent variable.

As argued earlier, subjects in the group assignment conditions have no opportunity to gain information for the construction process. However, they may well have a desire to validate their judgments. Validation implies choice of the agreeing other as a partner; she/he provides confirmation of one's own judgment. Thus, within the two group assignment conditions, regardless of the subject's perceived level of information, the agreer should be strongly preferred as a partner.

The perceived level of information variable should, however, have an impact upon partner selection within the individual assignment conditions. We have argued that construction tendencies should be most apparent for subjects with no information. Hence, these subjects will presumably desire to have as a partner someone who has more information about the light's movement. The agreeing other may be agreeing simply because she/he, too, is suffering from a lack of information. The disagreers, however, may be disagreeing because they have more information at their disposal than the subject has. As a result of these possibilities, preference for the agreer as a partner should be relatively low within this no information – individual assignment condition. For subjects with additional information, the need to gather more information (i.e., the desire to construct) should be weaker, and the desire to validate stronger, than for those subjects with no information. Hence, preference for the agreer as a partner should be stronger among subjects with, than among subjects without, additional information.

Method

Subjects

Forty Princeton University freshmen, naive to the true nature of the autokinetic effect, participated individually in the experiment for a payment of $1.50. Subjects were randomly assigned to one of the four conditions. The data from 2 subjects—1 from each of the group assignment conditions—were discarded. In one case the subject questioned the veridicality of the supposed other subjects' judgments. In the other, the subject did not perceive sufficient movement to permit the creation of one of the disagreeing responses. The final sample consisted of 10 subjects in each of the individual assignment conditions and 9 in each of the group assignment conditions.

Procedure

After making a few introductory comments, the experimenter informed the subject that a total of

four individuals would be participating in the experimental session. (This statement was the first of a number of steps that were taken to establish the plausibility of the presence of the fictitious group members.) Depending upon the punctuality of the subject's own arrival, the experimenter also commented upon how many of the (fictitious) others had already arrived. The subject was then ushered into an enclosed booth and instructed to read the materials that had been placed there earlier. While the subject read, the experimenter opened and closed doors at the appropriate times so as to lead the subject to believe that all three of the other subjects had arrived.

The materials that had been placed in the subject's booth described the study's supposed purpose and experimental design. The study purported to concern "the perceptual judgment of motion and distance" and was said to consist of three series of trials, "simulating three different ways in which judgments are often made." During Series 1, each individual was to make independent judgments about the distance the light moved. During Series 2, individual judgments were again to be made, but at the end of each trial the subject was to be informed of each of the other subject's responses. The third series was labeled a "coalition series" during which pairs of subjects would discuss their judgments. This conversation was supposedly to occur via a set of earphones and a microphone located in the booth. Subjects were told, "as much as possible, we will allow you to choose your partner."

Subjects also read that in order to investigate "how varying levels of information affect the accuracy of people's judgments," each subject would be assigned to one of three information levels for the duration of the experiment. On each trial, the experimenter would purportedly select a point within a distribution of distances specially prepared for that trial and would set a dial that would dictate the distance that the light moved. The information levels were said to relate to the distributions that the experimenter was ostensibly employing. The three conditions were described as: (a) a no information condition, in which subjects were to be provided with no additional information about the light's movement; (b) a partial information condition, in which subjects were to be provided with the interquartile range of the distribution of distances across all trials in the experiment; and (c) a complete information condition, in which subjects were to be given the interquartile range of the distributions for each and every trial of the experiment. Interquartile range was defined for the subject.

Assignment manipulation. One of two paragraphs followed the description of the information conditions. Half the subjects, those who had been randomly assigned to the individual assignment conditions, read the following paragraph.

> Each of you will be placed in one of these three conditions on a random basis. Thus, you may personally have at your disposal more or less information than the other people in the experiment.

For subjects in the group assignment conditions, the paragraph read:

> Each group of subjects in the experiment is randomly placed in one of these three conditions. Thus, whereas each of you in this experimental group will have the same amount of information, you and the other subjects in this group may have at your disposal more or less information than other groups in the experiment.

Once sufficient time had elapsed for the subject to read the materials and once the experimenter had feigned the arrival of each of the supposed other subjects, the experimenter began a series of private interviews with the actual subject and each of the supposed others. These interviews were conducted in the order of the subjects' purported arrival. To maintain and bolster the impression that three others were present, the experimenter opened the booth door for a supposed other. He then loudly requested that the person follow him, noisily moved the chair in the booth about, walked to the nearby interview room, and slammed the door shut behind himself. After allowing about a minute to pass, the experimenter noisily returned the supposed subject to the booth and closed the booth door. This procedure was repeated until all the supposed subjects had been "interviewed."

Perceived level of information manipulation. During the interview of the actual subject, the experimenter answered any questions that the subject had. He also made certain that the subject understood the descriptions of the three series and the three information conditions. The remainder of the interview concerned the information manipulation. The experimenter handed the subject a sheet of paper describing the condition to which the subject had been assigned. For half the subjects, this sheet simply read, "You have been randomly assigned to the NO INFORMATION condition." For the remaining subjects, the sheet stated, "You have been randomly assigned to the PARTIAL INFORMATION condition" and informed the subjects that the interquartile range of the distances the light would move was 2 to 6 inches. From pretesting, these figures had been determined to be within the range of distances subjects would perceive in the experimental setting. No subject was assigned to a complete information condition. Subjects were only led to believe that such a condition existed to ensure that the partial information subjects did not feel so informed that they would have no need to engage in social comparison.

Once the interviews had been completed, observations of the autokinetic effect began. The far wall of each booth contained a one-way mirror through which the subject observed the light in the next room. The light was located approximately 10 feet from the actual subject's mirror. For each trial the

light was illuminated for a period of 10 seconds. After each of the three trials in Series 1, the subject recorded a judgment of the distance to the nearest half-inch.[1] After each trial of Series 2, the subject also recorded a judgment. In addition, the subject wrote the judgment in large numerals on a blank sheet of paper and then held the sheet up to the mirror so that the experimenter could record each response. The experimenter walked from mirror to mirror and supposedly recorded the judgments of each of the subjects. Purportedly, he then made four copies and distributed one copy to each subject. Again, in order to maintain and bolster the deception that three others were involved, the experimenter made it appear that he was collecting a judgment from the occupant of each booth and that he was handing each subject a copy of the responses. This same procedure was followed for each of the three trials of Series 2.

Agreement manipulation. The supposed responses of the other subjects were fabricated in such a manner that one other subject generally appeared to agree with the subject and the two others generally disagreed. One disagreer (the positive disagreer) consistently perceived more movement than the subject; the other (the negative disagreer) reported observing less movement. On the first trial, the agreer was reported to have perceived 1/2 inch more movement than the subject, and the disagreers 2 inches more or less movement. On Trial 2, the agreer's response was 1/2 inch less than the subject's and the disagreers' responses were 1 inch more or less than the subject's. For Trial 3, the agreer's judgment was 1/2 inch more and the disagreers' 1 and 1/2 inches more or less than the subject's own response. All subjects were identified only by the booth identification letter (A, B, C, or D). The actual subject was always Subject B. For each subject, the location of the agreer, positive disagreer, and negative disagreer was randomly determined.

Dependent variables. After receiving the feedback for Trial 3 of Series 2, subjects were asked to complete a "Partner Selection Form" for Series 3. As a forced-choice measure, subjects were requested to select one other subject as a partner. In addition, in order to "accommodate all four of you as well as possible" and "in order to settle any conflict in partner selection that might occur," subjects were asked to rate the degree to which they would like to have each of the other subjects as a partner. These ratings were made on an 11-point scale where 0 indicated "definitely do not" prefer to have a particular subject as a partner and 10 meant "very much" prefer to have that person as a partner.

While the experimenter ostensibly reviewed the partner selection forms, subjects were asked to complete a short questionnaire. Two scales involved assessments of the degree to which construction and validation concerns were related to the subject's choice of a partner.[2] The first item asked subjects, "To what extent did you choose the partner you chose because he/she is likely to provide you with helpful information about the distance the light moves?" The second item read, "To what extent did you choose the partner you chose because he/she is likely to confirm and support your own judgments about the distance the light moves?" Each of these ratings was made on a 0–10 scale where 0 was labeled "not at all a consideration in my choice" and 10 "very much a consideration."

The last page of the questionnaire contained a check on the agreement manipulation. Subjects were asked to rate "how similar the responses of each of the other subjects were to your own responses during Series 2?" on a 0–10 scale where 0 was labeled "not at all similar" and 10 was "very similar". It should be noted that these agreement assessments were made while the subjects still had at their disposal the sheets that displayed each of the supposed subjects' judgments. Thus, this manipulation check measures only whether subjects could understand and correctly interpret the feedback when directly asked to do so. It in no way served as a memory check.

When the subject had completed the questionnaire, the experiment was terminated. The subject was carefully probed for suspicion with the aid of a postexperimental questionnaire. All deceptions were revealed, and the purpose of the study was explained to the subject.

Results

Manipulation Check

Responses[3] to the check on the manipulation of agreement suggest that the manipu-

---

[1] After each trial, the subject also indicated how certain he or she was of his or her judgment of the distance that the light moved. For each subject, the mean of the subject's three ratings during Series 1 (i.e., prior to any feedback) was computed. On a scale where 10 indicated extreme certainty and 0 reflected being not at all certain, the overall mean across all subjects was 4.25, suggesting that subjects were sufficiently uncertain to desire social comparison. An analysis of variance on these data revealed no significant differences between the conditions, however. Possibly, a floor effect was encountered; no information subjects may have been reluctant to admit any greater uncertainty.

[2] These scales were preceded by an open-ended measure in which subjects were asked to explain the reasons for their partner choices. The data from two judges who rated the extent to which each response indicated a concern with construction versus validation display the same pattern as the direct scale assessments. The same is true in Experiment 3. Details can be found in Fazio (1978).

[3] Two subjects, one from each of the individual assignment conditions, neglected to complete the manipulation check items.

Table 1
Mean Partner Preferences in Experiment 1

Perceived level of information and assignment	Agreer	Disagreer	
		Positive	Negative
No			
Individual	5.70	6.30	4.40
Group	8.11	4.22	4.22
Partial			
Individual	7.80	5.70	4.70
Group	8.11	6.56	2.67

Note. The higher the mean, the more that supposed other was desired as a partner.

lation was accomplished successfully. A 2 (No vs. Partial Information) × 2 (Group vs. Individual Assignment) × 3 (Agreer vs. Positive Disagreer vs. Negative Disagreer) analysis of variance, with the last factor as a repeated measure, revealed a strong effect-of-the-agreement variable, $F(2, 64) = 159.12$, $p < .001$. Furthermore, a contrast of the mean rating of the agreer ($M = 6.64$) with the average of the ratings of the two disagreers ($M = 4.21$) was highly significant, $t(64) = 17.72$, $p < .001$.

Partner Selection

Table 1 presents the mean preferences for each of the three others as a partner. A 2 × 2 × 3 analysis of variance revealed a main effect of the agreement variable (agreer $M = 7.43$, positive disagreer $M = 5.69$, negative disagreer $M = 4.00$), $F(2, 68) = 28.54$, $p < .001$. This general preference for the agreer is more characteristic, however, of the subjects in the group assignment condition than of those in the individual assignment condition, as is evidenced by a significant interaction between the assignment and agreement factors, $F(2, 68) = 4.12$, $p < .025$. As predicted, however, this two-way interaction is qualified by a significant higher order perceived level of information by assignment by agreement interaction, $F(2, 68) = 4.88$, $p < .02$. A preference for the agreer over the disagreers was apparent in each condition except the no information – individual assignment condition. In accordance with the conceptual framework proposed earlier, when others possibly had more relevant information and one's own perceived level of information was low, the agreeing other was not as strongly desired as a partner. Further confirming the hypotheses is the fact that multiple comparisons (by means of the least significant difference test) reveal that the agreer was preferred significantly less in the no information – individual assignment condition than in each of the other three conditions.

The forced-choice data display much the same pattern as the rating data. The proportion of subjects in each condition who chose the agreer as a partner (no information/individual = .50, partial information/individual = .90, no information/group = 1.00, partial information/group = .89) was noted, and these proportions were analyzed via an arcsin transformation (cf. Langer & Abelson, 1972). The analysis revealed a significant interaction between the perceived level of information and assignment variables, $z = 2.47$, $p < .02$. Subjects in the no information – individual assignment condition chose the agreer as a partner relatively less frequently than did the other subjects.

Motive for Comparison

Table 2 presents the mean ratings of desire for construction and validation in each condition. The data were analyzed as a 2

Table 2
Mean Construction and Validation Scores in Experiment 1

Perceived level of information and assignment	Construction	Validation
No		
Individual	7.00	3.80
Group	3.89	8.33
Partial		
Individual	5.20	6.00
Group	5.22	6.11

Note. The higher the mean, the more construction or validation was desired.

(Perceived Level of Information) × 2 (Assignment) × 2 (Construction/Validation) analysis of variance, with the last factor as a repeated measure. The analysis revealed an interaction between the assignment and construction/validation factors, $F(1, 34) = 6.82$, $p < .02$. More importantly, the predicted three-way interaction was also statistically significant, $F(1, 34) = 6.41$, $p < .02$, indicating that the reported motive for comparison was jointly dependent upon one's perceived level of information and the possibility that others possessed more information.

It should be noted that comparisons between construction and validation means within any given condition do not achieve statistical significance, with the exception of the no information – group assignment condition. Given that it is impossible to determine in advance whether the no information situation is sufficently lacking in information about the light's movement to produce a "pure" desire for construction, the present framework can only predict that the perceived level of information variable will affect desires to construct versus validate in a relative sense. The predictions call for differential construction versus validation interests when the no information – individual assignment cell is compared to each of the other cells. A series of interaction contrasts suggests that this is the case: for the comparison with the no information – group assignment condition, $t(34) = 3.65$, $p < .001$; with the partial information – individual assignment condition, $t(34) = 1.96$, $p < .10$; with the partial information – group assignment condition, $t(34) = 1.95$, $p < .10$.[4] Given that these measures of motive for comparison occurred after partner selection, these subsidiary data should be interpreted cautiously, however.

Discussion

The results of the experiment provide some support for the conceptual framework that was outlined earlier. Only when perceived level of information was low *and* the available others might possess more information were construction concerns relatively more apparent than validation concerns. Under these conditions, it appears that it may be advantageous to discover *why* another person disagrees. The disagreeing other may possess some useful information about the object. In contrast, individual assignment subjects whose perceived level of information was relatively high expressed little or no interest in discovering the reasons for another's disagreement. Rather, these subjects reported a relatively strong desire to confirm their own judgments and displayed a greater preference for the agreer than did no information subjects.

As the numerous interaction effects between the assignment and information variables show, the effects of perceived level of information differ under group versus individual assignment conditions. When the available others definitely possessed no more or less information than the subjects did, construction motives were thwarted. As a result, the interactive effect that was observed within the individual assignment conditions (no information subjects desiring construction more and validation less than did partial information subjects) is not apparent within the group assignment conditions. If anything, the data in the group assignment conditions tend in the reverse direction, with no information subjects desiring validation more and construction less than did partial information subjects, $t(34) = 1.65$, $p < .11$.

Regardless of the reported motives, subjects in both group assignment conditions displayed a strong preference for the agreer as a partner. One possible reason for this preference is a relatively greater concern for validation. Choosing an agreer, however, could reflect many possible concerns, including either validation or self-esteem enhancement. In fact, a number of researchers have drawn a distinction between informational and self-esteem motives for social comparison (e.g., Gruder, 1977; Hakmiller, 1966; Latane, 1966; Singer, 1966; Thornton & Arrowood, 1966). Experiments 2 and 3 more

[4] Each of these interaction contrasts reached a conventional level of statistical significance on the data from the open-ended measures of construction and validation.

closely address this issue of validation and its distinction from self-esteem processes.

Experiment 2

Goethals & Darley (1977) have suggested that two variables—agreement and similarity—may play a critical role in the selection of a comparison other. They invoked various attributional principles (Kelley, 1967, 1973) to note that a person seeking to determine whether his or her belief is correct wishes to determine whether his or her opinion is entity based (whether it accurately describes the entity) or whether it is person based (whether it derives from his own biases). Someone similar on attributes affecting the belief may agree simply because she/he is similar and is subject to the same biases that the individual is prone to. Correspondingly, disagreement from a dissimilar other can be attributed to the person's dissimilarity on the crucially relevant attribute.

Disagreement from a similar source and agreement from a dissimilar source appear to be more meaningful. If a similar other disagrees, the person's confidence that his opinion is entity based must decrease. Since the other is similar on the relevant factors, it would be difficult to attribute the discrepancy to the internal characteristics of that other. A dissimilar other who agrees, on the other hand, provides strong confirmation that one's belief is an accurate evaluation of the entity (cf. Goethals, 1972; Goethals & Nelson, 1973). Here is a person with wholly different biases who arrives at the same conclusion. The judgment which receives confirmation from a dissimilar other must, then, be due to the compelling qualities of the entity in question.

Experiment 2 examined whether attributions of this sort do actually occur. Subjects were asked to make causal attributions as to why each of four others (a similar and a dissimilar agreer and a similar and a dissimilar disagreer) agreed or disagreed with him. The predictions were: (a) the similarity/dissimilarity dimension would be assigned a greater causal role in determining the judgments of the similar agreer and the dissimilar disagreer than in determining the judgments of the similar disagreer and the dissimilar agreer and (b) the amount of information possessed would be assigned a greater causal role with respect to the judgments of the similar disagreer and the dissimilar agreer than with respect to the judgments of the similar agreer and the dissimilar disagreer.

Method

Subjects

Eighteen Princeton freshmen participated for a payment of $3.00. In the experiment, subjects were randomly assigned to the no or partial information conditions. One subject's data were discarded because the subject failed to understand how to complete the scales, leaving 9 subjects in the no information condition and 8 in the partial information condition.

Procedure

In a first session, all subjects were administered two paper and pencil tests. The first three-page, 18-item test was presented as a measure of one's depth perception ability. The test actually consisted of gestalt completion problems. Each item presented an incompletely and vaguely drawn picture. The subject's task was to identify the picture and to label it. The second test, presented as a measure of lateral perception, was also three pages and 18 items long. Each item presented an assortment of geometric figures. Subjects had to select which of five possible answers accurately displayed the smaller target figures in an arrangement that formed one larger figure.

Similarity manipulation. In order to provide subjects with feedback that was likely to confirm their own perceptions as to how well they had performed on the two tests, the number of items that the subject completed in the allotted time was counted for each test. These two numbers were placed on each of four graphs of page size, where two points of the abscissa were labeled *depth* and *lateral perception,* with the ordinate (8 inches in length) simply labeled *score.* A line connecting the two points represented the subject's profile. In addition, each graph presented the supposed scores of one other person. This other was identified by a letter that was to correspond to that person's booth identification letter during the second session. One similar other was shown to have a profile that was approximately ½ inch higher than the subject's on the depth perception test and ½ inch lower on the lateral perception test. Another profile showed the remaining similar other to be ½ inch lower than the subject on depth perception and ½ inch higher on lateral perception. The dissimilar others were portrayed by profiles with points 2 inches higher

or lower than the subject's. One dissimilar other was shown to have scored much better than the subject on depth perception and much lower on lateral perception, whereas the other was portrayed as having the reverse pattern. In this way, all the subjects were displayed as having approximately the same overall ability and differing on the specific components.

The procedure employed in the second session was virtually identical to that in the individual assignment conditions[5] of Experiment 1 and need not be repeated in detail. After reading the materials in his or her booth, the subject was ushered into the next room for the interview with the experimenter. During this time, the subject's questions were answered and the subject was notified of his or her information condition. The experimenter then handed the subject the test profiles and described how the graphs could be read. The experimenter further remarked that, whereas the perception of motion and distance involved both lateral and depth perception cues, people approach the perception of motion and distance with different styles. In this way, the profiles were presented as reflecting a dimension that was related to the task at hand (cf. Goethals & Darley, 1977; Zanna, Goethals, & Hill, 1975).

Following "interviews" with each of the supposed subjects, the autokinetic effect trials began. After each trial of Series 2, the subject received the bogus judgments of each of the supposed other subjects. One similar and one dissimilar other were portrayed as generally agreeing with the subject and the remaining similar and dissimilar others as generally disagreeing. These bogus judgments were presented via the same scheme used in Experiment 1. For half the subjects, the similar and the dissimilar others who had scored higher than the subject on the depth perception task were shown to agree with the subject, and the two others who had scored lower, to disagree. For the other half of the subjects, this pattern was reversed. Furthermore, for half the subjects, the similar agreer and the dissimilar disagreer were displayed as having seen more movement than the subject whereas the similar disagreer and the dissimilar agreer perceived less. For the other half of the subjects, the pattern was reversed such that the latter two individuals perceived more movement, and the former two less. Together, then, there were four possible arrangements of the others' responses. Approximately, one-quarter of the subjects in each condition received each of the possibilities. This "arrangement variable" produced no main or interactive effects on any of the variables assessed and accordingly receives no further consideration in this report.

Following Series 2, subjects were informed that they would be completing a questionnaire in which they would assess the reasons for the other subjects' judgments during Series 2. Subjects were asked to assess the importance of each of three dimensions in determining other subjects' judgments. The dimensions were the other's perceptual style, the amount of information the other possessed, and circumstances. Following Stevens and Jones (1976), the assessments were made on two sets of scales. The first set explicitly forced upon the subject the interdependence that is presumed to exist among the three dimensions. The subject was instructed to allocate a total of 100 points among the three possible reasons for the other subject's agreement or disagreement with his or her own judgments. A dimension was to be given a point value that reflected its judgments. Under the first listed dimension, "the subject's perceptual style," two options appeared: (a) "the similarity of the subject's perceptual style to my own" and (b) the "dissimilarity of the subject's perceptual style to my own." The subject was to choose (a) or (b), whichever he or she considered the most appropriate, and then assign the dimension a point value. Under the second listed dimension, "the amount of information the subject had about the light's movement," two options were also listed: (a) "the subject's lack of information" and (b) "the subject's possession of information." Again, the subject was to select the more appropriate option and assign the factor a point value. The third dimension listed was "the particular circumstances of the subject's observations" (e.g., the angle at which the subject viewed the light, his or her mood and attentiveness, etc.).

The second set of scales did not force any interdependence upon the subject. Instead, the subject rated the importance of the particular dimension on an independent scale. As the two sets of scales revealed the same effects, only the data from the interdependent scales shall be presented. The interested reader is referred to Fazio (1978) for a detailed presentation of the data from the set of independent scales.

Following completion of the scales, subjects were carefully probed for suspicion and were thoroughly debriefed.

Results

Subjects allocated a total of 100 points to the three attribution dimensions of perceptual style, information, and circumstances. Each dimension was assigned a point value that reflected its importance in determining the other person's judgments. Table 3 presents the means for each dimension. Because the value of one rating is perfectly determined by the sum of the other two ratings, only the perceptual style and information

---

[5] It was decided to run only the individual assignment conditions because perceived level of information was found to exert an influence in Experiment 1 only within the individual assignment cells.

Table 3
Means on Each of the Attribution Scales in Experiment 2

Attributional dimension and perceived level of information	Similar		Dissimilar	
	Agreer	Dis-agreer	Agreer	Dis-agreer
Perceptual style				
No	67.78	31.67	31.11	61.11
Partial	61.25	43.13	43.75	55.63
Information				
No	15.00	35.00	47.22	18.33
Partial	21.25	37.50	38.75	23.13
Circumstance				
No	17.22	32.22	21.67	20.56
Partial	17.50	20.63	17.50	21.25

Note. The higher the mean, the more importance assigned to that dimension.

dimensions were analyzed. A 2 (No vs. Partial Information) × 2 (Similar vs. Dissimilar) × 2 (Agreer vs. Disagreer) multivariate analysis of variance revealed a strong Similarity × Agreement interaction, $F(2, 14) = 8.16$, $p < .005$. Perceptual style was viewed as more important in determining the judgments of those others whose responses were congruent with their perceptual style (the similar agreer and the dissimilar disagreer) than in determining the judgments of the incongruent others (the similar disagreer and the dissimilar agreer), univariate interaction $F(1, 15) = 13.60$, $p < .005$. Least significant difference tests revealed that similar agreers ($M = 64.52$) were perceived as having been affected by their perceptual style significantly more than were similar disagreers ($M = 37.40$). Likewise, dissimilar disagreers ($M = 58.37$) were thought to have been affected by perceptual style significantly more than were dissimilar agreers ($M = 37.43$). A univariate analysis of the information data revealed a Similarity × Agreement interaction such that the amount of information possessed was assigned a greater causal role in determining the judgments of the incongruent others than in determining the judgments of the congruent others, $F(1, 15) = 16.99$, $p < .001$. Similar disagreers ($M = 37.25$) were thought to have been affected by the information dimension significantly more than were similar agreers ($M = 18.13$), and dissimilar agreers ($M = 42.99$) were perceived as having been affected significantly more than were dissimilar disagreers ($M = 20.73$). No multivariate or univariate interactions with perceived level of information approached statistical significance.[6]

The results confirm Goethals & Darley's (1977) attributional analysis. Relative to similar disagreers and dissimilar agreers, similar agreers and dissimilar disagreers were perceived as having been influenced by perceptual style. The judgments of similar disagreers and dissimilar agreers, on the other hand, are thought to emanate more from their possession of information than are judgments of the two remaining others. Furthermore, at least in the present setting, these attributional patterns appear to hold regardless of one's perceived level of information and one's motive for social comparison.

Experiment 3

Experiment 3 examined the implications of the attributional patterns found in Experiment 2 for actual choice of comparison other. From an attributional perspective, only a similar disagreer and a dissimilar agreer are potentially meaningful and informative comparison choices. It was predicted that these two sources differ in the degree to which they appeal to a person concerned with construction versus a person concerned with validation.

The person seeking construction perceives himself to suffer from a lack of information with which to "construct reality." Experiment 1 suggested that those seeking construc-

[6] These data were also analyzed in terms of scores coded for the direction of the various attributions. The analysis showed that the judgments of similar agreers and dissimilar disagreers were attributed largely to similar and dissimilar perceptual styles, respectively. The disagreement of a similar other and the agreement of a dissimilar other were attributed largely to the others' possession of information. No significant effects were observed on the circumstances scale. The analysis is reported in detail in Fazio (1978).

tion are more interested in discovering why some other person disagrees with them than are those people seeking validation. These differential preferences should be particularly evident when it is clear that this disagreeing other is similar on relevant attributes. Given this similarity, this other person should be agreeing, unless he is aware of some information that the subject himself lacks. To the extent that the partner preferences of no information subjects in the present experimental situation are dictated solely by construction concerns, the similar disagreer should be more strongly preferred than any of the available others. At a minimum, no information subjects should display a greater preference for the similar disagreer than partial information subjects, since the former group should be relatively more interested in construction and less interested in validation than the latter group.

Analogously, it can be hypothesized that those seeking validation should display a greater preference for the dissimilar agreer. Such a person, having what he considers sufficient information to "construct reality," is interested in determining the validity of his judgment. The dissimilar agreer is particularly useful in this regard. Agreement from a dissimilar source informs the individual that he has based his judgment upon the "facts" and not upon any person biases. Thus, to the extent that partial information subjects are concerned solely with validation, they should display a stronger preference for the dissimilar agreer than for any of the three others. Given their relatively greater interest in validation over construction, partial information subjects should, at a minimum, prefer the dissimilar agreer more than no information subjects.

Method

Subjects

Twenty-four Princeton freshmen participated for a payment of $3. Subjects were randomly assigned to no or partial information conditions. The data from 2 subjects were excluded from the analysis because the subjects were skeptical of either the agreement or the similarity feedback. The final sample consisted of 11 subjects in each of the two conditions.

Procedure

With the exception of the dependent variables, the procedure was identical to that employed in Experiment 2. After the manipulations had been accomplished, subjects in Experiment 3 completed the same dependent measures as in Experiment 1. Partner selection ratings were made for Series 3. Measures concerning motivations behind partner selection were also completed.

These subjects then completed checks on the manipulations of similarity and agreement. On a 0 (very dissimilar) to 10 (very similar) scale, subjects rated "how similar the depth and lateral perception test results of the other subjects were to your own test results." Also, on a 0 (strongly disagreed) to 10 (strongly agreed) scale, subjects rated "the degree to which the responses of each of the other subjects during Series 2 agreed or disagreed with your own responses." It should again be noted that subjects were not expected to have committed all the similarity and agreement information to memory. The profiles and judgments of each of the others were still available to the subject. These measures merely served to check whether subjects could correctly interpret the information provided about the others.

Results

Manipulation Checks

Similarity and agreement. Subjects in Experiment 3 completed two measures designed to check that the similarity and agreement information was understandable to both these subjects and those in Experiment 2. A 2 (No vs. Partial Information) × 2 (Similar vs. Dissimilar) × 2 (Agreer vs. Disagreer) analysis of variance, with the last two factors as repeated measures, was conducted on each of the manipulation checks. A strong main effect of the similarity variable was found on the similarity ratings (similar $M = 8.41$, dissimilar $M = 2.23$), $F(1, 20) = 491.25$, $p < .001$. Correspondingly, the agreement manipulation strongly affected the ratings of the degree to which the others' judgments agreed with the subject's own judgments (agreer $M = 8.21$, disagreer $M = 3.41$), $F(1, 30) = 141.42$, $p < .001$. Thus, subjects were capable of drawing the desired distinctions among the supposed four others present.

Certainty. Perceived level of information has been presumed in this report to relate to certainty. The more information an individual believes himself or herself to possess about the object in question, the more certain that per-

Table 4
Mean on Each Certainty Measure in Experiments 2 and 3

Measure	Experiment	Perceived level of information: No	Perceived level of information: Partial
Difference ratio	2	.84	.68
	3	.84	.72
Symmetry	2	.57	.21
	3	.38	.26

Note. On both measures, the higher the mean, the greater the uncertainty.

son is apt to feel about his/her judgment of that object. Because of subjects' reticence to admit of extreme uncertainty, direct measures of certainty tend to be relatively unpowerful. In Experiment 1, no effects were found with a direct measure of certainty (see Footnote 1). Thus, Experiments 2 and 3 involved an indirect measure of certainty. In addition to their best estimates of the distance that the light moved, subjects were asked to indicate the maximum and minimum possible distances that the light moved.

From this information, a two-component index of certainty was derived:[7] (a) the mean (across the three trials of Series 1) of the ratios of the difference between the maximum and minimum estimates to the best estimate and (b) the proportion of the three trials on which the subject placed his or her best estimate symmetrically between the minimum and maximum judgments. For each component, higher scores reflect less certainty.

Multivariate analysis of the data was performed in order to ask the question "Is there any linear combination of the difference ratio and the symmetry measure which discriminates between the conditions?" A 2 (Experiment 2 vs. 3) × 2 (No vs. Partial Information) multivariate analysis of variance revealed a main effect of the perceived level of information variable, $F(2, 34) = 4.66$, $p < .02$, with no main effect of experiment ($F < 1$) nor an interaction effect, $F(2, 34) = 1.04$. Table 4 presents the univariate means on each of the certainty measures. Although the univariate analysis of variance of the difference ratio measure revealed no significant effects, the univariate analysis of the symmetry measure revealed a main effect of the information factor, $F(1, 35) = 7.01$, $p < .02$, no main effect of experiment ($F < 1$), and no interaction effect, $F(1, 35) = 1.99$, $p > .15$.

Although the two certainty measures appear to be reasonable, numerous other plausible measures can probably also be derived. Nevertheless, the present measures and the above multivariate analysis do provide some evidence that subjects in the no information condition may have been less certain of their judgments than were the partial information subjects.

Partner Selection

Table 5 presents the mean ratings of the extent to which each of the supposed others

[7] These two components were derived from the notion that the experimental task of indicating a best estimate and minimum and maximum estimates can be approached in two very different manners. A subject may make his or her best estimate of the distance the light moves and then decide upon a minimum and a maximum, in effect placing a confidence band around his or her best estimate (e.g., "I think it moved 3 inches, but it could be anywhere from 1 to 5 inches"). If this approach were taken, then the difference between the minimum and maximum estimates can be considered to reflect a confidence interval. Since the magnitude of the difference between maximum and minimum estimates is dependent upon the distance perceived, the ratio of the difference to the best estimate of the distance the light moved was computed. The mean ratio across the three trials of Series 1 forms the first component.

The task may have been approached in another manner, however. A subject may first establish a minimum and maximum and then set his or her best estimate somewhere within this range. If this approach were taken, the difference between maximum and minimum estimates could not be considered to reflect a confidence interval. With this approach, the logical recourse when one is not sure where within the range to locate the best estimate is to simply average the minimum and maximum estimates (e.g., "Let's see, it moved a minimum of 1 inch and a maximum of 5 inches. I'll say it moved 3 inches"). Thus, a subject might be said to be uncertain of his or her response if he/she simply chooses as the best estimate the average of the earlier established minimum and maximum distances, that is, if he/she simply sets a judgment symmetrically between the minimum and maximum. This appproach is reflected in the second component that was computed.

was desired as a partner. A 2 (No vs. Partial Information) × 2 (Similar vs. Dissimilar) × 2 (Agreer vs. Disagreer) analysis of variance, with the last two factors as repeated measures, revealed a main effect of each of the two within variables. In general, similar others ($M = 6.07$) were preferred over dissimilar others ($M = 5.27$), $F(1, 20) = 5.37$, $p < .05$, and the agreers ($M = 7.27$) were preferred over disagreers ($M = 4.07$), $F(1, 20) = 20.54$, $p < .001$. Each of these within-subjects factors interacted with perceived level of information. The general preference for a similar other over a dissimilar other was apparent in the no information condition (similar $M = 6.59$, dissimilar $M = 4.77$) but not in the partial information condition (similar $M = 5.55$, dissimilar $M = 4.77$), $F(1, 20) = 8.88$, $p < .01$. Correspondingly, the preference for an agreer over a disagreer was stronger in the partial information condition (agreer $M = 8.18$, disagreer $M = 3.14$) than in the no information condition (agreer $M = 6.36$, disagreer $M = 5.00$), $F(1, 20) = 8.76$, $p < .01$. However, the three-way interaction, which was to be expected if no information subjects had been concerned solely with construction and partial information subjects solely with validation, did not occur, $F < 1$.

Some confirmation of the hypotheses is provided by a priori t ratio comparisons of the no vs. partial information means. These comparisons rely on the notion that despite any concerns extraneous to informational social comparison that may have arisen in the experimental situation, the no information and partial information subjects should have divergent concerns with construction versus validation and hence should display differential preferences for the similar disagreer and the dissimilar agreer. As predicted, no information subjects expressed a greater preference for the similar disagreer than did partial information subjects, $t(20) = 3.57$, $p < .005$, and partial information subjects preferred the dissimilar agreer significantly more than did no information subjects, $t(20) = 3.23$, $p < .005$. The two conditions did not differ in their ratings of the similar agreer, $t(20) = 1.08$, or the dissimilar disagreer, $t(20) < 1$.

Inspection of Table 5 reveals that the data were not as was to be expected if subjects had been concerned only with informational social comparison. In each condition, the comparison other who was predicted to be most preferred, given that preferences were to be dictated only by informational motives, was not in fact most highly preferred. Although no information subjects did significantly ($p < .05$, by a least significant difference test) prefer the similar disagreer to the dissimilar disagreer, their ratings of the similar disagreer and the dissimilar agreer did not differ, and they tended ($p < .20$), if anything, to prefer the similar agreer to the similar disagreer. Partial information subjects did significantly prefer the dissimilar agreer to both the similar disagreer and the dissimilar disagreer, but they displayed equal preference for the two agreers. Thus, although the data were supportive in the sense that they suggest that desires to compare with a similar disagreer and desires to compare with a dissimilar agreer each vary as a function of perceived level of information, the similar disagreer was not the most preferred partner among the no information subjects, and the dissimilar agreer was not the most preferred other among the partial information subjects.

Table 5
Mean Partner Ratings in Experiment 3

Perceived level of information	Similar		Dissimilar	
	Agreer	Disagreer	Agreer	Disagreer
No	7.27 (6)	5.91 (3)	5.46 (1)	4.09 (1)
Partial	8.18 (7)	2.91 (0)	8.18 (4)	3.36 (0)

Note. The higher the mean, the more that person was preferred as a partner. Numbers in parentheses are the number of subjects who chose that person as a partner on the forced-choice measure.

The forced-choice data are also presented in Table 5. The data display the same general pattern as the ratings data. However, none of the No versus Partial Information comparisons are statistically significant, largely because of the overwhelming preference

among both groups of subjects for the similar agreer.

Motive for comparison. A 2 (No vs. Partial Information) × 2 (Construction vs. Validation) analysis of variance, with the last factor as a repeated measure, was performed on the ratings of desire for construction and validation. As in Experiment 1, the predicted interaction achieved statistical significance, $F(1, 20) = 4.93$, $p < .05$. In the no information condition, the construction and validation means were 7.64 and 5.27, respectively; in the partial information condition, these means were 7.00 and 8.09, respectively. Within-condition comparisons of construction means to validation means did not reach statistical significance.

Discussion

The value of similar disagreers versus dissimilar agreers appears to depend upon one's perceived level information. While these two sources are each potentially meaningful from an attributional perspective, they differ in the degree to which they appeal to an individual interested in construction vs. validation. As perceived level of information increases (and, correspondingly, as interest in construction tends to decrease and interest in validation tends to increase), preference for a similar disagreer decreases and preference for a dissimilar agreer increases. The similar disagreer is relatively more appealing to the person interested in construction than to the person interested in validation. If one is in a hypothesis-generating stage, it appears to be important to discover why a similar other, who should be agreeing, actually disagrees. The dissimilar agreer, on the other hand, is preferred relatively more by a person interested in validation than by a person interested in construction. Possessing what she/he believes to be a sufficient level of information, the person interested in validation may find the dissimilar agreer a useful comparison other. His or her agreement can only be attributed to their common sharing of the "facts" and informs the individual that his or her own judgment is not based upon some personal idiosyncrasy.

The difference in preference for a dissimilar agreer that was observed between no and partial information conditions in Experiment 3 also serves to clarify an interpretational difficulty encountered with respect to the data in Experiment 1. As pointed out earlier, selection of the agreer as a partner in Experiment 1 could have been motivated by many factors —only one of which is validation. For example, the relatively stronger preference for an agreer in the partial information conditions could have been due to self-esteem motives. This same difficulty is not encountered with respect to Experiment 3 because of the addition of the similarity manipulation. A dissimilar agreer appears to be the relevant choice for determining whether one's judgment is accurate. The fact that this person is more preferred by partial than by no information subjects provides evidence of an increased concern for validation in the partial as opposed to the no information conditions. The similar agreer would appear to be the relevant partner choice if one were concerned with self-esteem maintenance or enhancement. Since the two conditions did not differ with respect to their preferences for the similar agreer, it appears that self-esteem motives do not vary as a function of perceived level of information.

It should be recalled that the subjects in Experiment 3 did not display absolute preferences for the predicted comparison other. That is, no information subjects did not most strongly prefer the similar disagreer. Nor did partial information subjects most strongly prefer the dissimilar agreer. These absolute preferences were to be expected to the extent that partner choice was dictated solely by informational concerns. Apparently, the experimental situation was not successful in creating such "pure" motives. Whether "pure" motives and absolute preferences occur in any given situation is likely to depend upon a number of additional parameters. In particular, it is likely to depend upon the importance of the task to the individual (cf. Jones & Regan, 1974). There are many reasons to choose a specific other as a partner. The less important the given task and informational social comparison reasons are in a

given situation, the more likely it is that these other factors will affect absolute preferences. In the present case, the noncritical nature of the experimental task and the anticipated ease of discussing a judgment with a similar agreer clearly affected absolute preferences. Only when the data are considered in relative terms are informational social comparison motives and their dependence upon perceived level of information apparent.

Why, when the data are considered in absolute terms, did so many subjects in each of the conditions prefer to have the similar agreer as a partner? One possibility is that subjects who chose the similar agreer may have done so because the task threatened their self-esteem. That is, rather than being primarily concerned with informational social comparison, these subjects may have been largely concerned with their self-esteem. A further analysis of the data provides some interesting insight. An internal analysis was performed in order to examine whether subjects who chose the similar agreer felt more uncertain of their judgments than did subjects who made the predicted partner choice. The discriminant weights from the multivariate analysis of variance on the difference ratio and symmetry measures of certainty were employed in order to construct a weighted sum that could serve as an index of certainty. Within the no information condition, those who chose the similar agreer tended to be less certain of their judgments than those who chose the similar disagreer ($p < .10$, by a two-tailed Mann–Whitney test). Among the partial information subjects, a similar pattern was evident. Those who chose the similar agreer tended to be less certain than those who selected the dissimilar agreer ($p < .10$). Collapsed across the two conditions, there was a significant difference between the subjects who chose the similar agreer and those who chose either the similar disagreer or the dissimilar agreer ($p < .01$). The results of this internal analysis suggest that an individual will not engage in either process of informational social comparison if he/she is very uncertain of his or her judgment. Rather than reveal the context of his or her uncertainty, the individual may simply ignore the need for informational social comparison. Concerns with self-esteem maintenance may become primary. The similar agreer is likely to have been perceived as the least threatening potential partner. Discussion with someone who is similar in perceptual style and who has agreed on past trials may have been viewed as a relatively safe and easy alternative. In addition, such comparison may provide a sufficient bolstering of one's certainty that one may subsequently be willing to test one's judgment in a more objective manner via informational social comparison.

Although the above internal analysis is quite revealing, the finding is a bit troubling in that it runs counter to the basic premise that informational social comparison is to be expected under conditions of uncertainty. Apparently, extreme uncertainty leads, not to a desire for information, but a feeling of threat and a desire to protect one's self-esteem. (See Conolley, Gerard, & Kline, 1978, for a similar discussion of the relation between uncertainty and informational vs. self-esteem motives within the context of ability comparison.) The current findings suggest that once a threshold of certainty is achieved such that the individual no longer feels threatened, the person is willing to "risk" informational social comparison. Whether construction or validation processes then occur will depend upon the person's level of uncertainty.

Future research will need to address the issue of informational versus self-esteem comparison further. Presumably, however, concerns for self-esteem among individuals who feel extremely uncertain would be overridden in situations where the individuals are making important and consequential decisions and need information to do so. That is, the threshold certainty level that must be achieved for a person to "risk" informational social comparison may vary as a function of the importance of the judgment in question. Furthermore, in such important situations, one of the comparison others whose value tends to vary as a function of perceived level of information (i.e., the similar disagreer or the dissimilar agreer) may become the most strongly preferred comparison other. That is, the relative effects observed in Experiment 3

when no and partial information conditions were compared may also become evident in absolute terms when important judgments are being made.

References

Conolley, E. S., Gerard, H. B., & Kline, T. Competitive behavior: A manifestation of motivation for ability comparison. *Journal of Experimental Social Psychology,* 1978, *14,* 123–131.

Fazio, R. H. *Motives for social comparison: The evaluation-validation distinction.* Unpublished doctoral dissertation, Princeton University, 1978.

Festinger, L. A theory of social comparison processes. *Human Relations,* 1954, *7,* 117–140.

Goethals, G. R. Consensus and modality in the attribution process: The role of similarity and information. *Journal of Personality and Social Psychology,* 1972, *21,* 84–92.

Goethals, G. R., & Darley, J. M. Social comparison theory: An attributional approach. In J. M. Suls & R. L. Miller (Eds.), *Social comparison processes: Theoretical and empirical perspectives.* Washington, D.C.: Hemisphere, 1977.

Goethals, G. R., & Nelson, R. E. Similarity in the influence process: The belief–value distinction. *Journal of Personality and Social Psychology,* 1973, *25,* 117–122.

Gruder, C. L. Choice of comparison persons in evaluating oneself. In J. M. Suls & R. S. Miller (Eds.), *Social comparison processes: Theoretical and empirical perspectives.* Washington, D.C.: Hemisphere, 1977.

Hakmiller, K. L. Threat as a determinant of downward comparison. *Journal of Experimental Social Psychology,* 1966, Supplement 1, 32–39.

Jones, S. C., & Regan, D. T. Ability evaluation through social comparison. *Journal of Experimental Social Psychology,* 1974, *10,* 133–146.

Kelley, H. H. Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 15). Lincoln: University of Nebraska Press, 1967.

Kelley, H. H. The process of causal attribution. *American Psychologist,* 1973, *28,* 107–128.

Langer, E. J. & Abelson, R. P. The semantics of asking for a favor: How to succeed in getting help without really dying. *Journal of Personality and Social Psychology,* 1972, *24,* 26–32.

Latane, B. Studies in social comparison—Introduction and overview. *Journal of Experimental Social Psychology,* 1966, *Supplement 1,* 1–5.

MacNeil, M. K., & Sherif, M. Norm change over subject generations as a function of arbitrariness of prescribed norms. *Journal of Personality and Social Psychology,* 1976, *34,* 762–773.

Mettee, D. R., & Smith, G. Social comparison and interpersonal attraction: The case for dissimilarity. In J. M. Suls & R. L. Miller (Eds.), *Social comparison processes: Theoretical and empirical perspectives.* Washington, D.C.: Hemisphere, 1977.

Sherif, M. A study of some social factors in perception. *Archives of psychology,* 1935, *27* (No. 187).

Singer, J. E. Social comparison—Progress and issues. *Journal of Experimental Social Psychology,* 1966, *Supplement 1,* 103–110.

Stevens, L., & Jones, E. E. Defensive attribution and the Kelley cube. *Journal of Personality and Social Psychology,* 1976, *34,* 809–820.

Thornton, D. A., & Arrowood, A. J. Self-evaluation, self-enhancement, and the locus of social comparison. *Journal of Experimental Social Psychology,* 1966, *Supplement 1,* 40–48.

Zanna, M. P., Goethals, G. R., & Hill, J. F. Evaluating a sex-related ability: Social comparison with similar others and standard setters. *Journal of Experimental Social Psychology,* 1975, *11,* 86–93.

Received September 27, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 10, 1699-1712

The Social Learning Theory of Sex Differences: Imitation is Alive and Well

David G. Perry
University of Alberta, Edmonton, Canada

Kay Bussey
Stanford University

The hypothesis that sex role development depends in part upon children's tendencies to imitate same-sex individuals more than opposite-sex models is central to most theories of sex typing. Yet Maccoby and Jacklin, in an influential recent review of the literature, conclude that the hypothesis has been disconfirmed. In the present article, it is argued that the research on which Maccoby and Jacklin based their conclusion is weak both methodologically and conceptually. This article presents a modified social learning theory account of the contribution of imitation to sex role development. It is suggested that children learn which behaviors are appropriate to each sex by observing differences in the frequencies with which male and female models as groups perform various responses in given situations. Furthermore, children employ these abstractions of what constitutes male-appropriate and female-appropriate behavior as models for their imitative performance. A first experiment confirmed that children engage in these processes. A second experiment extended the validity of the formulation by showing how it accounts for children's imitation of individual adult models. Specifically, it was shown that a child's imitation of an adult is strongly influenced by the degree to which the child believes that the adult usually displays behaviors that are appropriate to the child's sex (that is, behaviors that occur with greater frequency among individuals of the child's sex than among opposite-sex persons). In sum, the present research reinstates same-sex imitation as a viable mechanism of sex role development.

Sex role development, or sex typing, is the process by which children come to adopt the attitudes, feelings, behaviors, and motives that are culturally defined as appropriate [1] for their sex (Hetherington, 1967; Mischel, 1970). Virtually every leading theory of sex typing assigns a prominent place to imitation in the process. Psychoanalytic theory (Freud, 1949), cognitive-developmental theory (Kohlberg, 1969), cognitive consistency theory (Kagan, 1964), and social learning theory (Bandura & Walters, 1963; Mischel, 1966, 1970) all suggest that psychological differences between the sexes are at least in part perpetuated by the fact that boys and girls as groups are more inclined to imitate responses displayed by same-sex models than they are to imitate opposite-sex models. Furthermore, within each sex, individual differ-

This research was supported by an Australian Research Grant to the first author and an Australian Commonwealth Postgraduate Scholarship to the second author. The authors thank Judy Leah, Louise Perry, and Ray Pike for commenting on an earlier draft and the Department of Education, Brisbane, for providing a subject pool.

Requests for reprints should be sent to David G. Perry, Department of Psychology, University of Alberta, Edmonton, Alberta, Canada T6G 2E9.

[1] The question of whether a given response is more "appropriate" to one or the other sex can be determined in any of a number of ways. These include consensus by a panel of judges, empirical determination of the relative frequency with which the response is performed by the two sexes, and differences in the consequences boys and girls receive for performing a given response. Regardless of the manner in which one chooses to operationalize sex-appropriateness of behavior, however, the processes through which sex role learning is effected are likely to be similar, and it is with process that the present article is concerned.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3710-1699$00.75

ences in children's masculinity and femininity (or the degree to which a child displays responses appropriate, respectively, to the male or female sex) are believed to stem in part from the strength of children's tendencies to prefer imitation of one sex model over the other.

Recently, the validity of the hypothesis that psychological differences between the sexes are due in any significant degree to differential imitation of same-sex models has come under serious attack, notably by Maccoby and Jacklin (1974). These authors argue forcefully in their influential volume that the research to date fails to support the idea that differential imitation of like-sex models is probably a mechanism through which sex differences eventuate.

In arriving at their conclusion, Maccoby and Jacklin (1974) reviewed more than 20 imitation studies in which both sex of subject and sex of model had been included as factors in the research design, and they discovered that the vast majority of such studies failed to find a significant interaction of sex of observer and sex of model on children's imitation. A more recent review (Barkley, Ullman, Otto, & Brecht, 1977) included more than 80 studies and reached essentially the same conclusion. Furthermore, the same null result holds regardless of whether the modeled behavior is traditionally sex-typed or not and of whether the models are strange adults or familiar persons such as the child's own parents. Maccoby and Jacklin therefore concluded that it is necessary to search for factors other than same-sex imitation (e.g., biological determinants or direct elicitation and reinforcement of sex-appropriate behavior) to account for sex typing. If accepted, the Maccoby and Jacklin conclusion would constitute a devastating blow to most psychologists' understanding of the sex-typing process since, as has been pointed out, most theories assign a prominent role to the imitation of same-sex models.

We believe that it is premature to abandon the same-sex imitation hypothesis altogether. The results of these literature reviews suggest, however, that if the same-sex imitation hypothesis is to be retained, then a reconceptualization of the way in which same-sex imitation contributes to sex role development may be in order.

It is our contention that the typical investigation designed to explore same-sex imitation has employed an experimental paradigm that is conceptually remote from how imitation actually contributes to sex role development and therefore that the null results of much of the previous research on same-sex imitation constitute an inappropriate base for rejecting the same-sex imitation hypothesis. In the typical investigation, children are exposed to a single (often unfamiliar) male and/or female performing (often novel) responses and are then tested for imitation of the models' responses. The same-sex imitation hypothesis, in its usual form, predicts that children will view the actions of the single same-sex model as a better guide for their own behavior than the actions of the opposite-sex model and will thus imitate the former model more. This is actually quite unlike the way in which children use the responses of same- and cross-sex models as guides to their own behavior in naturalistic settings. In the real course of development, children discern what behaviors are appropriate for each sex by watching the behavior of *many* male and female models and by noticing differences between the sexes in the frequency with which certain behaviors are performed in certain situations. They then use these abstractions of sex-appropriate behavior as "models" for their own imitative performance. This is quite a different view from assuming that children develop such widely generalized tendencies to imitate same-sex models that they readily take the behavior of a single same-sex model as a more reliable guide to their own behavior than the behavior of an opposite-sex model, even when the same-sex model is unfamiliar or displays novel responses, or when the child has no idea about whether the model's behavior is likely to be consensually validated by others of the model's sex.

When conceptualizing the contribution of imitation to sex-role development, we find it useful to draw on the distinction made by social learning theorists between the observational learning and performance of modeled

behavior (Bandura, 1965, 1977; Mischel, 1970). According to social learning theory, through the observation of the behavior of models of both sexes, children cognitively acquire the potential to perform a vast repertoire of social behaviors that includes not only behaviors appropriate to the child's own sex but also (at least most of) the behaviors appropriate to the opposite sex. This is accomplished through such observational learning strategies as attending to the models' actions, coding the models' responses into covert imaginal or symbolic representations, and mentally rehearsing the models' actions. We propose that an additional and crucial (for sex typing) aspect of the observational learning process involves children's coding, or organizing in memory, of various responses as male-appropriate or female-appropriate on the basis of their having witnessed different proportions of available male and female models performing the responses. For example, if children observe that 80% of available male models perform a particular response in a given situation but only 5% of female models perform it, they are likely to code the response as male-appropriate or masculine.

Not everything learned is performed, however. Thus, although most boys know how to put on a dress, apply makeup, and primp in front of a mirror, few boys actually choose to perform these behaviors. Children are most likely to perform behaviors they have coded as appropriate for their own sex. According to social learning theory, which responses children select to perform from their repertoires depends primarily on the response consequences the children anticipate. Children should prefer to perform behaviors exhibited by same-sex models, because they have more often been rewarded and less often criticized for imitating same-sex models and because they gradually learn that the social environment holds similar expectations and imposes similar reinforcement contingencies on them and others of their sex (Bandura, 1969; Bussey & Perry, 1976; Mischel, 1970). In sum, although children acquire, via observational learning, behaviors that are appropriate to both sexes and label certain responses as same-sex-appropriate and others as opposite-sex-appropriate, they prefer spontaneously to imitate actions coded as same-sex-appropriate. This is because they infer that what is expected or permissible behavior for others of their sex is also likely to be appropriate behavior for them.

Our first experiment constitutes a laboratory analog study of how we hypothesize same-sex imitation to operate in naturalistic settings. In real life, male and female models as groups frequently diverge in the responses they display in a given situation. When confronted with a similar situation and the option to perform one of the responses they have seen male and female models exhibit, children should be inclined to choose an option they recognize as same-sex appropriate. Consider, for example, a 10-year-old boy who has recently transferred to a new school. In his first days, he observes that during recess the majority of girls play hopscotch, whereas most of the boys play baseball; that during library periods, most boys check out a Hardy Boys mystery, whereas most of the girls read Nancy Drew; that when given a choice of technical electives, boys choose woodwork, whereas girls choose sewing. Even if the boy has not previously observed sex differences in the performance of these various responses, he is now likely to code the responses as masculine or feminine and to use these codings as guides to his own response selection. Experiment 1 was designed to determine whether these suggestions are tenable.

Experiment 1

Method

Overview. Boys and girls were assigned to one of three modeling conditions or to a no-model control condition. In all the modeling conditions, children viewed 8 adult models (4 male, 4 female) individually and indicated their preferences on a series of two-choice preference tasks. In a first experimental condition, on every trial all the male models chose one of the items and all of the women chose the other item. In a second condition, on every trial 3 of the men and 1 of the women chose one item and the remaining man and women chose the other item. In a third condition, on every trial 2 of the men and 2 of the women chose one item and the other 2 men and 2 women chose the other item. Subjects in the no-

model condition skipped this phase. Subsequently, children in all groups were asked to indicate their own preferences for each item pair. The hypothesis was that children's endorsement of an item would be a function of the difference in the frequency with which same- and opposite-sex models had displayed a similar preference—the greater the difference, the more likely same-sex imitation was hypothesized to be.

Participants. Subjects were 96 8- and 9-year-old children, half boys and half girls, who attended a public school in a middle-class suburb of Brisbane, Australia. Models were four men and four women, all in their twenties. The experimenter was a female in her early twenties.

Stimuli. The preference task involved 16 pairs of objects. Within each pair the items belonged to the same class (e.g., a banana and an apple; a small plastic cow and a small plastic horse). Models were shown (on a black and white video monitor) making their choices from among the actual items, which were presented to them by an adult female (not the experimenter). The preference task was presented to subjects by showing them 16 color slides, each depicting one item pair. We deliberately selected responses as modeling stimuli that did not have clearly established histories of being more appropriate for one sex than the other. This was because our major goal was to demonstrate in the laboratory a major process through which previously sex-neutral responses *become* sex-linked. Furthermore, the use of already sex-typed responses could introduce potentially troublesome and irrelevant factors (e.g., the relative interest value of masculine and feminine behaviors) that might render unambiguous interpretations of the results in terms of cause and effect relations quite difficult.

Procedure. Each child was brought individually by the experimenter to a spare room on the school premises and, if assigned to a modeling condition, was seated before the video monitor. To modeling subjects, the experimenter explained that they would watch some people telling which things they "like better." In all modeling groups, the children saw, on the video monitor, the experimenter place the first item pair on a table before the eight models, individually solicit a preference from each model, and then continue to the next item pair until all 16 trials were completed. The models, who were asked for their preferences in a different order on every trial, always indicated their preference by pointing to the item of their choice and saying, "I like the (item) better." On each trial, four models always chose one item and the other four the remaining item. This equated for all groups the number of models endorsing a given item. Twelve boys and 12 girls were randomly assigned to each of three modeling conditions or to a no-model control condition which were run concurrently in a randomized blocks procedure.

In the first modeling condition, or *high (100%) within-sex consensus* modeling condition, for every item pair all the men chose one item and all the women chose the other item. Two modeling tapes were made for this condition, with half the subjects seeing one tape and the remainder seeing the other. The tapes were designed to counterbalance for the sex of model displaying a given item preference. For the first tape, determination of whether the men or the women chose a given item was determined randomly for each trial. On the second tape, the sex of the models choosing a given item was opposite to the sex of the models who had chosen that item on the first tape.

In the second modeling condition, or *moderate (75%) within-sex consensus* modeling condition, on every trial three men and one woman chose one item and the remaining three women and one man chose the other item. The identity of the man who chose as did the majority of the women and of the woman who chose as did the majority of the men changed from trial to trial, with each model serving four times as the "odd person out" for his or her sex. Half of the subjects watched a tape in which the items preferred by the three male models were identical to those preferred by all the men in the first tape used in the high within-sex consensus condition; the remaining subjects saw a tape in which the items preferred by the three male models were those that had been preferred by all the men in the second tape used in the high within-sex consensus condition.

In the third modeling condition, or *low (50%) within-sex consensus* modeling condition, on each trial two of the men and two of the women chose one item, whereas the remaining models chose the other item. Again, two tapes were used, with half the subjects viewing each. In the first tape, two men and two women were randomly drawn, for each trial, to make the same item choices as those made by all the men for that trial on the first tape used in the high within-sex consensus condition. The remaining models chose the other item. In the second tape, the two men and the two women who had been drawn for a given trial made the choice for that trial that had been made by all the male models in the second tape of the high within-sex consensus condition.

The remaining 12 boys and 12 girls were assigned to a no-model condition and proceeded directly to the imitation test after arrival at the experimental room. Children in the modeling groups proceeded to the imitation test immediately after the models had made their choices for the 16th item pair and the video equipment had been shut off.

In the *test for imitative performance,* children were shown the 16 slides depicting the item pairs and were told to indicate for each slide (verbally as well as by using a pointer) which item they preferred.

Subjects in the three modeling conditions were then given a *recall test.* The experimenter once again displayed the 16 slides, pointed to one of the two items for each slide, and instructed the child to tell if more of the men had liked that item, if more of the women had liked it, or if an equal number of men

and women had liked it. Children were promised a token (exchangeable for small prizes) for each correct answer.

Children in all groups were told they had performed very well and were awarded little prizes that were distributed when testing was completed.

Results

Imitative performance. An imitation score was derived for every child in the study by counting the number of the child's choices that matched the choices that had been unanimously *displayed by the male models in the high within-sex consensus modeling condition.* These scores of course could vary from 0 to 16 for a given child. The reasoning behind choosing imitation of the responses displayed by the male models in the high within-sex consensus condition as the datum to be compared across treatments is the following. According to the hypothesis of the research, boys' imitation of these responses should be highest in the high within-sex consensus condition and girls' imitation should be lowest in this condition. This follows from the prediction that children's endorsement of an item will be a function of the percentage of same-sex models who had displayed the preference. In this condition, 100% of the male and 0% of the female models endorsed these items. To see how changes in the percentage of same- and opposite-sex models endorsing an item alter imitation, it is necessary to examine imitation of the *same responses* but under conditions where the percentages of same- and opposite-sex models endorsing them have changed. This is most easily done by maintaining imitation of the male models' responses from the high within-sex consensus condition as the dependent variable across treatment conditions. As we move from examining children's imitation of these responses in the high within-sex consensus condition to the moderate and then the low within-sex consensus conditions, boys' scores should steadily decrease, and girls' should steadily increase. This is because the percentage of male models endorsing the preferences exhibited by the male models in the high within-sex consensus condition decreases as one moves to the moderate and then the low conditions, whereas the percentage of female models endorsing these responses increases as we move in this direction.

Recall that within each of the three modeling conditions, two modeling tapes were used (in order to counterbalance the models' choices), with half the children in each condition viewing one tape and the other half seeing the other tape. This affected computation of children's imitation scores in two ways. First, imitation scores for subjects in each modeling condition who viewed the first modeling tape for their condition (described in the method section) were computed by counting the number of the children's choices that matched the choices unanimously made by the male models in the first modeling tape of the high within-sex consensus modeling condition; imitation scores for subjects in each condition who viewed the second tape for their condition were computed by counting the children's choices that matched the choices unanimously made by the male models in the second modeling tape of the high within-sex consensus modeling condition. Second, imitation scores for subjects in the no-model group were calculated as follows: For a random half of the subjects, scores were taken as the number of choices that matched those of the male in the first modeling tape of the high within-sex consensus modeling condition; for the remaining no-model subjects, scores were taken as the number of choices that matched those of the male models in the second modeling tape of the high within-sex consensus modeling condition.

The mean imitation scores are presented separately for boys and girls for each treatment condition in Table 1. It may be seen that the data lend strong support to the hypothesis that children's imitation of a response increases as the percentage of same-sex models displaying the response increases. It is interesting to note that essentially identical results were achieved for girls. The fact that girls imitated, on the average, only 2.8 of the males' responses in the high within-sex consensus condition of course means that they were imitating 13.2 of the choices that 100% of the female models had made. As the percentage of female models endorsing these re-

Table 1
Imitative Performance Means of Boys and Girls in Each Treatment Condition (Experiment 1)

Child	Treatment condition			
	High (100%) within-sex consensus	Moderate (75%) within-sex consensus	Low (50%) within-sex consensus	No-model control
Boys	13.9_e	11.6_d	8.7_c	8.5_c
Girls	2.8_a	5.8_b	8.5_c	8.2_c

Note. For all treatment conditions, a subject's imitation score was the number of his or her preferences that matched those made by the male models in the high within-sex consensus condition. A subject's score could range from 0 to 16. Each mean is based on scores of 12 subjects. Cells not sharing a common subscript differ significantly from each other ($p < .05$) by a two-tailed t test.

sponses declined, so too did the girls' imitation of them. It can be seen, then, that the imitative performance of both boys and girls is strongly affected by changes in the frequency with which same- and opposite-sex models display the behavior.

An analysis of variance was performed on the means in Table 1, with sex of child and treatment condition as factors. There was a strong main effect of sex of child, $F(1, 88) = 162.87$, $p < .001$, as well as a significant interaction of sex of child and treatment condition, $F(3, 88) = 58.50$, $p < .001$. The sex main effect is, of course, due to the fact that imitation scores were determined by children's imitation of preferences given by the male models in the high within-sex consensus condition. If imitation of the preferences given by the female models in this condition had arbitrarily been chosen as the dependent variable, then of course a highly significant main effect of sex in the opposite direction would have been obtained. The significant interaction effect and its pattern clearly substantiate the conclusions reached above.

A final comment on the performance results is in order. When children skipped the modeling phase and therefore had no information about the sex-appropriateness of the responses, their imitation of the responses was not markedly different from when they observed equal numbers of male and female models display them (as in the low within-sex consensus condition). In fact, within each of the two rows of Table 1, all pairwise comparisons between means are significant (by two-tailed t tests) at $p < .05$, except for the comparison between the mean of the low within-sex condition and that of the no-model control condition.

Recall. Children in each modeling condition were asked, for each item choice displayed by the male models in the high within-sex consensus condition, whether more of the men, more of the women, or equal numbers of men and women had chosen it. The number of correct answers the children gave were analyzed in a 2×3 analysis of variance with sex of subject and modeling treatment condition as factors. The only significant effect was the main effect of treatment condition, $F(2, 66) = 7.78$, $p < .002$. Mean correct responses for the high, moderate, and low within-sex consensus conditions, respectively, were 14.4, 12.6, and 11.9. Clearly, the more consensus in choices there is among models of a given sex, the better able children are to remember whether and how the behavior is sex-typed. Boys and girls do not seem to differ in their propensities to do this.

Discussion

The research clearly supports our revised version of the same-sex imitation hypothesis. It is quite clear that when children observe same-sex models as a group exhibiting a response that diverges from the response made collectively by opposite-sex models under the same circumstances, the children are far more likely to imitate the response made by same-sex models. Obviously, the research restores same-sex imitation as a plausible mechanism of sex role development and reverses the Mac-

coby and Jacklin (1974) decision. However, the study suggests that the same-sex model for a child's imitation is the child's abstracted definition of appropriate behavior for the child's sex based on the child's observation of multiple examples of each sex behaving in a particular situation.

Experiment 2

Previous research on the same-sex imitation hypothesis has been guided by the assumption that children develop strongly generalized tendencies to imitate same-sex models. Recent reformulations of social learning theory (Bandura, 1977; Bandura & Barab, 1971; Mischel, 1968, 1973), however, stress the discriminating nature of children's selection of responses and models from among the vast array available to them. The negative results of research on the same-sex imitation hypothesis suggest that children do not automatically assume that a strange same-sex adult is a more appropriate role model for them than an opposite-sex model. Children probably learn quite early that individual differences exist among male and female models in terms of their masculinity and femininity and thus their suitability as models. Children should imitate a model when they believe the model's behavior is appropriate for them, too, but in order to conclude that a same-sex model is an especially reliable guide for their own behavior, the children must have some fairly unambiguous indication that the model is behaving typically for his or her sex (and is therefore an appropriate model). Experiment 2 tested this possibility.

Method

Overview. In an initial phase, children observed 8 adult models, 4 male and 4 female, indicating their preferences on a series of two-choice preference tasks. On each trial, 3 of the men and 1 woman chose one item, whereas the other man and the remaining women chose the other item. Across the trials, it was always the same man who chose as the 3 women did, and it was always the same woman who chose as the 3 men did. Thus, it was expected that over the trials children would learn that 3 of the men consistently concurred in their choices and thus were behaving sex-appropriately, whereas 1 of the men consistently chose with the majority of the women and thus was behaving sex-inappropriately. Similarly, children were expected to learn that the behavior of 3 of the women was sex-appropriate but that that of the fourth was sex-inappropriate. Some children (assigned to a no-premodeling control group) skipped this first phase and proceeded straight to the second phase.

In a second phase, children viewed 1 of the 8 models making choices alone on a new series of preference trials. For half of the children who had participated in the first phase, this model was one of those who had behaved appropriately for his or her sex during the first phase; for the other half of the children who had participated in the first phase, the model was one who had behaved sex-inappropriately in the first phase. Children who skipped the first phase (no-premodeling control subjects) also saw one of the models from the first phase, but of course they had no idea of the degree to which the model's behavior had previously agreed or disagreed with that of other members of his or her sex. Half of the children in each of these three groups saw a male model and the other half a female model.

In a third phase, children were tested for imitation of the preferences the single model had displayed in the second phase. It was predicted, first, that same-sex imitation effects would not occur for children assigned to the no-premodeling control group because these children (as in most previous studies) would not know whether the model was behaving appropriately or inappropriately for his or her sex. Second, it was predicted that same-sex imitation effects would be extremely powerful among children exposed to a same-sex model whose behavior the children knew had a history of concurrence with other same-sex models.

Participants and design. Subjects were 84 8-year-olds, half boys and half girls, enrolled at a school in a middle-class suburb of Brisbane. Six boys and six girls were assigned to cells in a 2 × 3 × 2 design involving Sex of Child × Sex-Appropriateness of the Single Model Seen in Phase 2 (appropriate, inappropriate, or no-premodeling control) × Sex of Model. An additional six boys and six girls were assigned to a no-model control condition. These children skipped the first two phases (that is, both modeling phases) but were asked for their own preferences during the third phase just as the other children were. Children had an equal chance of being assigned to any of the experimental design groups or to the no-model group, which were run concurrently in a randomized blocks procedure. The models and the female experimenter were the same as in the first experiment.

Stimuli. Stimuli used to solicit models' preferences during the first, multiple-modeling phase were different from but similar to those used in the first experiment. Again, there were 16 item pairs. The same set of 16 item pairs was used during the second, single-modeling phase as had been used in Experiment 1. Models in both phases were shown making choices from the actual items on a black-and-white video monitor. Children were asked for

their own preferences in the third phase and were shown the 16 color slides used in Experiment 1.

Procedure. Subjects were tested individually in a spare room at the school. Subjects participating in the first phase (multiple modeling) viewed the eight models making their choices, as in Experiment 1, on the video monitor. Three of the men always chose one item and three of the women always chose the other item; the remaining man and woman (inappropriate models) always chose as the majority of the opposite-sex models did. It should be clear that this treatment was identical to that received by subjects in the moderate within-sex consensus condition of Experiment 1, with the exception that in the present experiment the identity of the odd model out for his or her sex remained constant across trials. No-premodeling control children and no-model control children skipped this phase.

Children who had participated in the first phase or who were assigned to the no-premodeling control group participated in a second phase (single modeling). Children saw one of the male or one of the female models from the multiple modeling video tape indicate his or her preferences on a series of 16 new item choices. This modeling sequence was also presented on videotape. For children who had participated in the first, multiple-modeling phase, this model was one whom they had previously seen behave either appropriately or inappropriately for his or her sex. Children in the no-premodeling control condition observed one of the same models that children who had participated in the multiple-modeling phase saw, but as noted, they had no basis for knowing whether the model had a history of behaving appropriately or inappropriately for his or her sex.

To minimize contamination of the results by idiosyncratic attributes of the particular model that children viewed during this second, single-modeling phase, different models served as appropriate and inappropriate models for different groups of children. To have given each of the eight models a chance at participating in each of the various roles he or she could have assumed would have necessitated making a prohibitively large number of modeling tapes. As a compromise, the following procedure was used. Two multiple-modeling tapes were made for use in the first phase. Half the subjects participating in this phase saw one tape and the other half saw the other tape. For each of the two tapes, a different male or female model served as the model who made inappropriate choices for his or her sex. Children who were assigned to see an inappropriate model in the second phase of course viewed whichever model had served as the inappropriate model in the tape they had seen in the first phase (with half seeing the male and half seeing the female model). Children who were assigned to see an appropriate model in the second phase, however, viewed whichever model had served as the inappropriate model in the tape they had *not* seen (but who, of course, had served as one of the three appropriate models in the tape they *had* seen). For children assigned to the no-premodeling control condition, half viewed one of the two male models who had served as either an appropriate or an inappropriate model for children in the other conditions, and the other half viewed one of the two female models who had served in one of these roles for the other children. These procedures equated the frequency across conditions with which children viewed a particular model in the second phase as well as permitted more than one of the male and one of the female models to play the sex-inappropriate role.

In this study all the models made the same choices in the second phase. Extra tapes depicting the models making opposite choices were not used because examination of the data from Experiment 1 indicated that similar results were obtained regardless of the models' actual choices. The model's choices in the present experiment were the same as the choices seen by the children who saw the first tape in the high within-sex consensus modeling condition of the last experiment.

All children, including those assigned to a no-model control condition, were then shown the 16 slides depicting the item pairs used in the second phase and were told to give their own preferences, as in Experiment 1.

Children in the modeling conditions were asked to recall the choices the model had made in the second phase and were promised rewards for doing so. Finally, children were asked to respond on 7-point scales to the questions, "How much do you like (the model)?", "How much are you like (the model)?" "Do you think that most people would like (the model)?", and "How good is (the model) at making choices?" Children received small prizes for participating and were returned to the classroom.

Results

Imitative performance. A $2 \times 3 \times 2$ analysis of variance with sex of child, sex-appropriateness of the model, and sex of model as factors was performed on imitation scores of all the children except those assigned to the no-model control condition. Significant effects included sex of subject, $F(1, 60) = 9.32$, $p < .005$; sex-appropriateness of model, $F(2, 60) = 19.55$, $p < .001$; the interaction of sex of subject and sex of model, $F(1, 60) = 9.92$, $p < .005$; and sex of model with sex-appropriateness of model, $F(2, 60) = 3.27$, $p < .05$. However, the three-way interaction among all the variables was also highly reliable, $F(2, 60) = 41.15$, $p < .001$. The means for this analysis are given in Table 2. The least significant difference required for

Table 2
Imitative Performance Means for Interaction of Sex of Subject, Sex-Appropriateness of Model, and Sex of Model (Experiment 2)

Subject and model	Sex-appropriateness of model		
	Appropriate	Inappropriate	No-premodeling control
Boys			
Male model	14.7_{f}	4.3_{a}	13.2_{def}
Female model	3.5_{a}	9.3_{be}	12.0_{de}
Girls			
Male model	8.2_{b}	11.2_{cd}	12.7_{def}
Female model	14.3_{ef}	8.3_{b}	12.8_{def}

Note. Each mean is based on six subjects' scores. Cells not sharing the same subscript are significantly different from one another ($p < .05$) by a two-tailed t test.

any two means in this table to be reliably different from one another at $p < .05$ by a two-tailed t test is 2.62.

The results confirm the hypotheses. First, examine the data for children assigned to the no-premodeling control condition by glancing down the third column in Table 2. None of the means in the column is significantly different from any other mean in the column. This substantiates the prediction that when children do not know how the behavior of a model compares with that of other members of the model's sex, the same-sex imitation hypothesis is not confirmed. Also note that the means in this column are all relatively high. This suggests the possibility of a ceiling effect. When children see only one model displaying item preferences in an experimental setting and they do not know whether the model is a "good" or "bad" example for his or her sex, they may assume that the model—regardless of sex—is likely to be a relatively safe example to follow in that setting. That the high scores in the no-premodeling groups are in fact due to imitation is apparent from the fact that children assigned to the no-model control condition made choices that matched the models' at a substantially and significantly ($p < .05$) lower rate (M for boys in no-model group = 8.0; M for girls = 8.2).

The second hypothesis was that same-sex imitation effects would occur when the children were exposed to a model whose behavior, they had learned, was typical of others of the model's sex. It is evident, by glancing down the first column of Table 2, that this hypothesis received strong support, especially for boys. Note how boys who saw a sex-appropriate female model actually chose fewer of the model's responses than they did when they had not seen any model perform at all ($p < .05$).

Two other features of the pattern of means are noteworthy. First, glance down the middle column of means in Table 2, where data are presented for children's imitation of a model whom they had previously seen behave sex-inappropriately. Apparently, when children know an opposite-sex model is a more reliable indicator of behavior appropriate to the child's own sex than a same-sex model, they are more likely to imitate the opposite-sex model. Interestingly, cross-sex imitation is really evident only among girls, because the imitation scores displayed by girls exposed to an inappropriate male model ($M = 11.2$) were significantly ($p < .05$) higher than the imitation scores of no-model control girls ($M = 8.2$). In contrast, boys' imitation scores were not significantly higher with an inappropriate female ($M = 9.3$) than in the no-model control condition ($M = 8.0$). Boys may ignore rather than imitate females who behave masculinely.

A second point is that boys more actively inhibited imitation of sex-inappropriate behavior than girls. Imitation scores of boys who saw either an inappropriate male model ($M = 4.3$) or an appropriate female model ($M = 3.5$) were lower ($p < .05$ in each case) than those of no-model control boys. In con-

trast, girls did not inhibit imitation of male-appropriate behavior, when it was displayed by either a male ($M = 8.2$) or a female ($M = 8.3$) model.

Recall. Correctly recalled choices of the model were subjected to a similar $2 \times 3 \times 2$ analysis of variance. No effect attained significance (Grand $M = 15.1$).

Questionnaire data. Children's responses to the four postexperimental questions were subjected to similar three-way analyses of variance. No significant effects were obtained on the first question, which asked children to indicate their attraction to the model. For the second question, which asked children to rate their similarity to the model, several effects were significant, but they were all subsumed by a significant highest order interaction, $F(2, 60) = 14.03$, $p < .001$. The means indicated that when children saw a sex-appropriate model, there was stronger perceived similarity to a same-sex model for both boys and girls; when children saw a sex-inappropriate model, boys actually expressed greater similarity to the female model than the male model (no parallel result occurred for girls); among children in the no-premodeling conditions, there were no differences in perceived similarity to the model. It is, of course, important to remember that these data were collected after the subjects' imitation test and may thus have been influenced by their imitative performance. On the third question, which asked children to tell whether most people would like the model, there were no significant effects. On the final question, which asked children to indicate how good they thought the model was at making choices, there was a significant three-way interaction, $F(2, 60) = 3.97$, $p < .05$. It appeared that the only major contributor to this interaction was the fact that boys in the no-premodeling control condition perceived the female model as more competent than the male at making choices. Why this result emerged is unknown. However, in sum it appears that neither perceptions of the model's competence nor attraction to the model can account for the imitation results. On the other hand, perceived similarity to the model does appear as an influential possibility.

General Discussion

The major contribution of the present research is to reinstate same-sex imitation as a viable mechanism of sex role development. Maccoby and Jacklin's (1974) conclusion that differential imitation of same-sex models contributes negligibly to sex role development was based on the results of studies that lacked an adequately developed conceptual framework of how imitation contributes to the social learning of sex differences. The present article demonstrates quite clearly that children discern behaviors appropriate to the two sexes by observing differences between the sexes in the frequency with which they perform various responses in a given situation and, furthermore, that children use these abstractions concerning sex-appropriate behaviors as guides to their own performance in similar situations.

The research also demonstrates that a child's imitation of a single model is in part governed by the degree to which the child has come to believe that the model's behavior is ordinarily appropriate or inappropriate to the child's sex. In other words, children are most likely to imitate persons whom they perceive to be good examples of their sex role. It is of interest that some previous research has shown that girls are especially likely to imitate a warm or nurturant model, whereas boys are particularly likely to imitate a dominant model (Hetherington, 1967). This may be in part because children usually perceive warmth as indicative of femininity and dominance as reflective of masculinity (Best, Williams, Cloud, Davis, Robertson, Edwards, Giles, & Fowles, 1977; Rothbaum, 1977; Williams, Bennett, & Best, 1975). Clearly, then, children judge the likely appropriateness of an individual model's behavior by recalling how well that person's attitudes and behaviors usually match up with others of his or her sex.

According to social learning theory, children acquire (via observational learning) responses appropriate to both sexes but prefer to perform responses exhibited by same-sex models. In other words, model sex affects performance more than observational learning. Results of the second study are quite in

line with this proposition. In this experiment, there were no significant effects of the experimental variables upon children's recall of the model's behavior, although the same factors strongly affected children's imitative performance. These results mesh with those of Williams et al. (1975), who found that boys and girls do not differ in their knowledge of cultural stereotypes for male and female behavior which, we could argue, are primarily learned via observational learning. We do not wish to overstate the case, however. Several studies do show that children sometimes attend more closely to, as well as learn more about, the behavior of models consistent with their own sex role (Grusec & Brinker, 1972; Maccoby & Wilson, 1957; Maccoby, Wilson, & Burton, 1958; Perry & Perry, 1975; Slaby & Frey, 1975). Nadelman (1974) also found that children were able to identify correctly the sex-appropriateness of items appropriate for their own sex better than they were able to identify the sex-appropriateness of items traditionally associated with the opposite sex. Thus, although the preference for performing same-sex behaviors may well be stronger than the preference for learning same-sex behavior, it is not as though the latter phenomenon is nonexistent.

Our analysis of same-sex imitation in sex role development suggests that children must master certain cognitive achievements before their sex typing can to any significant degree be influenced by imitation of like-sex models. Before children will encode responses they see performed more often by members of a given sex as male- or female-appropriate, they must realize that human beings are divided into males and females. Before they will prefer to perform responses coded as same-sex-appropriate, they must realize that not only do they belong to one of these sexes, but they are subject to a similar set of expectations and reinforcement contingencies as others of their sex. Children may learn that other people are divided into males and females before they establish their own gender identities (Thompson, 1975). Thus children may proceed through an initial stage in which they acquire some knowledge of male- and female-appropriate behavior (via observational learning) but, not realizing they belong to one of these sexes, may fail to display a preference for imitating same-sex-appropriate behavior. Such a possibility should be tested. Once children attain gender constancy they do attend more to same-sex models (Slaby & Frey, 1975). However, developmental research that tracks the emergence of children's gender identities and their beliefs that they are expected to behave similarly to members of a particular sex and relates these variables to children's imitation of same- and cross-sex models is needed.

Because the present formulation emphasizes that children imitate responses that they abstract as appropriate for their sex as a group, it clearly assigns less importance to children's imitation of particular individuals (e.g., the like-sex parent) than do more traditional theories, especially Freudian theory (1949) or various other views that hold that children achieve sex typing via imitation of a warm and/or socially powerful parental figure (Hetherington, 1967; Mussen & Distler, 1959; Parsons, 1955; Sears, Rau, & Alpert, 1965). Although this is true, we do not mean to imply that children's sex typing is never influenced by their especially strong imitation of "significant others," for it undoubtedly is. However, we would expect that children who initially adopt responses by imitating a highly nurturant and/or dominant same-sex parent would ultimately drop the responses from their active repertoires if they eventually realize that no other same-sex individuals perform the responses but that many opposite-sex persons do.

Although in the first experiment boys and girls did not differ in their propensities to imitate abstracted same-sex-appropriate behavior, there was clear evidence in the second experiment that boys were more concerned than girls with matching their behavior to a model known to exemplify sex-typical behavior. This concurs with previous findings that a variety of socializing agents, including parents and peers, enforce stronger sanctions on boys than on girls for conforming to culturally defined sex-role prescriptions (Fagot, 1977; Fling & Manosevitz, 1972) and that boys are more likely to conform to peer pressure than are girls (Maccoby &

Jacklin, 1974). It may indeed be the case, as Lynn (1969) suggests, that boys become sex-typed through "masculine role identification" by acquiring and imitating cultural stereotypes of male-appropriate behavior learned by observing a variety of male figures both within and beyond the home. Girls, on the other hand, are less concerned with conforming to abstract standards.

Although this research indicates that imitation of abstracted sex-appropriate behavior is a plausible mechanism of sex role development, it is not implied that other factors (e.g., genetics, direct elicitation, and shaping of sex-typed behavior) are unimportant. The present formulation is probably a weak or inadequate explanation of early sex differences, say, those occurring in the first year or two of life. As early as age 1, there are sex differences in children's toy preferences, though the exact qualities of the toys (e.g., softness, faceness, number of moving parts, manipulability) responsible for the differences have yet to be isolated (Jacklin, Maccoby, & Dick, 1973). It seems unlikely that such early differences are due to abstracted differences about sex-appropriate behavior learned from observation of multiple models. The cognitive capacities required to notice that the sexes differ in their frequency of performing certain behaviors, to register the information in memory as male- or female-appropriate, and to learn that one is expected to conform to sex-appropriate role behavior take time to develop. Obviously, other factors must be involved in early sex differences.

One might wish to level the criticism against the present research that results are attributable to "demand characteristics" of the experimental procedures. In particular, it might be argued that using multiple models and testing the child in the presence of an experimenter who knew how the models had responded placed stronger demands on the child to display same-sex imitation than would exist under more natural circumstances and therefore that the results are artifactual. We would argue that many sex-typed behaviors are indeed public, with people in a position to view a performer's behavior just as aware as the performer of the sex-appropriateness of various response options in the situation. Consequently, it is entirely appropriate to examine imitation of sex-appropriate and -inappropriate behavior in the laboratory under circumstances similar to those that exist outside of it (i.e., in the presence of an observer—here, the experimenter). Moreover, the results of a third experiment conducted by the authors indicate that same-sex imitation does occur even in the absence of the experimenter.[2]

In the procedures used in the present study, endorsement of a choice preferred by the majority of one sex of model was confounded with rejection of the choice preferred by the majority of opposite-sex models. To illustrate, consider the models' choices in the high within-sex consensus condition of the first experiment. The endorsement by 100% of the male models of a particular item was, of course, coupled with implicit rejection of the item unanimously chosen by the female models. In any two-item forced-choice situation, endorsement of one item implies rejection of the other. In the present research, it is not possible to determine, for example, if a boy's imitation of responses endorsed by male models was a function of an active desire to emulate the male models or a wish not to display a choice indicated

---

[2] To examine the generality of some of the findings from the first two experiments reported, the authors conducted a third experiment. The design was identical to that of Experiment 2 of the present report. However, the procedure was altered in three major ways. First, peers instead of adults were used as models. Second, although the first, multiple-modeling phase of the study was conducted exactly as in Experiment 2 (with models indicating their preferences from among item pairs), the second modeling phase involved the modeling of novel action and verbal sequences rather than the modeling of additional item preferences. Third, the test for imitation involved leaving the child alone with the objects the model had used in displaying the action and verbal responses in the second phase, with instructions to "see if you can do some interesting things with them" while the experimenter was gone. Children's imitation of responses displayed by the model in the second phase was assessed by a time-sampling technique while the child was observed through a one-way vision screen for a 4-minute period. Results were virtually identical to those obtained in Experiment 2.

by the female models. In real-life settings, of course, many choice situations do exemplify the two-pronged situation facing children in our study where one choice was masculine and the other feminine. For theoretical reasons however, it would be interesting to develop a methodology that permits separate manipulation and assessment of a propensity to accept same-sex models' behavior, on the one hand, and a tendency to reject opposite-sex models' behavior, on the other. For instance, the stronger same-sex imitation shown by boys in the present research may stem from simultaneous desires to accept masculine behavior and to reject feminine behavior. Perhaps girls are also prone to accept same-sex behavior but are less repulsed by the thought of displaying cross-sex behavior.

Conclusion

The present research restores the same-sex imitation hypothesis as a plausible mechanism of social learning of sex differences, though in a modified conceptual context. Many important questions remain, however. These include the nature of the processes accounting for sex differences in the early years; the exact roles played by parents, peers, unfamiliar adults, and other classes of social models in the process; the sequencing of such cognitive accomplishments as children's realization that the social world is divided into males and females, that they belong to one of these groups, that they and others of their sex are subject to a similar set of expectations and reinforcing contingencies from the social environment; the relationship of these factors to children's observational learning and performance of sex-type behavior; and whether acceptance of same-sex behavior and rejection of opposite-sex behavior contribute equally to same-sex imitation in children of both sexes.

References

Bandura, A. Influence of models' reinforcement contingencies on the acquisition of imitative responses. *Journal of Personality and Social Psychology,* 1965, *1,* 589–595.

Bandura, A. Social-learning theory of identificatory processes. In D. A. Goslin (Ed.), *Handbook of socialization theory and research.* Chicago: Rand McNally, 1969.

Bandura, A. *Social learning theory.* Englewood Cliffs, N.J.: Prentice-Hall, 1977.

Bandura, A., & Barab, P. G. Conditions governing nonreinforced imitation. *Developmental Psychology,* 1971, *5,* 244–255.

Bandura, A., & Walters, R. H. *Social learning and personality development.* New York: Holt, Rinehart & Winston, 1963.

Barkley, R. A., Ullman, D. G., Otto, L., & Brecht, J. M. The effects of sex typing and sex appropriateness of modeled behavior on children's imitation. *Child Development,* 1977, *48,* 721–725.

Best, D. L., Williams, J. E., Cloud, J. M., Davis, S. W., Robertson, L. S., Edwards, J. R., Giles, H., & Fowles, J. Development of sex-trait stereotypes among young children in the United States, England, and Ireland. *Child Development,* 1977, *48,* 1375–1384.

Bussey, K., & Perry, D. G. Sharing reinforcement contingencies with a model: A social-learning analysis of similarity effects in imitation research. *Journal of Personality and Social Psychology,* 1976, *84,* 1168–1176.

Fagot, B. I. Consequences of moderate cross-gender behavior in preschool children. *Child Development,* 1977, *48,* 902–907.

Fling, S., Manosevitz, M. Sex typing in nursery school children's play interests. *Developmental Psychology,* 1972, *7,* 146–152.

Freud, S. *An outline of psychoanalysis.* (1st ed., 1940). New York: Norton, 1949.

Grusec, J. E., & Brinker, D. B., Jr. Reinforcement for imitation as a social learning determinant with implications for sex-role development. *Journal of Personality and Social Psychology,* 1972, *21,* 149–158.

Hetherington, E. M. The effects of familial variables on sex typing, on parent–child similarity and on imitation in children. In J. P. Hill (Ed.), *Minnesota symposia on child psychology* (Vol. 1). Minneapolis: University of Minnesota Press, 1967.

Jacklin, C. N., Maccoby, E. E., & Dick, A. E. Barrier behavior and toy preference: Sex differences (and their absence) in the year-old child. *Child Development,* 1973, *44,* 196–200.

Kagan, J. Acquisition and significance of sex typing and sex role identity. In M. L. Hoffman and L. W. Hoffman (Eds.), *Review of child development research* (Vol. 1). New York: Russell Sage Foundation, 1964.

Kohlberg, L. Stage and sequence: The cognitive-developmental approach to socialization. In D. A. Goslin (Ed.), *Handbook of socialization theory and research.* Chicago: Rand McNally, 1969.

Lynn, D. B. *Parental and se x-role identification: A theoretical formulation.* Berkeley, Calif.: McCutchan, 1969.

Maccoby, E. E., & Jacklin, C. N. *The psychology*

of sex differences. Stanford, Calif.: Stanford University Press, 1974.

Maccoby, E. G., & Wilson, W. C. Identification and observational learning from films. *Journal of Abnormal and Social Psychology*, 1957, *55*, 76–87.

Maccoby, E. G., Wilson, W. C., & Burton, R. V. Differential movie-viewing behavior of male and female viewers. *Journal of Personality*, 1958, *26*, 259–267.

Mischel, W. A social-learning view of sex differences in behavior. In E. E. Maccoby (Ed.), *The development of sex differences*. Stanford, Calif.: Stanford University Press, 1966.

Mischel, W. *Personality and assessment*. New York: Wiley, 1968.

Mischel, W. Sex-typing and socialization. In P. H. Mussen (Ed.), *Carmichael's manual of child psychology*. New York: Wiley, 1970.

Mischel, W. Toward a cognitive social learning reconceptualization of personality. *Psychological Review*, 1973, *80*, 252–283.

Mussen, P. H., & Distler, L. Masculinity, identification, and father–son relationship. *Journal of Abnormal and Social Psychology*, 1959, *59*, 350–356.

Nadelman, L. Sex identity in American children: Memory, knowledge, and preference tests. *Developmental Psychology*, 1974, *10*, 413–417.

Parsons, T. Family structure and the socialization of the child. In T. Parsons & R. F. Bales (Eds.), *Family socialization and interaction process*. Glencoe, Ill.: Free Press, 1955.

Perry, D. G., & Perry, L. C. Observational learning in children: Effects of sex of model and subject's sex role behavior. *Journal of Personality and Social Psychology*, 1975, *31*, 1083–1988.

Rothbaum, F. Developmental and gender differences in the sex stereotyping of nurturance and dominance. *Developmental Psychology*, 1977, *13*, 531–532.

Sears, R. R., Rau, L., & Alpert, R. *Identification and childrearing*. Stanford, Calif.: Stanford University Press, 1965.

Slaby, R. G., & Frey, K. S. Development of gender constancy and selective attention to same-sex models. *Child Development*, 1975, *46*, 849–856.

Thompson, S. K. Gender labels and early sex role development. *Child Development*, 1975, *46*, 339–347.

Williams, J. E., Bennett, S. M., & Best, D. L. Awareness and expression of sex stereotypes in young children. *Developmental Psychology*, 1975, 635–642.

Received October 5, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 10, 1713-1722

Telling Lies

Bella M. DePaulo and Robert Rosenthal
Harvard University

Men and women (20 each) were videotaped while describing someone they liked, someone they disliked, someone they were ambivalent about, someone they were indifferent about, someone they liked as though they disliked him or her, and someone they disliked as though they liked him or her. Accuracy at detecting that some deception had occurred was far greater than accuracy at detecting the true underlying affect, and people who were good at detecting that deception was occurring were not particularly skilled at reading the speakers' underlying affects. However, people whose deception attempts were more easily detected by others also had their underlying affects read more easily. Speakers whose lies were seen more readily by men also had their lies seen more readily by women, and observers better able to see the underlying affects of women were better able to see the underlying affects of men. Skill at lying successfully was unrelated to skill at catching others in their lies. A histrionic strategy (hamming) was very effective in deceiving others, and this strategy was employed more by more Machiavellian people, who also tended to get caught less often in their lies. Methodological considerations and systematic programs for future research are discussed.

Studies of skill at detecting lies from verbal and/or nonverbal cues usually focus on observers' ability to distinguish truthful responses from deceptive ones (e.g., Ekman & Friesen, 1974; Fay & Middleton, 1941; Harrison, Hwalek, Raney, & Fritz, 1978; Kraut, 1978; Littlepage & Pineault, 1978; Maier & Janzen, 1967; Maier & Thurber, 1968; Zuckerman, DeFrank, Hall, Larrance, & Rosenthal, in press; Krauss, Geller, & Olson, Note 1). Most published studies of human lie detection have found that people are substantially more accurate than chance at distinguishing veracity from mendacity (Fay & Middleton, 1941; Feldman, 1976; Harrison et al., 1978; Kraut, 1978; Littlepage & Pineault, 1978; Maier, 1966; Maier & Janzen, 1967; Maier & Thurber, 1968; Matarazzo, Wiens, Jackson, & Janaugh, 1970; Zuckerman et al., in press; see also DePaulo, Zuckerman, & Rosenthal, in press).

Discerning the degree of deceptiveness of a given response, however, is only one of several senses in which one might be said to be skilled at detecting a lie. In the case of a lie that involves the cloaking of a felt emotion with a feigned one, for example, skill at detecting the lie might be conceptualized as the ability to see through to the underlying affect that the deceiver is trying to hide. Using Ekman and Friesen's (1969) terms, this latter skill (identifying the concealed information) will be called *leakage* accuracy, whereas the former (recognizing that deception is or is not occurring) will be labeled *deception* accuracy. Although it might seem quite plausible that people who are especially good at detecting deceptiveness

We thank Russell A. Eisenstat and Susan Finkelstein for their help in collecting these data, and Judith A. Hall, Charles M. Judd, and Miron Zuckerman for their helpful comments on an earlier version of this manuscript.

Preparation of this article was supported by the Milton Fund of Harvard University, Biomedical Sciences Support Grant 5S07 RR07046-12 from the National Institutes of Health to Harvard University, and the National Science Foundation. During the conduct of this research, the first author was supported by a National Science Foundation Fellowship.

Requests for reprints should be sent to Bella M. DePaulo, who is now at the Department of Psychology, Gilmer Hall, University of Virginia, Charlottesville, Va. 22901.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3710-1713$00.75

will also be particularly skilled at reading leakage, this empirical question has not yet been addressed.

Two-tiered lies, in which an underlying or "true" affect is covered up by a dissimulated or "sent" affect, come in many varieties. For instance, they may involve the covering up of a positive emotion with a negative one or the covering up of a negative emotion with a positive one. In the few studies that have examined these kinds of lies (Ekman & Friesen, 1969, 1974; Feldman, 1976), the liars were always attempting to send positive cues. In the Ekman and Friesen (1974) study, for instance, nurses watching a very disconcerting film pretended to be watching a very pleasant one; in Feldman's study, teachers either liked or disliked a student who had performed either well or poorly, but in all conditions teachers were instructed to give only positive feedback to the student. We cannot know from these studies whether skill at detecting these sugar-coated lies (in which the feigned affect is positive) is related to skill at detecting lies in which the dissimulated affect is negative.

Paralleling our lack of knowledge about the generality of skills at detecting different kinds of lies is a lack of knowledge about the generality of skills at decoding different kinds of liars. For instance, are people who are especially skilled at identifying deceptiveness in women also particularly successful in recognizing deceptiveness in men?

The same kinds of questions that are asked about people trying to detect lies can also be asked about people who are attempting to spread lies. That is, are those persons who are especially adept at hiding their attempts at deception also especially successful at concealing their true feelings? Are people who are especially good at hiding an underlying positive affect also particularly skilled at hiding an underlying negative affect? Are people who are particularly transparent to women also especially transparent to men? Finally, are people who are good liars also good lie-detectors; that is, does it take one to know one?

To study these and other questions, we asked subjects to describe several people whom they knew: someone they liked, someone they disliked, someone they felt ambivalent about, and someone they felt indifferent about. To elicit deception, we asked them to describe the liked person while pretending that they really disliked him or her, and we also asked them to describe the disliked person while pretending that they really liked him or her. Thus, the design included underlying positive affect covered by feigned negative affect and underlying negative affect covered by feigned positive affect. It also included "pure" positive affect (liking), "pure" negative affect (disliking), and simultaneously occurring positive and negative affect (ambivalence). We asked the same subjects to return to judge these various descriptions. We believe that this paradigm, in which the observer is called upon to discriminate among a wide variety of affects, is more valid ecologically than is the scenario characteristic of most deception studies, in which the subject is asked merely to decide whether the speaker is lying or telling the truth.

This design also permitted us to study a particular style of lying—one that we call *hamming*. We think that a reasonable operational definition of a ham is someone who when pretending to like someone he or she actually dislikes, expresses *more* liking than when describing a person he or she really does like. Another kind of ham is one who exaggerates dissimulated disliking—that is, one who expresses more disliking when pretending to dislike someone he or she actually likes than when describing someone he or she really does dislike.

We wanted to know several things about this style of lying:

1. How common is this hamming strategy?

2. Is there really just one kind of ham, or might there be at least two? That is, are the people who exaggerate liking the same people who exaggerate disliking?

3. Does hamming work? That is, do people who ham it up get caught in their lies less frequently than those who use a less exaggerated style? Conceivably, hamming could backfire if an observer got the impression that the speaker "doth protest too much."

Because Machiavellianism is commonly thought to be linked to success at lying—

and in one study (Exline, Thibaut, Hickey, & Gumpert, 1970) has in fact been shown to be so related—we administered the Mach scale (Christie & Geis, 1970) to all of our subjects.

Method

Subjects

Subjects were 40 Harvard summer school students (20 males and 20 females) recruited for a study of "person descriptions" and paid for their participation.

Procedure

Subjects were asked to take 1 minute to describe each of the following persons: someone they liked, someone they disliked, someone they felt ambivalent about, and someone they felt indifferent about. (Ambivalence was defined as strong feelings of both liking and disliking; indifference was defined as no strong feelings of liking or disliking.) To elicit deception, subjects were also asked to describe the person they liked as if they really disliked him or her (like-as-though-dislike or LD condition) and to describe the person they disliked as if they really liked him or her (dislike-as-though-like or DL condition). These six descriptions were given in one of nine different orders (randomly assigned). Half of the subjects (10 males and 10 females, randomly assigned) described males; the others described females.

During the sessions, the experimenter remained behind a one-way mirror and videotaped the descriptions. Subjects were aware of this, and also knew that the experimenter was not informed of the sequence of their descriptions. They were urged to try to be very convincing in all of their descriptions.

From these descriptions, two hour-long videotapes were made, composed of the middle 20 sec of each description, plus rating pauses. The order of appearance of the senders on these tapes was randomized. (Only the middle 20 sec were used to keep the rating task—described below—to a more manageable length.) All subjects returned to judge one of these videotapes. Subjects always judged a videotape on which they did not appear. Subjects saw the videotape they were judging twice: the first time, they rated the descriptions on 9-point scales of liking, ambivalence, and deception; the second time, they rated the descriptions on 9-point scales of disliking, discrepancy, and tension.[1]

Accuracy Scores

An accurate decoder of deception would rate the deceptive descriptions as very deceptive and the pure like and dislike descriptions as not very deceptive at all. Thus, deception accuracy was defined as a subject's mean deception rating of the like-as-though-dislike and dislike-as-though-like descriptions minus their mean deception rating of the pure like (L) and dislike (D) descriptions. Separate subscores were computed analogously for the LD and DL descriptions and for the decoding of male and female speakers.

$$\text{Overall deception accuracy} = \frac{\text{deception ratings of LD} + \text{DL}}{2} - \frac{\text{deception ratings of L} + \text{D}}{2};$$

$$\text{Deception accuracy for LD} = \text{deception ratings of LD} - \text{deception ratings of L};$$

$$\text{Deception accuracy for DL} = \text{deception ratings of DL} - \text{deception ratings of D}.$$

Accuracy of encoding deception was defined analogously. A speaker whose lies were easily detected would be rated as very deceptive in the LD and DL conditions but not very deceptive at all in the liking or disliking conditions. Thus the encoding deception score was defined as the mean deception rating (by all judges who rated the speaker) of the LD and DL descriptions minus the mean deception rating of the pure like and dislike descriptions. Separate subscores were computed analogously for LD and DL and for judgments made by male and female observers.

$$\text{Overall deception accuracy} = \frac{\text{rated deception on LD} + \text{DL}}{2} - \frac{\text{rated deception on L} + \text{D}}{2};$$

$$\text{Deception accuracy for LD} = \text{rated deception on LD} - \text{rated deception on L};$$

$$\text{Deception accuracy for DL} = \text{rated deception on DL} - \text{rated deception on D}.$$

A judge more skilled at seeing underlying affects (decoding leakage) would rate a speaker pretending to like someone she/he actually disliked as disliking

[1] Scale endpoints were labeled as follows: speaker does not like (dislike) the person at all (1), speaker likes (dislikes) the person very much (9); speaker is not very ambivalent (deceptive, discrepant, tense) (1), speaker is very ambivalent (deceptive, discrepant, tense) (9). Discrepancy was defined as the simultaneous communication of several different emotions.

the person relatively more than liking him or her. Similarly, a more accurate decoder of leakage would see relatively more liking than disliking in the LD descriptions. Thus, LD accuracy was defined as a subject's mean liking rating of the LD descriptions minus his or her mean disliking rating of those descriptions; for DL, liking was subtracted from disliking. These scores were also computed separately for judgments of male and female speakers.

Leakage accuracy for LD
= liking ratings of LD
− disliking ratings of LD;

Leakage accuracy for DL
= disliking ratings of DL
− liking ratings of DL.

Encoding leakage (or betraying one's own true feelings) was defined as follows: for LD, the mean of judges' disliking ratings was subtracted from the mean of their liking ratings; for DL, liking was subtracted from disliking. Subscores were computed for judgments made by male and female judges.

Leakage accuracy for LD
= rated liking on LD
− rated disliking on LD;

Leakage accuracy for DL
= rated disliking on DL
− rated liking on DL.

Hamming

A person who hams in the DL condition expresses greater liking of the truly disliked person (whom he or she is pretending to like) than of the truly liked person. Thus, ham like was defined as judges' mean liking rating of a speaker's DL description minus their mean liking rating of that speaker's true or pure *like* description. Analogously, "ham dislike" was defined as judges' mean dislike rating of a speaker's LD description minus their mean *dislike* rating of that speaker's true or pure *dislike* description.

Ham like = rated liking on DL − rated liking on L;

Ham dislike = rated disliking on LD
− rated disliking on D.

Results

Accuracy

To examine the relative accuracy of communication of different kinds of lies (leakage, deception, LD, DL), a 2 (Leakage/Deception) × (LD/DL) analysis of variance (ANOVA) was computed, with repeated measures on both factors and accuracy scores as the dependent variable.

The main effect of leakage versus deception, $F(1, 36) = 132.04$, $p < .001$; $d = 3.83$,[2] showed that accuracy at recognizing when deception was occurring (deception accuracy) was substantially greater than accuracy at reading a speaker's underlying affect (leakage accuracy). Or, from the point of view of the speaker, the tendency to have one's attempts at deception accurately diagnosed was significantly greater than the tendency to have one's true underlying affect correctly identified. The main effect for LD versus DL and the interaction of LD/DL × Leakage/Deception were both very small, showing that accuracy at communicating L did not differ significantly from accuracy at communicating DL, and that the magnitude of the accuracy difference between leakage and deception did not vary significantly according to the particular affect that was feigned or concealed.

The overall level of deception accuracy differed significantly from zero, which is the value that would be expected under the null hypothesis of no accuracy, $t(38) = 5.06$, $p < .001$; $d = 1.64$. Thus, observers were significantly more accurate than chance at detecting when deception was occurring. Accuracy at reading the underlying affect was significantly *worse* than chance, however, $t(38) = 6.04$, $p < .001$; $d = 1.96$. This means that on the whole, observers tended to read the affect that speakers *intended* to communicate rather than the affect that speakers really felt.

Telling Different Kinds of Lies

The correlation between skill at detecting deception and skill at reading leakage was .05 for LD and −.11 for DL.[3] Both of these correlations were far from significant. Thus people who are especially good at knowing

---

[2] The index d is an estimate of the size of the effect expressed in standard deviation units (Cohen, 1977). As a rule of thumb, Cohen suggests regarding ds of .20, .50, and .80 as small, medium, and large effects, respectively.

[3] ns varied from 38 to 40 because there was one missing leakage subscore and one missing deception subscore.

when deception is occurring are not necessarily particularly skilled at knowing the speaker's true affective state.

In terms of encoding, however, the relationship between leakage and deception is strongly positive (for LD, Pearson's $r = .50$, $p < .01$; for DL, $r = .54$, $p < .01$). Thus people whose deception attempts are readily detected tend also to leak their true underlying feelings.

The ability to recognize deception when the speaker is hiding positive affect (LD) is not significantly related to the ability to recognize deception when the speaker is hiding negative affect (DL; $r = -.18$). However, the ability to read a leaked positive cue is very strongly related to the ability to spot a leaked negative cue ($r = .75$, $p < .001$). For encoding, the tendency to get caught telling an LD lie is correlated only .13 (*ns*) with the tendency to get caught telling a DL lie. The analogous leakage correlation is similarly small and nonsignificant (.19), showing that people who leak their underlying positive feelings don't necessarily leak their underlying negative feelings.

Different Kinds of Liars and Lie Detectors

Speakers who get caught lying by women also tend to get caught lying by men (for LD, $r = .62$, $p < .001$; for DL, $r = .54$, $p < .001$), and speakers whose leaked cues are picked up by women also have their leaked cues picked up by men (for LD, $r = .76$, $p < .001$; for DL, $r = .77$, $p < .001$). Similarly, observers who are skilled at reading the leakages of women also tend to be skilled at reading the leakages of men (for LD, $r = .72$, $p < .001$; for DL, $r = .78$, $p < .001$). However, skill at knowing when women are lying is *not* significantly correlated with skill at knowing when men are lying (for LD, $r = .06$; for DL, $r = .27$).

Skill at Telling Lies and Skill at Telling Lies: Are They Related?

The four different kinds of ability to tell (detect) lies (detecting leaked liking and leaked disliking; detecting LD and DL lies) were correlated with the four different kinds of ability to tell (get away with) lies. The 16 correlations ranged from .20 to −.24, and all were statistically insignificant. The median correlation was exactly zero. Thus, the ability to get away with one's own lies seems to be completely unrelated to the ability to catch other people in their lies.

Hamming

Only 13 of the 40 subjects earned *ham like* scores greater than zero; only 10 earned *ham dislike* scores greater than zero. Thus, the tendency to express more liking (or disliking) when lying than when really liking (or disliking) was not particularly common in this sample. Further, *ham like* scores were not significantly correlated with *ham dislike* scores ($r = .24$), suggesting that there are probably at least two different kinds of hams. Finally, hams who exaggerate feigned liking and hams who exaggerate feigned disliking are both tremendously successful at their deception attempts (between *ham like* and encoding deception, or getting caught in one's attempts to lie, $r = -.60$, $p < .001$; the r for ham dislike was identical.

Machiavellianism

A median split was calculated for Machiavellianism scores, and a Machiavellianism (High/Low) × Type of Affect (LD/DL) × Type of Lie (Leakage/Deception) ANOVA was computed with encoding scores as the dependent variable. A main effect for Machiavellianism showed that high Machs were slightly more successful at getting away with their lies than were low Machs, $F(1, 36) = 3.04$, $p = .09$; $d = .58$. The interaction of Machiavellianism with type of affect showed that high Machs were especially successful at deceiving when pretending to dislike someone they really liked, $F(1, 38) = 5.96$, $p < .05$; $d = .79$. The interaction of Machiavellianism with type of lie was not significant.

An analogous 2 × 2 × 2 ANOVA was computed using decoding scores as the dependent variable. The main effect for Machiavellianism was very small, and the interactions of type of

lie and type of affect with Machiavellianism were also nonsignificant.

In this sample, Machs also tended to be hams—significantly so for *ham like* ($r = .31$, $p = .055$), though only weakly for *ham dislike* ($r = .14$, *ns*).

Discussion

Consistent with at least ten previous studies of human lie detection (Fay & Middleton, 1941; Feldman, 1976; Harrison et al., 1978; Kraut, 1978; Littlepage & Pineault, 1978; Maier, 1966; Maier & Janzen, 1967; Maier & Thurber, 1968; Matarazzo et al., 1970; Zuckerman et al., 1979), we found that observers were able to identify accurately the occurrence of deception. However, we also found that observers were markedly unable to see through to a speaker's underlying affect when that affect was covered over by some other dissimulated display. Feldman (1976), too, in a study that examined both leakage and deception skill, found that observers could pick up the deception cues but not the underlying affect. Perhaps a convergent finding is the near-zero correlation between identifying deception and reading leakage. Apparently, an observer can have a keen sense of when "something fishy is going on," without necessarily being able to differentiate the feigned feeling from the true affect.

Except for the very small relationship between skill at hiding liking (LD) and skill at hiding disliking (DL) ($M\ r = .16$), the ability to lie successfully appears to be a fairly general one. People who effectively mask their deception attempts also effectively conceal their true affective states ($M\ r = .52$). Further, people who are especially adept at fooling women (whether in terms of leakage or deception) also tend to be particularly skilled at deceiving men ($M\ r = .67$).

There was also a considerable degree of consistency in the ability to detect different kinds of leakages. People who were especially skilled at noticing an underlying negative affect were also particularly adept at discerning an underlying positive affect. Similarly, people who were especially likely to read the leakages of women were also especially likely to read the leakages of men ($M\ r = .75$).

The tendency to read leaked cues, however, is not significantly related to the ability to recognize the occurrence of deception. Further, the ability to detect deception appears to be a much more disparate set of skills than is the ability to perceive leaked affects. People who can tell that a lie is occurring when liking is being concealed cannot necessarily detect a lie when disliking is hidden. Further, observers who know when women are lying do not necessarily recognize dissimulation by men. Using a "To Tell the Truth" format, Littlepage and Pineault (1978) also found a near-zero correlation between the ability to point to the impostrous male panelists and the ability to identify the impostrous women. Perhaps observers operate on the assumption that there is a single set of clues that tip people off to deception, when in fact women's deceit is revealed in ways different from men's. There is already some evidence suggesting that people's lay theories about clues to deception are not particularly accurate (Ekman & Friesen, 1974; Kraut, 1978; Maier & Janzen, 1967; Krauss et al., Note 1). The second part of the hypothesis, about sex differences in clues to deception, is still essentially an unanswered question. With the exception of a study by McClintock and Hunt (1975) and one by Mehrabian (1971, Experiment 2), most studies of deception cues tend to use subjects of only one sex or to exclude sex as a factor in the analyses (Ekman & Friesen, 1972; Ekman, Friesen, & Scherer, 1976; Harrison et al., 1978; Knapp, Hart, & Dennis, 1974; Kraut, 1978; Streeter, Krauss, Geller, Olson, & Apple, 1977). Both Mehrabian (1971) and McClintock and Hunt (1975) present evidence suggesting that there may in fact be sex differences in at least some of the behaviors emitted during deception.

The relative homogeneity of encoding skills compared to decoding skills is consistent with Kraut's conclusion that people are consistently successful or unsuccessful as liars but are not so consistent as lie detectors. The exception in our data to the general consistency of skills in encoding lies was the very weak relationship between lying successfully about underlying positive feelings (LD) and lying successfully about underlying negative feel-

ings (DL). The importance of the particular kind of affect that is being hidden or dissimulated was also apparent in (a) the weak correlation between detecting LD lies and detecting DL lies; (b) the nonsignificant correlation between exaggerating (hamming) feigned liking and exaggerating feigned disliking; and (c) the interaction between Machiavellianism and the type of affect being encoded. This suggests that conclusions drawn from studies involving just one variety of hidden affect or one kind of "sent" cue may in some cases be of limited generality.

The ability to lie effectively to others was entirely unrelated to the ability to detect the lies of others. Zuckerman et al. (in press) reported a similarly low and insignificant correlation (.20) between encoding and decoding deception. These results are consistent with the results of 17 studies of pure (nondeceptive) nonverbal communications: for those studies, the median correlation between encoding skill and decoding skill is .16 (DePaulo & Rosenthal, in press). The ability to express one's emotions accurately appears to be quite distinct from the ability to interpret the emotions of others accurately, whether those emotions are real or dissimulated.

The Machiavellian subjects in our study, like those in the Krauss et al. (Note 1) study, were not especially skilled at detecting deception; like the Machs in the study by Exline et al. (1970), however, they were fairly successful as liars. We also found that Machs seem to favor an especially theatrical style of deceiving, particularly when feigning positive regard: they, more than their less Machiavellian counterparts, tend to be hams when they lie. Moreover, the hamming strategy appears to work quite well: people who tend to exaggerate sentiments of liking that they do not really feel, as well as people who exaggerate feigned feelings of disliking, are much less likely to get caught in their lies than are the less histrionic sorts. The future of the study of hamming looks promising. Already there is evidence suggesting both sex differences (Rosenthal & DePaulo, in press) and developmental differences (Feldman, Jenkins, & Popolla, in press) in this style of lying.

Methodological Considerations

The present study points to the advantages of distinguishing deception from leakage, of sampling a wide range of encodings, and of using designs that completely cross the affects that are actually experienced with those that are dissimulated or "sent." These design features should help to separate deceptive messages from discrepant but nondeceptive ones and to distinguish deception from stress or from unpleasant affects.

In our study, the lie detectors were the same persons who had previously served as liars; hence, they might have been predisposed to use particular kinds of cues or strategies. Moreover, in our study as in earlier investigations (e.g., Streeter et al., 1977), judges knew exactly what proportion of the time the senders would be lying. Research paradigms in which the judges are not given this information—and, furthermore, are not specifically informed about the content of the lie—might be more realistic. Finally, comparisons of posed deceit with spontaneous lies will shed light on the generalizability of the kinds of deceptions we most often stage in our laboratories, as will a more extensive sampling of deceptions in naturalistic settings.

Future Research in Deception: Toward a More Programmatic Approach

Studies of the abilities to detect lies and to deceive successfully comprise just one area in the psychological literature on deception. (We concern ourselves here only with lie detection that is unaided by physiological measurements or detection devices; for a more physiologically oriented review, see Lykken, 1974, 1979, Orne, Thackray, & Paskewitz, 1972, or Podlesny & Raskin, 1977.) A second line of inquiry looks at accuracy of lie detection as a function of differential access to different channels or modalities such as the face, the body, the tone of voice, the words, and various combinations of these cues (e.g., Ekman et al., 1976; Ekman & Friesen, 1969, 1974; Feldman, 1976; Littlepage & Pineault, 1978; Streeter et al., 1977; Zuckerman et al., in press; Krauss et al., Note 1). The appeal of these studies probably derives in part from

the suggestion that certain channels, such as the face, that under ordinary circumstances are extremely informative, can be especially misleading under conditions of deception. These studies demonstrate the kinds of information that *can* be gleaned from a particular channel when access is restricted to that channel; a further step would be to demonstrate whether this same information is in fact extracted when information from other channels is available, too (cf. Streeter et al., 1977). If it can be established that particular channels are especially misleading under conditions of deception, it might then be asked whether people adjust their attentional strategies in ways that seem to acknowledge these modality effects (DePaulo, Rosenthal, Eisenstat, Rogers, & Finkelstein, 1978), whether particularly skilled lie detectors are especially likely to make these adjustments, and whether attentional instructions or training procedures can increase accuracy of detection.

A third approach has concerned itself with identifying particular cues, such as smiling, speech errors, and pitch, which distinguish deceptive from nondeceptive responses (e.g., Ekman et al., 1976; Ekman & Friesen, 1972; Harrison et al., 1978; Knapp et al., 1974; Kraut, 1978; Luria, 1932; McClintock & Hunt, 1975; Mehrabian, 1971; Streeter et al., 1977; Alker, Note 2). This line of research might be advanced by a consideration of the relationships among (a) the cues that actually do distinguish truth from deception; (b) the cues that people say they use to distinguish truth from deception; and (c) the cues that people actually do utilize in their judgments (see Baskett & Freedle, 1974; Kraut, 1978).

Studies of veridical clues to deception, when they do examine comparable clues, sometimes yield disappointingly inconsistent results. The question "What are the clues to deception?" is probably too broad; instead, we may need to ask, "What are the clues to which kinds of deceit by which kinds of deceivers to which kind of receivers in which kinds of situations?"

A Matrix for Generating and Organizing Research on Deception

We can start, then, by sketching a taxonomy of lies, liars, lie detectors, and social settings. Then, by locating specific studies of clues to deception within this matrix of lie types, setting types, deceiver types, and detector types, perhaps higher order consistencies will emerge. This four-dimensional approach should be equally relevant to the other two lines of inquiry. Thus, for example, we might want to know whether certain channels are especially misleading when communicating certain kinds of lies, or whether people are especially effective liars or lie detectors in same-sex or same-status dyads.

Among the many kinds of lies are ordinary white lies, proffered in the service of smoothing social interaction, lies of self-aggrandizement for purposes of impression management or personal gain, and lies in which the deceived party is the intended beneficiary, as in physicians' careful cloaking of grim prognoses or parents' imaginative fabrications about sex and death.

Lies may involve facts that may or may not be verifiable by external criteria (as in police investigations), or they may involve attitudes or opinions or feelings about oneself or others. Lies vary also in the degree of guilt or stress that they cause for the deceiver, the degree of involvement that they engender in the topic of the lie, the consequences of getting caught in the lie, and whether the lies were planned and rehearsed or unpremeditated.

An important dimension of social settings is the degree of normative or situational support that they offer to the deceiver. Those who lie to parole officers (especially in "tough" neighborhoods) or to captors in a military setting probably find their deceit more readily justifiable than those who lie to priests, doctors, and close friends. Structural characteristics of contexts, such as the degree of formality or informality or of democracy or autocracy, may also importantly affect the cues used to detect deceit, the modalities used to convey deceit, and the overall ease with which lies are discovered or perpetuated. Still another important aspect of social settings is the people in those settings—the number of participants, their structural relationship to the deceiver and the deceived (e.g., whether they are outside observers or are directly involved in the interaction); their affective

relationship (whether they are more closely allied with the deceiver or with the deceived); and their expectations about the probability of occurrence of deception.

Relevant characteristics of both the deceiver and the detector include demographic characteristics (e.g., age, sex, race, socioeconomic status); personality characteristics (e.g., adjustment, anxiety, Machiavellianism); social competence (e.g., nonverbal sensitivity); and cognitive competence (e.g., intelligence). The liar's and the lie detector's appraisal of success in deception (i.e., whether they view effective liars as clever and well adjusted or as immoral and psychopathic) might also predict the outcomes of deceptive transactions. These effects are probably best studied as two-way interactions in designs that cross characteristics of the deceiver with analogous characteristics of the detector.

Trait versus state aspects of lying might also mediate deception effects—that is, the deceiver may be an inveterate liar or an occasional liar who at the present moment is or is not telling the truth; similarly, the lie detector may be a chronic paranoid or a naive and unsuspecting soul who at the present moment is or is not suspicious of the deceiver's intent. The relationship of the deceiver to the deceived—spatially; temporally (is this a one-time meeting or will they see each other again in the future?); and personally (are they strangers, friends, intimates, or enemies? what is their relative status and power?)—should also be considered in future research efforts.

Reference Notes

1. Krauss, R. M., Geller, V., & Olson, C. *Modalities and cues in the detection of deception.* Paper presented at the meeting of the American Psychological Association, Washington, D.C., September 1976.
2. Alker, H. A. *Mystification and deception in presidential press conferences.* Paper presented at the meeting of the American Psychological Association, Washington, D.C., September 1976.

References

Baskett, G. D., & Freedle, R. O. Aspects of language pragmatics and the social perception of lying. *Journal of Psycholinguistic Research,* 1974, *3,* 117–131.

Christie, R., & Geis, F. (Eds.), *Studies in machiavellianism.* New York: Academic Press, 1970.

Cohen, J. *Statistical power analysis for the behavioral sciences* (rev. ed.). New York: Academic Press, 1977.

DePaulo, B. M., & Rosenthal, R. Ambivalence, discrepancy, and deception in nonverbal communication. In R. Rosenthal (Ed.), *Skill in nonverbal communication.* Cambridge, Mass.: Oelgeschlager, Gunn & Hain, in press.

DePaulo, B. M., Rosenthal, R., Eisenstat, R. A., Rogers, P. L., & Finkelstein, S. Decoding discrepant nonverbal cues. *Journal of Personality and Social Psychology,* 1978, *36,* 313–323.

DePaulo, B. M., Zuckerman, M., & Rosenthal, R. Humans as lie detectors. *Journal of Communication,* in press.

Ekman, P., & Friesen, W. V. Nonverbal leakage and clues to deception. *Psychiatry,* 1969, *32,* 88–106.

Ekman, P., & Friesen, W. V. Hand movements. *Journal of Communication,* 1972, *22,* 353–374.

Ekman, P., & Friesen, W. V. Detecting deception from the body or face. *Journal of Personality and Social Psychology,* 1974, *29,* 288–298.

Ekman, P., Friesen, W. V., & Scherer, K. R. Body movement and voice pitch in deceptive interaction. *Semiotica,* 1976, *16,* 23–27.

Exline, R., Thibaut, J., Hickey, C., & Gumpert, P. Visual interaction in relation to machiavellianism and an unethical act. In R. Christie and F. Geis (Eds.), *Studies in machiavellianism.* New York: Academic Press, 1970.

Fay, P. J., & Middleton, W. C. The ability to judge truth-telling or lying from the voice as transmitted over a public address system. *Journal of General Psychology,* 1941, *24,* 211–215.

Feldman, R. S. Nonverbal disclosure of teacher deception and interpersonal affect. *Journal of Educational Psychology,* 1976, *68,* 807–816.

Feldman, R. S., Jenkins, L., & Popolla, O. Detection of deception in adults and children via facial expressions. *Child Development,* in press.

Harrison, A. A., Hwalek, M., Raney, D. F., & Fritz, J. G. Cues to deception in an interview situation. *Social Psychology,* 1978, *41,* 156–161.

Knapp, M. L., Hart, R. P., & Dennis, H. S. An exploration of deception as a communication construct. *Human Communication Research,* 1974, *1,* 15–29.

Kraut, R. E. Verbal and nonverbal cues in the perception of lying. *Journal of Personality and Social Psychology,* 1978, *36,* 380–391.

Littlepage, G., & Pineault, T. Verbal, facial, and paralinguistic cues to the detection of truth and lying. *Personality and Social Psychology Bulletin,* 1978, *4,* 461–464.

Luria, A. R. *The nature of human conflicts.* New York: Liveright, 1932.

Lykken, D. T. Psychology and the lie detector industry. *American Psychologist,* 1974, *29,* 725–739.

Lykken, D. T. The detection of deception. *Psychological Bulletin,* 1979, *86,* 47–53.

Maier, N. R. F. Sensitivity to attempts at deception

in an interview situation. *Personnel Psychology*, 1966, *19*, 55–65.

Maier, N. R. F., & Janzen, J. C. The reliability of reasons used in judgments of honesty and dishonesty. *Perceptual and Motor Skills*, 1967, *25*, 141–151.

Maier, N. R. F., & Thurber, J. A. Accuracy of judgments of deception when an interview is watched, heard, and read. *Personnel Psychology*, 1968, *21*, 23–30.

Matarazzo, J., Wiens, A., Jackson, R., & Manaugh, T. Interviewer speech behavior under conditions of endogenously present and exogenously-induced motivational states. *Journal of Clinical Psychology*, 1970, *26*, 141–148.

McClintock, C. C., & Hunt, R. G. Nonverbal indicators of affect and deception in an interview setting. *Journal of Applied Social Psychology*, 1975, *5*, 54–67.

Mehrabian, A. Nonverbal betrayal of feeling. *Journal of Experimental Research in Personality*, 1971, *5*, 64–73.

Orne, M. T., Thackray, R. I., & Paskewitz, D. A. On the detection of deception. In N. S. Greenfield & R. A. Sternbach (Eds.), *Handbook of psychophysiology*. New York: Holt, Rinehart & Winston, 1972.

Podlesny, J. A., & Raskin, D. C. Physiological measures and the detection of deception. *Psychological Bulletin*, 1977, *84*, 782–799.

Rosenthal, R. & DePaulo, B. M. Sex differences in accommodation in nonverbal communication. In R. Rosenthal (Ed.), *Skill in nonverbal communication*. Cambridge, Mass.: Oelgeschlager, Gunn & Hain, in press.

Streeter, L. A., Krauss, R. M., Geller, V., Olson, C., & Apple, W. Pitch changes during attempted deception. *Journal of Personality and Social Psychology*, 1977, *35*, 345–350.

Zuckerman, M., DeFrank, R. S., Hall, J. A., Larrance, D. T., & Rosenthal, R. Facial and vocal cues of deception and truth. *Journal of Experimental Social Psychology*, 1979, *15*, 378–396.

Received August 20, 1978 ■

New JPSP Format Delayed Until April

Recent issues of this journal carried an announcement that the *Journal of Personality and Social Psychology* would appear in January 1980 under new policies and in a new sectioned format. However, in order to publish the backlog of manuscripts accepted under present policies, the first three issues of 1980 will continue to carry articles accepted under Clyde Hendrick's Acting Editorship.

This journal will appear in April as a sectioned journal edited by Melvin Manis, Ivan D. Steiner, and Robert Hogan, as previously announced.

Robert S. Daniel, Chair, APA
Publications and Communications Board

Journal of Personality and Social Psychology
1979, Vol. 37, No. 10, 1723-1731

Attributional Information Processing: A Response Time Model of Causal Subtraction

Eliot R. Smith
Department of Sociology
University of California, Riverside

Frederick D. Miller
New York University

Response time measures have been used occasionally in social psychology, but rarely as direct probes of information processing. A study collecting response time data in a near-exact replication of McArthur's classic attribution study sheds light on the information processing involved in subjects' responses. The process is analyzed into two stages: (a) encoding or comprehension of the stimulus sentence and the consensus, distinctiveness, and consistency information and (b) attributional processing per se. In the second stage, response time analyses suggest that perceivers operate by *subtracting* causes from an initial set to arrive at a response, rather than by adding causal components (person, stimulus, and circumstances) until an adequate cause is obtained. Subtraction is theoretically related to the salience model of attribution. Response time measures promise to expand greatly the ability of social psychologists to build process models of causal attribution and other kinds of social perception and cognition.

Researchers on causal attribution have expressed increased interest recently in the formulation of models of attributional processing, as distinct from the prediction of attributional judgments or outcomes (Pryor & Kriss, 1977). A model of attributional information processing permits the prediction of social cognition and behavior in a range of new situations (assuming that the same process will continue to operate), whereas studies that just measure attributional outcomes are hampered by the fact that any particular outcome could be generated by a large number of different processes, with different implications for new situations (Anderson, 1976, p. 11). In addition, information-processing models of attribution may find a place within the general models of human cognitive functioning that are currently being developed in cognitive psychology (e.g., Anderson, 1976; Schank, 1975), leading us closer to a fruitful union of cognitive and social psychology (Simon, 1976; S. Taylor, 1976).

Response time (RT) measures, a frequently used tool for the development of cognitive process models, are just beginning to be used in social psychology. Pryor and Kriss (1977) used response times to validate a manipulation of salience or availability: they demonstrated that the first noun presented in a sentence (whether subject or object) was more quickly recognized in a response time test than the second noun presented. They concluded that the first noun was more available. Markus (1977) used response times to demonstrate that people with "self-schemata" relevant to a given adjective identified those adjectives more quickly than did subjects without such self-schemata. However, social psychologists have not yet used response time methods to probe processing directly (cf. Posner, 1978), as cognitive psychologists have used them (e.g., Carpenter & Just, 1975; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977).

Two of the ways in which response time measures can probe processing will be con-

This research was supported by the University of California Academic Senate and by a New York University Challenge Grant.

Requests for reprints should be sent to Eliot R. Smith, Department of Sociology, University of California, Riverside, California 92521.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3710-1723$00.75

sidered here (for a review, see D. Taylor, 1976). First, the time for a response may vary depending upon experimentally manipulated factors. The studies by Carpenter and Just (1975) on verification of affirmative and negative sentences illustrate this possibility. Negative sentences take longer to verify, and this can be interpreted to indicate that they are transformed to affirmative before the verification process begins. The second possibility is that a relationship may exist between the content of a response and the time needed for the response. For example, the time for "not match" judgments about the identity of complex stimuli is greater than for "match" judgments in the data of Trabasso, Rollins, and Shaughnessy (1971). This is used as evidence for a serial feature-matching process, in which the features of the stimuli are compared for identity one at a time, with a mismatch of features leading to several time-consuming processes such as preparing a "not match" response and rechecking features.

The current study uses these two types of analysis to examine the processes involved in causal attributions in a replication of the now-classic McArthur (1972) attribution study. McArthur presented her subjects with sentences like "Sue is afraid of the dog," followed by three types of additional information. High (low) consensus is information about the actions of other people: "Almost everybody (nobody) else is afraid of the dog." High (low) distinctiveness concerns this person's responses to other objects: "Sue is not afraid of almost any other dog (is afraid of almost every other dog)." High (low) consistency describes the past history of the response: "In the past Sue has almost always (never) been afraid of the dog." The three types of information are manipulated in a $2 \times 2 \times 2$ design, and subjects' responses as to what caused the event to occur are obtained. McArthur's dependent measure offers subjects a choice of four causal possibilities:

1. Something about the *person* (e.g., Sue) probably caused her to make response X (e.g., be afraid) to stimulus X (e.g., the dog).

2. Something about *stimulus X* probably caused the person to make response X to it.

3. Something about the *particular circumstances* probably caused the person to make response X to stimulus X.

4. Some combination of 1, 2, and 3 above probably caused the person to make response X to stimulus X. (The subject is asked to check either $1 + 2$, $1 + 3$, $2 + 3$, or $1 + 2 + 3$; McArthur, 1972, p. 175.) McArthur's study was originally designed to test Kelley's (1967) analysis of variance (ANOVA) model of attribution and has been extremely influential in the attribution research tradition.

The addition of a response time measure to this study will allow two types of analysis. The first purpose of this study is to investigate the influence of consensus, distinctiveness, and consistency information on the time taken for attributional processing. It may be hypothesized that certain patterns of information will be easier to encode and process than others. For example, one prediction is that certain "standard" patterns of information (Orvis, Cunningham, & Kelley, 1975, suggest the three patterns high/high/high consensus/distinctiveness/consistency, low/low/high, and low/high/low) are consistent with perceivers' expectations and should be more quickly processed than other patterns. In general, relationships between the manipulated information factors and RTs give insight into subjects' encoding and processing of the information.

The second aim of this study is to examine the relations between the content of causal judgments and the time taken to arrive at them in order to clarify the procedure the attributor uses. Several procedures are possible. One is a "look-up" process, in which the perceiver refers to memory for a particular attributional response that is stored with each possible configuration of information (or at least with some of the possible patterns), much as one looks up a name in a phone book to locate a number (the response). That is, one may have stored the fact that low/low/high information points to the person as cause; that high/high/high implies stimulus; that low/high/low implies circumstances, and so forth. The process of looking up each information configuration should take about the same amount of time, yielding a prediction of no relation between

mean RT and the content of the response. Alternatively, one could predict (following Orvis et al., 1975, p. 606) that perceivers may have stored *only* these three "standard data patterns," so that responses of person, stimulus, or circumstance attributions to those patterns are quick, whereas other responses are slower, since the perceiver must engage in further processes (perhaps comparing the observed information pattern to more than one of the standard patterns to arrive at a compromise response).

A second type of process is additive. Here the perceiver begins by considering a single causal factor (most likely the person) and adds other factors to it (a process that takes time) when and if the available information indicates that they are required. That is, if a quick check shows that the current response is adequate, it is given. On the other hand, if a mismatch appears between the tentative causal response and the given information, another component must be added to the tentative response (e.g., changing from person to person plus circumstances), and a response change has been found to require processing time (cf. Newell, 1973, p. 490; Trabasso et al., 1971). It may also be true, as Trabasso et al. (1971) found, that the detection of a mismatch between response and information requires more time than detection of a match. The overall prediction is thus that single-component responses (or at least the person response) should be quick and multi-component responses should be slower; ideally, this model would give a *positive,* linear relationship between the number of components in the response and the response time, if detecting a mismatch and adding any component to the response takes the same increment of time. This sort of model is suggested by the way in which psychologists tend to think of attributional responses. Examples are the treatment by Orvis et al. (1975) of information patterns for the three single-component responses as "standard," and Kassin and Hochreich's (1977) labeling of single-factor responses as "simple" and multi-factor responses as "complex."

A third suggested process is the inverse of the second. Instead of beginning with a single component as a tentative response and engaging in processing that may add components to the response, the perceiver might begin with a three-component tentative response and engage in processing that may delete or subtract components. As in the second model, the subject is presumed to check the tentative response against the given information; again, a match is presumed to lead quickly to the emission of that already prepared response. A mismatch leads to processing, the effect of which is to delete a causal component from the response. As in the second model, the changing of the tentative response is presumed to be time-consuming. The overall prediction is that the three-component response of person plus stimulus plus circumstances should be the quickest and single-component responses the slowest, as the latter involve the most deletions (two). Ideally, one would obtain a *negative,* linear relationship between the number of causal components in the response and the response time, indicating that the mismatch detection and subtraction of any component from the response takes about the same amount of time. The predictions of these three models are summarized in Table 1.

Thus this study will investigate the relationship between the reported attribution and the time taken to report that attribution to enable a choice among the three processes suggested above. From McArthur's (1972) study and many other investigations we know the content of the causal attributions that subjects produce for each information combination. However, each of the three processes outlined above could generate that same set of outcomes (cf. Anderson, 1976, p. 11), and thus the entire body of prior research gives inconclusive evidence as to the nature of the process, the issue considered here.

Method

The procedure of this study generally followed that of McArthur (1972), which should be consulted for further detail. To permit the recording of RTs, the study was conducted with the aid of a computer. The stimulus materials (sentences) were displayed on a televisionlike cathode-ray tube screen and subjects reported their attributions by pressing response keys. Timing started when the

Table 1
Response Time Predictions for Three Models

Model	Response		
	1-component	2-component	3-component
Look up	←Same time for any response→		
(Orvis et al.[a] variant)	Short	Moderate to long→	
Additive	Short	Moderate	Long
Subtractive	Long	Moderate	Short

[a] Orvis, Cunningham, & Kelley, 1975.

sentence appeared and ended when a key was pressed; the times thus include reading and comprehension time as well as attributional processing. The computer controlled the generation and presentation of stimuli, recorded the responses and RTs, and output the data for analysis.

Materials

The stimulus materials are described above and in McArthur (1972). The 32 sentences used fall into four categories based on the type of verb: action, accomplishment, opinion, and emotion. Like McArthur, we will be concerned primarily with the distinction between the first two categories versus that between the third and fourth (manifest vs. latent verbs).

As in the McArthur study, consensus, distinctiveness, and consistency information were manipulated by additional sentences that followed the sentence describing the event that was the object for attributions. These sentences were always presented in the same order for two reasons: (a) to replicate McArthur's study, her procedure was followed in this regard, and (b) the different weights that the three items of information have when their order is randomized (Ruble & Feldman, 1976) are not of central concern here. We are chiefly interested in the response times rather than in the responses' content, as long as the responses indicate that a successful replication of McArthur's results has been achieved.

The question to which subjects responded was that used by McArthur: "Your task is to decide, on the basis of the information given, what probably caused the event to occur" (1972, p. 174). McArthur's response format was slightly modified in this study in order to treat all responses more symmetrically and to avoid the second level of response necessitated when "combination" was chosen in the original format. Just three response keys were used. They represented the person, stimulus, and circumstances causal factors, and subjects were permitted to respond with just one, any two, or all three of the keys to indicate combination responses. Subjects in pilot testing indicated that this response format posed no difficulty, so it was adopted for the data collection. Subjects began each trial with three fingers of the preferred hand resting on the three keys. Timing stopped when the computer sensed the first key press; if any further key presses took place within one-half sec, they were recorded as part of a multikey combination response (the keys of a combination response were nearly always pressed simultaneously by subjects). The assignment of the labels *person*, *stimulus*, and *circumstances* to the three keys was counterbalanced between subjects.

Procedure

The 24 subjects, paid undergraduate volunteers, were run individually. After subjects had been instructed in the use of the response keys and given practice on 8 practice sentences of the type used in the study, the actual trials began. On each trial, the subjects read the sentence and additional consensus, distinctiveness, and consistency information; responded with an attribution by pressing a key or keys; and then signaled when they were ready for the next trial. The intertrial interval was not timed. Trials were administered in two blocks of 16 each, with a 3–5 min rest between blocks. The 32 stimulus sentences were given to each subject in a different random order; thus no systematic order effects can exist in the data.

Analyses

The results were analyzed in what is effectively a Verb Type (Manifest vs. Latent) × Consensus × Distinctiveness × Consistency (2 × 2 × 2 × 2) analysis of variance (ANOVA), with the appropriate interactions with subject and with sentence forming the components of the denominator mean square for quasi-F ratios. Such tests confirm that an effect is reliable not only with a new sample of subjects (using the same experimental materials, e.g., sentences) but with a new sample of subjects and a new sample of materials as well. The routine use of simple F statistics with only subject as an error term provides no evidence for the replicability of effects beyond the specific set of experimental materials used, a practice Clark (1973) criticizes for language research. Kenny and Smith (Note 1) de-

scribe the technical aspects of this analysis more fully, including the procedure for assigning information conditions to sentences for each subject and for recovering them in the analysis.

Analyses were performed for several dependent variables, including (a) response times; (b) the response categories McArthur used, which were person, stimulus, circumstances, person plus stimulus, and "other" attributions (all other types combined); and (c) the person, stimulus, and circumstances components of the response. The components (i.e., the responses on the three keys treated as three separate dependent variables) have the advantage that they are statistically independent, which is not the case with McArthur's categories, where each response has to fall into exactly one category. Furthermore the analysis of components may more accurately represent the content of subjects' responses. For example, a response of person plus circumstances will be counted here as (both) a response involving person and a response involving circumstances, whereas the original category system would treat such a response with "others" as unrelated to either person or circumstances.

Logs of RTs were analyzed because a slope of nearly 1 related the raw RT cell means to the cell standard deviations. Winer (1970, p. 400) suggests using a log transformation in such cases, and it practically succeeded in eliminating the cell mean–standard deviation relationship.

Results

Response Times

The analyses of variance by information condition using log response time as the dependent variable yielded only one significant effect, a distinctiveness–consistency interaction. Under high distinctiveness – high consistency and low distinctiveness – low consistency, responses were significantly slower ($M = 13.44$ sec) than under the other two combinations, $M = 12.32$ sec, quasi-$F(1, 7) = 18.9$, $p < .01$. The verb-type factor had no significant main effects or interactions affecting response time—thus manifest and latent verbs do not seem to differ in the amount of time required for attribution.

Relations Between Response and Response Time

Collapsing across information conditions, one can analyze the relationship between the response given by the subject and the RT for that response. This analysis should indicate the attributional procedure followed by the subject, as discussed above. Ignoring conditions in this analysis is appropriate, since subjects receive conditions in a random order and so do not know until the stimulus appears which condition it represents. Thus subjects are unable to apply different attributional procedures for different conditions. (In addition, if subjects did use different procedures for different conditions, one would expect to find only confused relationships between responses and RTs, as a result of averaging across different procedures.) The results of this analysis are shown in Tables 2 and 3. The data show a very clear negative, linear trend in log response time depending upon the number of components in the response. Subjects take the shortest time, approximately 11.07 sec (7.009 in log units) to make the "complex" person-plus-stimulus-plus-circumstances response, and the longest time, approximately 13.64 sec, for the "simple" single-factor responses. Each component that is subtracted adds approximately .097 in log units to the RT. Deviations from this linear trend are nonsignificant, so one would not gain by assuming that different components might take different amounts of time to subtract.

Table 2
Response Time by Response Analysis, Using Logs

Response	*M* response time	
	Log	Sec
Person	7.223	13.71
Stimulus	7.209	13.52
Circumstances	7.222	13.69
Person + stimulus	7.142	12.64
Person + circumstances	7.171	13.01
Stimulus + circumstances	7.116	12.32
Person + stimulus + circumstances	7.009	11.07

Attributions

The results with McArthur's (1972) dependent measures paralleled those of the earlier study nearly exactly, except that the analyses in this study have less power be-

Table 3
Analysis of Variance of Logs

Effect	Sum of squares	*df*	*MS*	*F*	*p*
Between responses	3.570	6	.595	3.57	.002
Linear trend in number of components	3.327	1	3.327	19.92	.001
Deviations from linear trend	.243	5	.049	<1	*ns*
Within groups	126.786	760	.167	—	

Note. Regression summary of linear trend: Log RT = 7.315 − (number of components) .097.

cause of their reliance on quasi-*F* ratios rather than simple *F*s. Across the 28 ANOVA effects in the two studies (4 Dependent Measure Categories × 7 Main Effects and interactions of the informational variables), 9 effects were significant in both studies; 10 effects were significant in McArthur's study and not here, and none was significant here and not in McArthur's investigation. All effects that were significant in both studies had mean differences in the same direction.[1]

Analyses of the person, stimulus, and circumstances components produced somewhat simpler results, as predicted. In contrast to the analysis of the McArthur dependent measures, no interaction effects were found with the components. Person attribution was increased by low consensus, low distinctiveness, and high consistency. Stimulus attribution was increased by high consensus. Circumstances attribution was increased by low consistency. Means and quasi-*F* tests appear in Tables 4 and 5.

Table 4
Means for Significant Effects of Informational Factors on Causal Components

Information/ level	Proportion of responses including: Person	Stimulus	Circumstances
Consensus			
High	.552	.685	*ns*
Low	.724	.471	
Distinctiveness			
High	.523	*ns*	*ns*
Low	.753		
Consistency			
High	.714	*ns*	.276
Low	.562		.654

Discussion

The near-exact replication of McArthur's results (when her measures are analyzed) in this study indicates that the changes in design (a different set of sentences and a new response format) did not materially affect the results of subjects' information processing. However, the addition of the response time dependent measure yields valuable information about the processes that underlie the attributional judgments made by subjects. The results will be discussed in the context of a proposed model of information processing containing two stages, the times for which are added to produce the total RT (cf. D. Taylor, 1976). The stages are *encoding* the presented information (stimulus sentence plus the three types of information) and *attribution processing*. The logical necessity for these two stages in a model for the current task is clear. The subject must begin by reading and understanding the stimulus sentences, transforming them into some internal representation in memory. This encoding stage must take time. Second, though some inferences about causal relations may take place during encoding (cf. Kintsch, 1974; Schank, 1975), the encoded information must still be used in some fashion to generate and check an appropriate response to the attributional question; three alternative processes for this stage were outlined above. Of course, other processing may occur as well—for example, the generation of the

[1] The results of these analyses of McArthur's dependent variables are not presented in full; they are relied upon simply to establish that a successful replication of McArthur's results has occurred. The analyses reported in the next paragraph and in Table 3 represent a simpler and more direct presentation of the attributional results than do the analyses of McArthur's dependent variables.

physical movements involved in making the response—but if their time is roughly constant across conditions, the conclusions drawn for the two-stage model are not affected (only the grand *M* RT will be changed). The current data cannot indicate whether the two stages are purely sequential. At least some encoding must precede attributional processing, but the two stages might overlap to some extent; our conclusions are unaffected by this variable. (For a comparable two-stage model for a concept verification task—comprehension and processing—see Trabasso et al., 1971.)

Table 5
Quasi-F *Tests for Significant Effects of Informational Factors on Causal Components*

Effect	*df*	*F*	*p* <
Consensus on person	1, 13	17.36	.001
Consensus on stimulus	1, 11	23.03	.001
Distinctiveness on person	1, 6	81.52	.001
Consistency on person	1, 9	20.64	.001
Consistency on circumstances	1, 13	47.68	.001

Note. All effects not shown are not significant at the .05 level. Denominator *df*s for a quasi-*F* test depend on the data and are not simply the *df* for a denominator *MS* in the analysis of variance table.

Encoding

The effects of the informational variables on RT can be interpreted as representing encoding or comprehension processes. In particular, since distinctiveness and consistency interact with this particular task, they must be presumed to influence the same stage (Sternberg, 1969; D. Taylor, 1976). The reasoning behind this conclusion is based on the fundamental assumption of RT models, that of additive stages. If two factors affected the times of *different* stages, their effects on the total RT would be additive (and thus noninteractive). Hence, factors that interact must influence the time of a single stage. Consensus does not interact with the other two information factors, but this need not cause its assignment to a different stage (D. Taylor, 1976, p. 190).

The distinctiveness–consistency information combinations high–low and low–high take less time to encode than do the other two combinations. Why is this? Subjects' expectations may provide an answer. In two studies Orvis et al. (1975) presented subjects with single information items (consensus, distinctiveness, or consistency alone) and asked the subjects to predict the levels of the other information factors. Subjects given high consistency predicted that distinctiveness would be low, in both of the studies by Orvis et al.; subjects given low consistency predicted that distinctiveness would be high in both studies. Similarly, high distinctiveness led to predictions of low consistency (in only one of the two studies, however), and low distinctiveness led to predictions of high consistency (in both studies). Thus, in seven of these eight comparisons the high–low and low–high combinations seem to go together or to be consistent with subjects' expectations. They are therefore partially redundant (carry less information) and so should be encoded more quickly. The other two combinations, high–high and low–low, do not have this redundancy; they tend to contradict subjects' expectations and should take more processing time to encode. The distinctiveness–consistency interaction effect on overall RT is thus explicable as a function of encoding processing.

Attribution

The second stage is the inference of a cause using the encoded forms of the presented information. The negative relationship obtained between the number of components in the response and the response time suggests that the process is subtractive. The subject appears to start by considering person plus stimulus plus circumstances as a cause. If that is found to be the appropriate response, it is given very quickly; otherwise, components are successively subtracted from that set as they are found to be unnecessary. Each detection of a mismatch between the tentative response and the information, followed by the subtraction of a component, takes about the same amount of time (.097 in the log scale). Thus, two-component causes all take somewhat more time to give than the three-component response, and single-component causes take longest of all.

The pattern of results weighs against the additive model of attributional processing, which predicts that single-component responses should be the quickest rather than the slowest. The information look-up model is also inconsistent with the results; this model postulates that the perceiver looks up the causal response appropriate for each possible configuration of information. This model predicts that RT should be not at all dependent upon the response given or else (in the variant of Orvis, Cunningham, & Kelley) that single-component responses should be quickest.

Subtraction and Salience

The subtraction model of attributional processing faces a problem that has not been discussed yet, namely the origin of the specific causal elements included in the initial tentative response. Other models, for example the additive model, do not face this problem: for them, a causal component (e.g., person or circumstances) to be added to the tentative response is suggested by a consideration of available information (e.g., consensus or consistency) about the event under scrutiny. But the subtractive model holds that people start with a list of causal elements *prior* to examining the relevant information. Where do these causal possibilities come from? The reason this issue has not been discussed to this point is that its solution in the context of the McArthur experiment is obvious: subjects use the causal possibilities suggested by the dependent measure. If the subtractive model is to describe attributional processing outside this context, though, some other mechanism must be involved.

In fact, the literature contains evidence for several possible mechanisms. One is salience. A model of attribution proposed by Pryor and Kriss (1977) and Taylor and Fiske (1978) holds that perceivers seize upon a single sufficient cause that is salient or available to them and report it as their attribution. The current model of causal subtraction would go beyond that claim to hold that *salient causes are candidates for the initial list of causes*. They are then processed subtractively and if not ruled out are reported as the person's attribution. The subtraction model thus predicts the salience phenomena that have been empirically obtained by the above-mentioned researchers and others, since salient causes will be considered as candidate causes and will often be given as an attribution. The proposal advanced here, however, goes beyond prior proposals to put the role of salience into a broader context of attributional information processing.

A second possible mechanism is scripts. If the action to be explained is part of an expected sequence of events, then the script (Schank & Abelson, 1977) of which it is part may suggest one or more plausible candidate causes. For example, if the waiter comes to one's table in a restaurant, one can tentatively guess that he may be bringing a menu, preparing to take one's order, and so forth. Again, these causes should then be subtractively processed. Thus, *scripted causes are also candidates for the initial list*. There are other possible sources of causes as well; for example, a person may simply (for idiosyncratic reasons) enter a situation with some specific expectation as to the nature of the causal factors operating there.

In any case, the subtraction model's difficulty seems solvable. Even outside the experimental context, sources of candidate causal factors (salience, scripts, or expectations) that enable the person to put together an initial list of one or more tentative causes are frequently available. The model holds that this initial list is then processed by attempting to rule out causes. When this has been done to the perceiver's satisfaction (the criterion may be influenced by demands to be careful, anxiety, or time pressure, etc.), the perceiver reports a cause. It is not yet clear what will happen if the perceiver succeeds in ruling out all candidate causes from the list. Presumably a new list will be put together and subjected to the same process of subtraction, and this should be time-consuming.

Conclusions

The data reported here are interpreted by proposing a two-stage model of the cognitive processing elicited by the experimental task, encoding or comprehending the stimulus text and attributional processing itself. Examina-

tion of the relationship between RT and the content of the response led to the proposal that perceivers assign a cause by subtracting causal components after beginning by considering as candidates the full set of three components. The subtractive model was shown to be consistent with salience effects on attributions, since perceptual salience can suggest candidate causes for consideration.

The methods applied in this paper have been much used in cognitive psychology but have as yet found little application in social psychology. The work of Kintsch (1974; Kintsch & van Dijk, 1978) and others on text comprehension should be referred to for clues as to how social psychologists might increase their knowledge of the first stage, the comprehending or encoding of events for attribution. Response time methods can be extended to cover attributional processing in other types of situation and with different materials and response formats, allowing tests of the idea that the basic subtractive nature of the attribution process may be a stable, general feature.

Reference Note

1. Kenny, D. A., & Smith, E. R. *A note on the analysis of designs in which subjects receive each stimulus only once.* Unpublished manuscript, University of Connecticut, 1979.

References

Anderson, J. R. *Language, memory, and thought.* Hillsdale, N.J.: Erlbaum, 1976.

Carpenter, P. A., & Just, M. A. Sentence comprehension: A psycholinguistic processing model of verification. *Psychological Review,* 1975, *82,* 45–73.

Clark, H. H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior,* 1973, *12,* 335–359.

Kassin, S. M., & Hochreich, D. J. Instructional set: A neglected variable in attribution research? *Personality and Social Psychology Bulletin,* 1977, *3,* 620–623.

Kelley, H. H. Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 15). Lincoln: University of Nebraska Press, 1967.

Kintsch, W. *The representation of meaning in memory.* Hillsdale, N.J.: Erlbaum, 1974.

Kintsch, W., & van Dijk, T. A. Toward a model of text comprehension and production. *Psychological Review,* 1978, *85,* 363–394.

Markus, H. Self-schemata and processing information about the self. *Journal of Personality and Social Psychology,* 1977, *35,* 63–78.

McArthur, L. The how and what of why: Some determinants and consequences of causal attribution. *Journal of Personality and Social Psychology,* 1972, *22,* 171–193.

Newell, A. Production systems: Models of control structures. In W. G. Chase (Ed.), *Visual information processing.* New York: Academic Press, 1973.

Orvis, B. R., Cunningham, J. D., & Kelley, H. H. A closer examination of causal inference: The role of consensus, distinctiveness, and consistency information. *Journal of Personality and Social Psychology,* 1975, *32,* 605–616.

Posner, M. *Chronometric explorations of mind.* Hillsdale, N.J.: Erlbaum, 1978.

Pryor, J. B., & Kriss, M. The cognitive dynamics of salience in the attribution process. *Journal of Personality and Social Psychology,* 1977, *35,* 49–55.

Ruble, D. N., & Feldman, N. S. Order of consensus, distinctiveness, and consistency information and causal attributions. *Journal of Personality and Social Psychology,* 1976, *34,* 930–937.

Schank, R. *Conceptual information processing.* Amsterdam: Elsevier, 1975.

Schank, R., & Abelson, R. P. *Scripts plans goals and understanding.* Hillsdale, N.J.: Erlbaum, 1977.

Schneider, W., & Shiffrin, R. M. Controlled and automatic human information processing I: Detection, search, and attention. *Psychological Review,* 1977, *84,* 1–66.

Shiffrin, R. M., & Schneider, W. Controlled and automatic human information processing II: Perceptual learning, automatic attending, and a general theory. *Psychological Review,* 1977, *84,* 127–190.

Simon, H. A. Discussion: Cognition and social behavior. In J. S. Carroll & J. W. Payne (Eds.), *Cognition and social behavior.* Hillsdale, N.J.: Erlbaum, 1976.

Sternberg, S. Memory-scanning: Mental processes revealed by reaction time experiments. *Acta Psychologica,* 1969, *30,* 276–315.

Taylor, D. A. Stage analysis of reaction time. *Psychological Bulletin,* 1976, *83,* 161–191.

Taylor, S. E. Developing a cognitive social psychology. In J. S. Carroll & J. W. Payne (Eds.), *Cognition and social behavior.* Hillsdale, N.J.: Erlbaum, 1976.

Taylor, S. E., & Fiske, S. T. Salience, attention, and attribution: Top of the head phenomena. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 11). New York: Academic Press, 1978.

Trabasso, T., Rollins, H., & Shaughnessy, E. Storage and verification stages in processing concepts. *Cognitive Psychology,* 1971, *2,* 239–289.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1970.

Received October 25, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 10, 1732-1741

The Effects of Self-Esteem, Success–Failure, and Self-Consciousness on Task Performance

Joel Brockner
State University of New York College at Brockport

Previous research has demonstrated that the task performance of low self-esteem individuals (low SEs) suffers in the presence of self-focusing stimuli (e.g., a mirror). The present study was designed to determine if such stimuli must inevitably have adverse effects on low SEs. It was reasoned that if low SEs were provided with success feedback from a previous task, then the nature of their self-consciousness would be altered on a subsequent task. Specifically, low SEs should attend more to positive and less anxiety-provoking aspects of themselves than would low SEs who received failure feedback from the previous task. Under the former condition, the low SEs' subsequent task performance was expected to improve. For high self-esteem individuals (high SEs), who typically perform well, previous success–failure feedback was expected to have little effect on subsequent performance. In a three-factor design, subjects high and low in chronic self-esteem received false success or failure feedback from a previous task and completed a concept formation task in either the presence or the absence of a mirror. Whereas high SEs performed equally well following success or failure, low SEs in the success condition performed significantly better than low SEs in the failure condition (and just as well as high self-esteem–success participants). This Self-Esteem × Prior Feedback interaction was significant in the presence of the mirror, but not in its absence. In the absence of the mirror, however, this interaction was observed for subjects who were high in dispositional self-consciousness, but not for those who were low. Practical and theoretical implications of these findings are discussed.

Clinical practitioners and researchers have long been concerned with ways to enhance the self-evaluations of individuals suffering from low self-esteem (SE). Unfortunately, this task has proven quite difficult, as low self-esteem people (low SEs) seem to be trapped in a "vicious cycle" of negativity. For example, many investigators (Hamachek, 1971; Shrauger, 1972) have reported that low SEs do significantly worse in achievement settings (i.e., situations that have evaluative implications for one's SE) than do high self-esteem subjects (high SEs). Poor performance, in turn, would seem to set the stage for continued self-criticism and low SE.

What are the precise causes of the low SEs' often poor task performances? A recent series of experiments (Brockner, 1979; Brockner & Hulton, 1978) suggested that one important factor was the low SEs' focus of attention during task performance. In several studies the performance of the low SEs (but not of the high SEs) was quite malleable across conditions, depending on the manipulation of their attentional focus. Thus low SEs did quite poorly in the presence of stimuli designed to increase self-focused attention (i.e., an audience, a mirror, and a video camera). However, the low SEs' performance was also enhanced considerably by

The author thanks Topic Editor Charles S. Carver, John Eckenrode, Marc Riess, Larry Wallnau, Robert Wyer, and two anonymous reviewers for their astute comments on an earlier version of the manuscript. He is also indebted to Michelle Beirne, Kathy Dems, Bruce Machlica, and Ken Watson for their very able assistance in the collection of the data.

Requests for reprints should be sent to Joel Brockner, Department of Psychology, SUNY College at Brockport, Brockport, New York 14420.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3710-1732$00.75

a set of instructions intended to increase task-focused and decrease self-focused attention. These latter data have implications for breaking the vicious cycle of low SE. It seems entirely possible that if low SEs repeatedly improve their task performance their self-opinions may begin to become more positive.

Low SE, Focus of Attention, Anxiety, and Performance

Brockner (1979) suggested that the attentional focus manipulations produced differential levels of anxiety for low SEs that in turn could have mediated task performance. That is, the self-focusing stimuli appeared to make the low SEs more anxious, whereas the task-focusing instructions, by reducing self-consciousness, were thought to make them less anxious. Brockner also presented evidence consistent with the notion that anxiety and performance were inversely related—not surprising, given the fairly complex nature of the task (see Spence, Farber, & McFann, 1956).

If Brockner's (1979) analysis is correct (i.e., if the attentional manipulations influenced the low SEs' anxiety, and if their anxiety mediated performance), then several questions confront researchers who wish to discover methods to improve the task performance of low SEs. Perhaps most important, does self-focused attention always covary with anxiety for low SEs? Stated another way, must low SEs' anxiety level always increase in the presence of self-focusing stimuli? If self-focused attention and anxiety always covary for low SEs, then self-focusing stimuli should inevitably produce decrements for low SEs on tasks in which anxiety is known to impair performance. However, if low SEs' self-focused attention and perceived anxiety need not covary, then it may be possible for low SEs to perform well on such tasks, even in the presence of self-focusing stimuli. In fact, Brockner (1979) observed that the low SEs' performance does not always have to suffer in the presence of self-focusing stimuli. In one study low SEs were placed in front of a mirror *and were also* instructed to concentrate on the task. It was found that these subjects made significantly fewer errors than low SEs who performed the task in front of a mirror but were not told to focus on the task. It was suggested that by concentrating on the task, the former group's level of self-focus, and therefore anxiety, was reduced.

To reduce the anxiety of low SEs, is it always necessary to *reduce* their level of self-focused attention? Or is it also possible to minimize their anxiety (and thereby enhance performance) by changing *the self-aspect that is focused on* when attention is self-directed? Both Duval and Wicklund (1972) and Wicklund (1975) have suggested that the affective nature of self-focused attention largely depends upon what happened to the individual prior to his being made self-aware. Therefore, under certain conditions, self-focused attention can be affectively pleasant; for example, if the individual received positive feedback on some personality dimension, then self-focused attention actually enhanced positive affect (Ickes, Wicklund, & Ferris, 1973). Moreover, recent research by Scheier and Carver (1977) convincingly demonstrated that self-focused attention can intensify any emotional experience, regardless of its affective nature.

Present Study

In the present study, subjects high and low in chronic self-esteem first received either success or failure feedback on a social insight test and then completed a concept formation task in either the presence or absence of a self-focusing stimulus. Results obtained by Ickes et al. (1973) and Scheier and Carver (1977) suggest that low SEs who are provided with success feedback and then made self-aware should experience self-focus characterized by more positive and/or fewer anxiety-provoking thoughts. Under these circumstances low SEs should perform well on the concept formation task. Alternatively, if they are provided with failure feedback, low SEs will probably experience negative affect. The self-focusing stimulus accompanying the concept formation task should then cause an increase in anxiety awareness. Accordingly, low SEs would be expected to do quite poorly on the task in this condition.

On the basis of previous research, it was expected that the affective state of the high SEs would be less influenced by the success–failure feedback than the affective state of the lows. Given that high SEs are generally already low in anxiety (Crandall, 1973), the success feedback should not have such pronounced effects. Moreover, research by Shrauger and Rosenberg (1970) strongly suggests that the high SEs' greater ego strength would allow them to withstand the potentially anxiety-provoking effects of the failure feedback. High SEs, in addition to being less affectively influenced by success–failure feedback, also report lower levels of self-focused attention in the presence of self-focusing stimuli (Brockner, 1979; Brockner & Hulton, 1978). For these two reasons the self-focusing stimulus should not have the amplifying effects on affect and performance that it is expected to have on the low SEs. Rather, high SEs should perform well in the self-focus condition after either success or failure.

Because the low SEs' affective states were expected to be less polarized in the absence of self-focusing, the condition differences in performance in the no self-focus conditions, although similar to those expected in the self-focus conditions, may not attain statistical significance.

In sum, the present study consisted of a three-factor (SE × Prior Feedback × Attentional Focus) design. A SE × Prior Feedback interaction was predicted, such that (a) low SEs would perform worse in the failure than in the success condition, with no corresponding difference for high SEs, and (b) low SEs would perform worse than high SEs in the failure condition, but just as well as high SEs in the success condition. Moreover, this interaction effect was only clearly expected to be significant in the self-focus condition.

Subjects also completed a measure of dispositional self-consciousness (the private self-consciousness scale of Fenigstein, Scheier, & Buss, 1975) prior to performing the concept formation task. Previous research has shown that this individual difference variable, like situational manipulations of self-focused attention, is directly related to awareness of one's emotional state (Scheier, 1976; Scheier & Carver, 1977). Furthermore, it has been suggested that personality variables are most likely to be related to behavior when relevant situational stimuli are absent or ambiguous (Mischel, 1973). Therefore, it was expected that in the no self-focus condition, (a) the performance of subjects high in private self-consciousness would parallel the behavior already predicted for subjects in the self-focus condition, and (b) subjects low in private self-consciousness would be less apt to demonstrate the basic SE × Prior Feedback interaction effect hypothesized above.

Method

Participants

The 100 undergraduate participants were volunteers recruited on the campus of the SUNY College at Brockport. The data of 7 subjects who were suspicious about the experimental procedure and 3 who did not understand the instructions were not included.[1] The final sample consisted of 60 females and 30 males, who were randomly assigned to conditions. Subjects' SE and self-consciousness were assessed by scales completed immediately prior to the experiment.

Self-esteem. The self-esteem scale was identical to the one used by Brockner (1979) and Shrauger (1972). The scale measures the subjects' perceived competence across 16 various situations (academic, athletic, and social). Subjects were asked to indicate the percentage of time a particular positive behavior or outcome applied to them. The average score for all 16 items was computed for each subject, such that higher scores represented higher SE. Subjects were classified as high or low SEs on the basis of a median split.

Self-consciousness. Subjects also indicated on 6-point scales how much each of the private self-consciousness statements of Fenigstein et al. (1975) was characteristic of them. The scores for each subject were summed over the 10 items, with higher scores representing greater self-consciousness. A median split was employed to classify subjects as high or low in private self-consciousness. SE and private self-consciousness were not related, $r(88) = .06$.

Procedure

All participants were run one at a time. Upon entering the laboratory they were greeted by a

[1] Interestingly, all of the discounted subjects for whom there were data ($N = 7$) were in the failure condition. More will be said about their performance on the concept formation task in the Results section.

female experimenter, led to a cubicle, and told that this "problem-solving" experiment would consist of several parts. First, subjects were asked to complete the SE and self-consciousness inventories. Afterward they responded to the Chapin Social Insight Test (Gough, 1968).

Social insight test. For each of these 10 problems, subjects read a short paragraph describing a social situation. They then had to choose one of four alternatives that they thought was the best explanation of the behavior of the individuals described in the paragraph. Subjects were told that the test measures the ability to "see into" social situations and, to heighten their psychological involvement, that social insight is "correlated with a person's capacity to show empathy for others, to predict others' responses in social situations, and it relates to a person's general understanding of human nature. After completing the test you will be informed of how well you did on it." The test took 10–15 minutes to complete.

When subjects finished the test, the experimenter excused herself while she went to "score" their performance. Returning about 3 minutes later, she administered the false feedback. In the success condition subjects were told that they scored in the 90th percentile of a group "consisting of you and your peers." In the failure condition subjects were told that they had scored in the bottom 20% of their peer group. After leaving the subjects for 1 minute so that they would absorb the feedback, the experimenter returned and asked them to complete a manipulation check questionnaire.

Subjects were then led across the hall to "complete the last part of the experiment." In this second room, a male experimenter blind to the subjects' feedback condition and SE and self-consciousness scores seated the subjects at a desk where they found the concept formation task (described immediately below) and the accompanying instructions. Subjects read the instructions to themselves while the experimenter simultaneously read them aloud. Subjects were also given a practice trial to familiarize them further with the task. After the practice trial was completed, subjects estimated on a 41-point scale how well they expected to perform at the task (endpoints were "very poorly" [1] and "very well" [41]). These data were collected for exploratory purposes but will become relevant in the discussion of the results.

Concept formation task. The task was identical to the one employed by Brockner (1979; Brockner & Hulton, 1978). Specifically, 32 squares were attached to a $14 \times 25 \times .25$ inch ($36 \times 64 \times .6$ cm) wooden board. Each square had five characteristics: number, size, color, shape, and position. Beneath each square a small hole had been drilled in the wooden board. The subject was told that the concept consisted of one, two, or three characteristics. The experimenter selected a square that was an example of the correct concept. The subjects' task was to determine the concept by lifting up other squares, one at a time. If the square was an example of the correct concept, the word YES was written on a piece of oak tag below the board. If the square was not an example of the concept the word NO was written below. When subjects believed that they had ascertained the correct concept, they wrote it on a slip of paper provided to them. At no time were subjects ever informed of the accuracy of their responses. Subjects were told that although "we are interested in the speed and accuracy with which you determine the concepts, it is more important to be accurate."

Attentional focus. The manipulation of attentional focus was introduced just before subjects began the concept formation task. In the self-focus condition subjects were told:

> One more point. Since we are interested in how people go about forming concepts, the experimenter will be watching you as you are working on the task. Also, a mirror will be placed in front of you, making it easier for the experimenter to observe your behavior more closely.

With that, the experimenter placed a large mirror (24×36 inches—61×91 cm), which had hitherto been unexposed, directly in front of the subject. A large mirror was employed in order to strengthen its self-focusing effect. That is, subjects could see not only themselves but also the experimenter observing them as they worked on the task. In the no self-focus condition the mirror manipulation was never introduced.

In all conditions subjects then began to work on the task, which was comprised of seven trials. The experimenter, seated adjacent to the subjects, recorded the amount of time taken and the number of squares looked at for each trial. When the seventh trial was completed, subjects answered a postexperimental questionnaire. In addition to indicating the nature of their attentional focus and the anxiety they felt while completing the task, subjects wrote their hypotheses and suspicions about the study. Finally, because of the potentially sensitive nature of the subject matter under study, all participants were carefully debriefed.

Results

Performance [2]

Errors of both commission and omission were summed over the seven trials to deter-

[2] Only the error data will be reported. Other qualitative performance measures included the time taken to complete the trials and the number of squares looked at before subjects made their responses. Consistent with previous research (Brockner, 1979; Shrauger, 1972), there were no condition differences on the time measure. The only effect to emerge on the other measure was the Feedback × Attentional Focus interaction ($p < .05$). This effect will not be considered further because it is not consistent with any other findings in this study, and no previous researchers have ever found condition differences on this measure.

mine task performance. A preliminary t test revealed no effect for sex of subject, $t < 1$. Therefore, the data were analyzed with a three-factor unweighted-means analysis of variance. The only effect to attain significance was the SE × Prior Feedback interaction, $F(1, 82) = 4.22$, $p < .05$. Simple effects elucidated the nature of this effect. Specifically, low SEs performed worse than high SEs in the failure condition, $F(1, 82) = 7.39$, $p < .01$, but not in the success condition ($F < 1$).[3] Moreover, low SEs made significantly more errors in the failure than in the success condition, $F(1, 82) = 6.81$, $p < .02$, whereas high SEs were unaffected by the feedback manipulation ($F < 1$).

While the triple interaction did not achieve significance, the means in Table 1 suggest that the significant effects reported above were largely attributable to the results obtained in the self-focus condition. Indeed, planned comparisons (i.e., simple interaction effect analyses) demonstrated that the SE × Feedback interaction was significant in the self-focus, $F(1, 82) = 4.55$, $p < .05$, but not in the no self-focus condition ($F < 1$).[4] Additional simple effect analyses revealed that the only instance in which low SEs made significantly more errors than high SEs was in the failure–self-focus condition, $F(1, 82) = 10.25$, $p < .005$, whereas the only time significantly more errors were made in the failure than in the success condition was in the low SE – self-focus condition, $F(1, 82) = 4.02$, $p < .05$.

Table 1
Task Performance by Condition

Attentional focus/self-esteem	Prior feedback: Success n	M	SD	Failure n	M	SD
Self-focus						
High	11	3.64	2.34	14	2.21	2.36
Low	10	3.90	3.54	11	6.73	4.88
No self-focus						
High	14	3.36	2.13	10	4.20	4.21
Low	11	2.73	3.35	9	5.11	3.37

Note. Performance was measured by errors. Scores ranged from 0 to 18. Thus the higher the score, the poorer the performance.

Table 2
Task Performance by Condition, Including Private Self-Consciousness

Private self-consciousness/ self-esteem	Prior feedback: Success n	M	SD	Failure n	M	SD
Self-focus						
High/high	5	4.40	2.07	10	1.70	1.57
High/low	5	6.40	3.13	5	5.20	4.32
Low/high	6	3.00	2.53	4	3.50	3.70
Low/low	5	1.40	1.67	6	8.00	5.33
No self-focus						
High/high	6	5.00	1.55	4	3.25	2.36
High/low	6	2.17	3.06	4	7.50	3.42
Low/high	8	2.13	1.64	6	4.83	5.23
Low/low	5	3.40	3.19	5	3.20	1.92

Note. Performance was measured by errors. Scores ranged from 0 to 18. The higher the score, the poorer the performance.

Self-consciousness analysis. To test for the main and interaction effects involving the dispositional self-consciousness variable, a four-factor (Self-Consciousness × SE × Feedback × Attentional Focus) unweighted-means analysis of variance was performed. The only significant effect to emerge, in addition to the already reported SE × Feedback interaction,

[3] The data of seven subjects in the failure condition were discarded because of their suspicions about the validity of the social insight test. Perhaps because the impact of the failure feedback was lessened for these subjects, they tended to make very few errors on the task. Nevertheless, the four subjects who would have been low SEs still made more errors ($M = 2.50$) than the three would-be high SEs ($M = 1.33$).

[4] Previous researchers (Brockner, 1979; Brockner & Hulton, 1978; Shrauger, 1972) have found SE differences on this task in the presence, but not in the absence, of self-focusing stimuli. Furthermore, the a priori hypothesis of the present study, on theoretical grounds, was that the SE × Feedback interaction would be significant in the self-focus but not necessarily in the no self-focus condition. Winer (1971) clearly states that it is legitimate to compare specific means, if the comparisons were planned prior to the inspection of the data. The appropriate planned comparisons, therefore, were the simple interaction effect analyses in the self-focus and no self-focus conditions.

was the four-way interaction, $F(1, 74) = 7.25$, $p < .01$. By inspecting Table 2, one finds that there were nonsignificant performance differences between SE groups in the success conditions. In the failure conditions high SEs performed far better than low SEs, except in the absence of a source of self-focused attention (i.e., the no self-focus – low self-consciousness condition). Although the means in the self-focus condition suggest that the SE × Prior Feedback interaction was stronger for subjects low, rather than high, in dispositional self-consciousness, the simple triple interaction effect (comparing the SE × Feedback interaction for high vs. low self-conscious subjects) revealed that this difference was not significant, $F(1, 74) = 1.44$. In the no self-focus condition, high self-conscious subjects exhibited a similar SE × Feedback interaction shown by all subjects in the self-focus condition, whereas low self-conscious subjects did not. Indeed, the simple triple interaction effect did attain significance in the no self-focus condition, $F(1, 74) = 6.79$, $p < .02$. In short, in the no self-focus condition, subjects high in self-consciousness responded (a) like subjects in the self-focus condition and (b) unlike subjects low in self-consciousness. Stating the latter effect more generally, the personality variable only had an effect in the absence of the mirror, that is, the relevant situational stimulus (Mischel, 1973).[5]

Manipulation Checks

Prior feedback. After receiving their feedback but before starting to work on the concept formation task, subjects completed two 41-point scales measuring their affective states (Question 1: At this point, how socially insightful do you think you are? Anchor points: "not at all" [1] and "very" [41]; Question 2: How would you describe your current mood state? Anchor points: "very bad mood" [1] and "very good mood" [41]). Since the two items correlated significantly, $r(88) = .44$, $p < .01$, an affect index was computed for each subject by summing the scores. Because these measures were collected before the attentional focus variable had been manipulated, a two-factor unweighted-means analysis of variance was performed on the index. The analysis revealed a significant interaction effect, $F(1, 86) = 7.10$, $p < .01$. Low SEs expressed more positive affect in the success than in the failure condition, $F(1, 86) = 24.27$, $p < .001$, whereas high SEs did not, $F(1, 86) = 1.34$. Moreover, high SEs were only more positive than low SEs in the failure condition, $F(1, 86) = 12.10$, $p < .005$.

Attentional focus. Two 41-point self-report items were employed to measure the amount of positive and negative thoughts subjects were having about themselves while working on the task. Question 1 measured positive (e.g., feeling confident) and Question 2 negative (e.g., thinking I was doing poorly) self-evaluative thoughts. Three-factor unweighted-means analyses revealed a SE × Feedback interaction on the first question, $F(1, 82) = 3.98$, $p < .05$. Specifically, it was only in the failure condition that high SEs were having more positive thoughts than low SEs. On the second question the means were in the opposite direction, but all effects were nonsignificant. Thus, it was only in the failure condition that low SEs tended to have more negative thoughts than high SEs.

Anxiety. Subjects were also asked, "*While you were completing the task,* (a) how anxious did you feel, and (b) how much did the experimenter's presence bother you?" Since the two items were well correlated, $r(87)$[6] $= .42$, $p < .01$, an anxiety index was formed by summing each subject's scores. The analysis of variance revealed a significant SE main effect, $F(1, 81) = 4.85$, $p < .05$, with low SEs more anxious than high SEs, and a mar-

---

[5] Subjects were also measured on the other two subscales of the self-consciousness scale of Fenigstein et al. (1975), that is, public self-consciousness and social anxiety. The four-factor analysis involving the public self-consciousness variable yielded a marginally significant four-way interaction ($p < .10$), the nature of which was similar to that observed on the private self-consciousness analysis. The social anxiety analysis was not performed because an unbalanced distribution of that variable caused one of the conditions to contain only one subject.

[6] One subject inadvertently failed to complete the first anxiety measure. Thus, the degree of freedom is 87 rather than 88.

ginal SE × Feedback interaction, $F(1, 81) = 2.76$, $p < .10$. Interestingly, the mean condition differences on the index were remarkably similar to the performance data reported in Table 1. Although the triple interaction was not significant, simple interaction effect analyses showed that the SE × Prior Feedback interaction effect was significant in the self-focus condition, $F(1, 81) = 4.90$, $p < .05$, but not in the no self-focus condition ($F < 1$). Low SEs were only significantly more anxious than high SEs in the failure–self-focus condition, $F(1, 81) = 8.64$, $p < .01$, and the only instance in which failure subjects were more anxious than success subjects was the low SE – self-focus condition, $F(1, 81) = 4.90$, $p < .05$.

Correlational Data

The pattern of the SE × Feedback interaction was similar on measures of self-evaluative thoughts, anxiety, and task performance. Further suggestive evidence that task performance was mediated by subjects' self-evaluative thoughts and anxiety levels stems from a series of correlations. Specifically, subjects' positive thoughts were inversely related, $r(88) = -.42$, $p < .01$, and their negative thoughts were directly related, $r(88) = .36$, $p < .01$, to the number of errors they made. There was also a positive correlation between subjects' anxiety index scores and the number of errors they committed, $r(87) = .33$, $p < .01$.

Discussion

In sum, the major hypotheses of the study were supported. It was predicted that the feedback would cause low (but not high) SEs to attend to different aspects of themselves, such that the self-evaluative statements made by low SEs would be more anxiety-provoking following failure than following success. Self-focused attention was expected to intensify anxiety awareness, which in turn was thought to mediate task performance. Indeed, the task performance results showed that low SEs made more errors than high SEs following failure, as long as there was some source of self-focused attention (i.e., the situationally produced self-awareness or subjects' own tendencies towards self-consciousness). Following success, low SEs performed just as well as high SEs in all conditions.

The fact that low SEs performed better following success has implications for breaking the vicious cycle of low self-esteem. That is, previous investigators (Brockner, 1979; Brockner & Hulton, 1978; Shrauger, 1972) have demonstrated that low SEs' performance deteriorates in the presence of self-focusing stimuli. The present results suggest, however, that by reducing the low SEs' anxiety before making them self-aware, one can block the negative effects of the self-focusing stimulus on their performance. Of course, in the absence of a no feedback – self-focus condition, it is not known whether the success feedback reduced the low SEs' anxiety and enhanced their performance, whether the failure further heightened anxiety and impaired performance, or whether both occurred. However, in what would amount to a relevant no feedback – self-focus condition, Brockner (1979) found that high SEs made 2.50 errors, whereas the mean score for low SEs was 7.22. These data are very close to the means observed in the failure – self-focus condition in the present study. Thus, for low SEs in the self-focus conditions, it would appear that the success feedback reduced anxiety and enhanced performance, whereas the failure had only slight effects.

Alternative interpretation. It could be argued that the mirror produced its effects by serving as a source of general arousal. More specifically, consider the following logical sequence. Feedback may have manipulated the direction of attentional focus experienced by low SEs, so that they were more task-focused and less self-focused following success than failure. Accordingly, their performance should be better following success. The mirror may then have acted as an arousal source, enhancing performance differences by further restricting the range of cues to which subjects attended (Easterbrook, 1959). That is, restriction of cues among task-focused subjects would have increased task focus; cue restriction among self-focused subjects would have further decreased task focus. How might the mirror's

presence have enhanced arousal? Given the instructions that accompanied the introduction of the mirror ("so that the experimenter can watch you more closely") and the fact that subjects could see the exprimenter observing them in the mirror, the perceived importance and/or evaluative nature of the task may have been greater in the self-focus than in the no self-focus condition.

Although this alternative explanation warrants consideration, there is evidence against it. For example, if the mirror was a source of general arousal because it caused subjects to attribute more significance to the task, then they should have assigned greater importance to the task in self-focus than no self-focus conditions. Although no measures of perceived task importance were collected in this study, Brockner (1979) did measure "the importance of doing well at this task" as a function of SE and the presence of a mirror and video camera (using the same justification to introduce the self-focusing stimuli that was employed in the present study). There was no tendency for subjects (including low SEs) to attribute more importance to the task in the presence of the self-focusing stimuli.

In addition, it has been demonstrated that private self-consciousness does not correlate with individual difference measures of arousal/emotionality (e.g., general emotionality, test anxiety, social anxiety; Scheier & Carver, 1977). In examining the performance data of high private self-conscious subjects in the no self-focus condition (see Table 2), one finds that the SE × Feedback interaction was parallel to the one observed for *all* subjects who were exposed to the mirror. In attempting to account for such parallel effects of manipulated and dispositional self-attention, Scheier and Carver have suggested that "an interpretation based on arousal may provide a viable explanation for the effects of a mirror, but it cannot account for the effects of private self-consciousness. An interpretation based on self-focused attention can explain both the effects of a mirror and the effects of private self-consciousness. Thus, although different interpretations can be used to explain the effects of a mirror and private self-consciousness separately, at this time the self-awareness interpretation appears more reasonable, comprehensive, and parsimonious" (Scheier & Carver, 1977, p. 634).

A similar line of reasoning may be used to suggest that the mirror did not cause increased evaluation apprehension. That is, if the mirror enhanced evaluation apprehension, then subjects in the no self-focus condition who are more sensitive to the evaluation process (those high in social anxiety; Turner, 1977) should have performed as did those in the self-focus condition. As stated previously, the social anxiety analysis was not conducted because one of the cells (in the no self-focus – low social anxiety condition) contained only one subject. However, there were enough subjects in the no self-focus – high social anxiety condition to test this hypothesis. Interestingly, the results did not at all resemble the SE × Feedback interaction obtained in the self-focus condition. Rather, there was a slight tendency for both high and low SEs to make more errors following failure.

Self-Attention and Behavioral Self-Regulation

The present data are generally consistent with Carver's (1979) analysis of the consequences of self-directed attention. In essence, Carver suggests that when subjects are self-aware they will be more motivated to conform to the salient behavioral standard in a situation (e.g., to solve the concept formation task successfully). If arousal cues are not introduced, the subject should complete the task successfully. If fear or anxiety arousal cues become salient, however, the subject's task performance will be interrupted. At this point the individual will assess his/her likelihood of being able to complete the task. If the subject's expectations are positive, he/she will vigorously reattempt to match behavior to the standard. Subjects who feel unable to complete the task will respond with passivity and withdrawal.

To test this model, Carver, Blaney, and Scheier (1979a, 1979b) have presented subjects with fear or anxiety-provoking situations and have measured task persistence/performance. The results of these studies have generally supported the model: positive expect-

ancy subjects exhibit greater persistence/performance than negative expectancy participants, but only when self-focused attention is high.

One of the experiments of Carver et al. (1979b) is conceptually quite similar to part of the present design. In that study, all subjects were provided with failure feedback. Persistence at a subsequent task was measured as a joint function of self-awareness and expectancy for success at that second task. In the present study, some subjects were given failure feedback and then had to perform the concept formation task with no manipulation of expectancy. High SEs performed better than low SEs in the present research when attention was self-focused, but not when self-focus was lower. These results parallel those of Carver et al. if it can be assumed that (a) task persistence is related to task performance and (b) SE is positively related to expectancy for success at the concept formation task. The first assumption seems straightforward, and the second one was empirically supported in the present study. High SEs *did* expect to perform better than low SEs, $F(1, 86) = 6.77$, $p < .025$, particularly in the failure condition, $F(1, 86) = 5.80$, $p < .025$.

These performance data not only confirm those obtained by Carver et al. (1979a, 1979b), but they provide converging support for Carver's (1979) theoretical model. That is, to demonstrate that self-focusing is a necessary component of the model, Carver et al. employed situational manipulations of self-focused attention. In the present study, the fact that only subjects high in dispositional self-consciousness in the no self-focus condition performed as did those in the self-focus condition further attests to the importance of self-focused attention.

The lack of SE differences in performance in the success–self-focus conditions can also be explained by self-regulation theory. First, given the previous success, rather than failure, feedback, subjects' arousal level while completing the task may not have been high. Hence they would not necessarily be expected to interrupt task performance to assess the likelihood of successful task completion. Moreover, even if subjects did assess these expectancies, no SE differences in performance would be predicted, as high and low SEs did not reliably differ in their expectations for performance in the success condition, $F(1, 86) = 1.62$.

Conclusions

The task performance findings in this study are a logical extension of those observed in previous research (Brockner, 1979). In the earlier study, to enhance their task performance in the presence of a self-focusing stimulus, low SEs were told to concentrate diligently on the task. Thus, the previous study adopted a quantitative approach to combating self-focused attention for low SEs, in that the task-focus instructions were designed to reduce *the degree of* the low SEs' self-focusing. The present study entailed a more qualitative strategy to offset the negative effects on performance of the low SEs' self-focused attention. Rather than reduce the degree of their self-focusing, an attempt was made to change *the nature of* the low SEs' self-focused attention by providing them with success feedback from a previous task. This strategy also proved to be effective, as the low SEs' task performance was no longer impaired in the presence of the self-focusing stimulus. These performance data are consistent with and serve to extend recent theorizing about the consequences of self-directed attention.

References

Brockner, J. Self-esteem, self-consciousness, and task performance: Replications, extensions, and possible explanations. *Journal of Personality and Social Psychology*, 1979, *37*, 447–461.

Brockner, J., & Hulton, A. J. B. How to reverse the vicious cycle of low self-esteem: The importance of attentional focus. *Journal of Experimental Social Psychology*, 1978, *14*, 564–578.

Carver, C. S. A cybernetic model of self-attention processes. *Journal of Personality and Social Psychology*, 1979, *37*, 1251–1281.

Carver, C. S., Blaney, P. H., & Scheier, M. F. Focus of attention, chronic expectancy, and responses to a feared stimulus. *Journal of Personality and Social Psychology*, 1979, *37*, 1186–1195. (a)

Carver, C. S., Blaney, P. H., & Scheier, M. F. Reassertion and giving up: The interactive role of self-directed attention and outcome expectancy. *Journal of Personality and Social Psychology*, 1979, *37*, 1859–1870. (b)

Crandall, R. The measurement of self-esteem and related constructs. In J. R. Robinson and P. R. Shaver, *Measures of social psychological attitudes*. Ann Arbor, Mich.: Institute for Social Research, 1973.

Duval, S., & Wicklund, R. A. *A theory of objective self-awareness*. New York: Academic Press, 1972.

Easterbrook, J. A. The effect of emotion on cue utilization and the organization of behavior. *Psychological Review*, 1959, *66*, 183–201.

Fenigstein, A., Scheier, M. F., & Buss, A. H. Public and private self-consciousness: Assessment and theory. *Journal of Consulting and Clinical Psychology*, 1975, *43*, 522–527.

Gough, H. *The chapin social insight test*. Palo Alto, Calif.: Consulting Psychologists Press, 1968.

Hamachek, D. *Encounters with the self*. New York: Holt, Rinehart & Winston, 1971.

Ickes, W. J., Wicklund, R. A., & Ferris, C. B. Objective self-awareness and self-esteem. *Journal of Experimental Social Psychology*, 1973, *9*, 202–219.

Mischel, W. Toward a cognitive social learning reconceptualization of personality. *Psychological Review*, 1973, *80*, 252–283.

Scheier, M. F. Self-awareness, self-consciousness, and angry aggression. *Journal of Personality*, 1976, *44*, 627–644.

Scheier, M. F., & Carver, C. S. Self-focused attention and the experience of emotion: Attraction, repulsion, elation, and depression. *Journal of Personality and Social Psychology*, 1977, *35*, 625–636.

Shrauger, J. S. Self-esteem and reactions to being observed by others. *Journal of Personality and Social Psychology*, 1972, *23*, 192–200.

Shrauger, J. S., & Rosenberg, S. E. Self-esteem and the effect of success and failure feedback on performance. *Journal of Personality*, 1970, *38*, 404–417.

Spence, K. W., Farber, I. E., & McFann, H. H. The relation of anxiety (drive) level to performance in competitional and noncompetitional paired associate learning. *Journal of Experimental Psychology*, 1956, *52*, 296–305.

Turner, R. G. Self-consciousness and anticipatory belief change. *Personality and Social Psychology Bulletin*, 1977, *3*, 438–441.

Wicklund, R. A. Objective self-awareness. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 9). New York: Academic Press, 1975.

Winer, B. J. *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill, 1971.

Received October 27, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 10, 1742-1757

A New Round Robin Analysis of Variance for Social Interaction Data

Rebecca M. Warner
University of Miami, Coral Gables

David A. Kenny
University of Connecticut

Michael Stoto
Harvard University

Experimental social psychology has dealt primarily with situations that are not true social interactions; in a typical study, a subject responds to a fixed, artificial social stimulus such as a photograph, written description, or performance by a confederate. Although these artificial social stimuli provide experimental control over independent variables and can be analyzed using the types of statistical models originally developed for nonsocial experimental research, they provide little or no information about the interactive aspects of social behavior—the reciprocity or mutual contingency of the behavior of interaction partners. This paper describes a nonexperimental design specifically tailored to social interaction data that provides more information about individual differences and social influence in social interactions: a round robin design in which each person interacts with every other person. After a brief review of available models, a new and more general model for the analysis of social interaction data is presented, with an empirical demonstration using vocal activity data.

Experimental social psychology has often been limited to the study of artificial, one-sided social situations in which the subject responds to some fixed social stimulus created by the experimenter; this has been the case even in research on person perception and interpersonal attraction, where it is especially apparent that the behaviors and attitudes of persons are mutually contingent (i.e., A's liking for B affects B's liking for A, and vice versa). In research on subject reactions to fixed stimuli (rather than the reactions to other persons in the context of naturally occurring social interactions), information about the reciprocity of social behaviors is lost. Moreover, the use of standardized, artificial stimuli often precludes having stimuli that are representative of typical interactions. Thus, the generalizability of such studies may be rather limited.

A round robin design is one in which all possible pairs of subjects from some set of subjects interact (like the pairings formed in a tennis tournament). This term was introduced by Gleason and Halperin (1975). For each person paired with every other person, an observation is made of some social behavior (speech pattern, rating, or degree of attraction). Thus the "treatments" to which each subject reacts are the behaviors of other subjects. This design provides two kinds of information about social behavior—first, information about individual differences among subjects (in speech patterns, ratings, or other behaviors); second, information about the mutual influence that interaction partners have on each other's behaviors (for instance, the tendency to reciprocate positive feelings or to match the durations of certain kinds of pauses in speech).

Before introducing a new round robin analysis of variance, two simpler designs for social interaction data will be briefly reviewed. Each of these designs provides partial information about social behavior, and the round robin design can be understood as a model

Development of this model was supported in part by National Science Foundation Grant BNS 77-03271, David A. Kenny, principal investigator. Additional funds were provided by the Livingston Fund Award (Harvard Medical School).

Requests for reprints should be sent to Rebecca M. Warner, Department of Psychology, University of Miami, Coral Gables, Florida 33124.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3710-1742$00.75

that integrates and elaborates on these two simpler designs.

Intraclass Correlation

A common nonexperimental design for social interaction research is one in which a number of different dyads are formed and each subject's behavior toward one randomly assigned partner is observed. An example is a study reported by Welkowitz, Cariffe, and Feldstein (1976) on the congruence of vocal activity (i.e., the tendency for conversation partners to match the mean durations of their switching pauses, the pauses that occur after one person has stopped speaking and before the other person begins to speak). For each subject there is one observation, the mean switching pause duration in seconds. The observations are arranged by dyads, as illustrated by the hypothetical data displayed in Table 1.

To determine the strength of the tendency for partners to match pause durations, it is necessary to look at the correlation between the two columns of data, but an ordinary correlation coefficient is inappropriate here because the data do not consist of ordered pairs, that is, the assignment of any particular observation to the first or second column is arbitrary. This type of layout also occurs in examining twin data, as in research on IQ; here also there is no basis for labeling one twin X and the other twin X'. For this type of data layout an intraclass correlation is required (Snedecor & Cochran, 1967, 294–296). The r intraclass is calculated slightly differently from the ordinary product-moment r; each pair of observations is counted twice when computing the covariance term in the numerator—once as X, X' and again as X', X. Instead of calculating the variances of the X and X' groups separately, the denominator consists of the variance of all the observations. The r intraclass is just the ratio of these two terms. The intraclass correlation provides information about the matching or reciprocity of behaviors between partners, but it ignores individual differences among persons.

Table 1
Hypothetical Data Layout for Intraclass Correlation of Switching Pause Durations

Dyad	X	X'
1	.7	.9
2	1.2	1.1
3	1.5	.8

Note. Each entry is a hypothetical mean switching pause duration (in seconds). The sequence can of course be extended for Dyads 4, 5, 6, and so forth. X = switching pause duration for one of the speakers; X' = switching pause duration for the other speaker.

Table 2
Hypothetical Data Layout for an Analysis of Variance Design: Mean Proportion of Time Spent Speaking by Doctors

	Patients	
Doctor	Quiet, depressed	Talkative, anxious
1	23.4	15.7
2	36.6	24.1
3	45.9	26.3

Note. Design used by Goldman-Eisler (1952).

Ordinary Two-Way Analysis of Variance

A second approach to the study of social interaction involves the partitioning of variances to look for individual subject differences. Individual persons (or groups of persons) are treated as levels of subject factors in an ordinary analysis of variance. When there is some asymmetry of roles (such as doctor/patient or interviewer/interviewee), the data are easily translated into a two-way analysis of variance. An example of this design is a study by Goldman-Eisler (1952); the hypothetical data in Table 2 represent the mean proportion of time each doctor spent talking when paired with individual patients of two different types. This design can be treated as a two-way analysis of variance, and comparisons can be made of the following: individual differences among doctors in talkativeness; differences in the activity levels elicited from doctors by the two different types of patients; and interactions between doctor and patient type.

This design does deal with individual differences (both in the types of social behavior

Table 3
Round Robin Data Layout

Subject	Subject 1	2	3	4
1	—	X_{121}	X_{131}	X_{141}
	·	X_{122}	X_{132}	X_{142}
	—	X_{123}	X_{133}	X_{143}
2	X_{211}	—	X_{231}	X_{241}
	X_{212}	·	X_{232}	X_{242}
	X_{213}	—	X_{233}	X_{243}
3	X_{311}	X_{321}	—	X_{341}
	X_{312}	X_{322}	·	X_{342}
	X_{313}	X_{323}	—	X_{343}
4	X_{411}	X_{421}	X_{431}	—
	X_{412}	X_{422}	X_{432}	·
	X_{413}	X_{423}	X_{433}	—

Note. X_{ijk} = a behavior of Person i toward Person j on Day k.

that individuals exhibit and in the types of social behavior that groups of patients elicit); but it is less well suited to the study of social interactions among peers (i.e., social interactions that do not have a role asymmetry that makes it easy to assign some individuals to the row factor and others to the column factor); furthermore, it provides no information about mutual influence of social behaviors.

Round Robin Analysis of Variance

Model Specification

Round robin designs make it possible to obtain both kinds of information about social interactions—individual differences and mutual contingency. Several round robin–type designs have been developed (Bechtel, 1971; Gleason & Halperin, 1975; Lev & Kinder, 1957); all of these designs are applicable to data layouts in which each subject is paired with every other subject, but they differ with respect to the assumptions that they have made about the kinds of nonindependence among observations. Since these models have developed as a special application of the paired-comparisons design in psychophysical research, some of the assumptions of these models are rather restrictive or in other respects not ideally suited for some of the special difficulties that arise in social psychological research. The reason for developing a new model is to tailor the round robin design more specifically to the problems of social interaction and to make available a model with less restrictive assumptions about the type of dependence among observations. Depending on the research problem, however, one of these previously suggested models may be more appropriate; accordingly, the specifications of these other models will be described briefly later in this paper to assist readers in choosing the best model.

Round robin designs arise frequently in small-groups research, where records are kept of the frequency or duration of various types of acts that each group member addresses to each other group member; in studies of person perception, where each person rates every other person in a group; and such data may be generated by creating all possible pairs from a small subject pool and studying their behavior in isolation or by collecting data in group settings.

Any social psychological dependent variable can be used: summary information on speech, gaze, or body movements; frequency counts of the number of acts (aggressive, altruistic, or other); ratings, perceptions, or self-report measures of utility—the amount of some resource (money, time, materials, or whatever) that each person gives to each other person.[1]

A round robin data layout is illustrated in Table 3, where X_{ijk} is an observation of some social behavior of Person i toward Person j on Day or Time k.

The diagonal cells are empty, since ordinarily a person cannot be paired with himself or herself, and the matrix is asymmetric, since the behavior of Person i toward j is ordinarily different from that of j to i, even though they may be highly correlated. With some modifications, this will be treated as an $n \times n \times r$ analysis of variance, where n is the number of subjects and r is the number of observations made on each dyad. This is a random effects

[1] Note two limitations on appropriate data for the round robin analysis of variance—there should not be many cells with zero entries (apart from the diagonal cells), and neither the row nor column sums should be constant, as would occur if each subject allocated limited resources among other subjects (such as rank-order preferences or limited amounts of money).

model, that is, the set of n subjects is considered to be a very small random sample from a very large population of subjects; subjects were not selected to represent levels of some dimension (such as dominance). By treating this as a random effects model, we achieve greater generality. Formally stated, the model is as follows:

$$X_{ijk} = \underset{\text{mu}}{\mu} + \underset{\text{alpha}_i}{\alpha_i} + \underset{\text{beta}_j}{\beta_j} + \underset{\text{gamma}_{ij}}{\gamma_{ij}} + \underset{\text{epsilon}_{ijk}}{\epsilon_{ijk}}.$$

The α_i term represents the contribution of Person i as an actor, a source of behaviors; for instance, in a study of the proportion of time spent speaking in conversations, α_i represents Person i's talkativeness. The β_j term represents Person j's effect as a partner (i.e., the amount of talk that Person j tends to elicit from people when he/she is their partner). Note that these effects can be logically distinguished, although in the case of vocal activity there is a negative correlation between the proportion of time a person tends to talk, and the proportion of time he/she allows other persons to talk when he/she is their partner. For some other social variables, there might conceivably be a positive or zero correlation between the subject-as-actor and subject-as-partner effects. The γ_{ij} term is an interaction effect, representing the special adjustment which Person i makes in level of talkativeness when paired with Person j. As usual, ϵ_{ijk} represents the error term, which picks up variability in behavior at different times.

Looking at the layout in Table 3 may clarify the distinction made between actor and partner effect. The ith row mean represents the average behavior of Person i toward $n - 1$ different partners; thus the row variance indicates whether there are clear-cut and consistent individual differences among persons when their behavior is observed with a number of different partners. The jth column mean represents the average of behaviors of persons who have Person j as a partner; thus the column variance indicates whether there are significant differences among individuals as partners (that is, as social stimuli to which other persons react). These two main effects, row (actor) and column (partner) are both subject variables, and they are in fact based on the same set of subjects; in developing our new round robin model, we will distinguish the actor and partner factors and allow these two factors to be either positively or negatively correlated. Among the other round robin designs, different approaches are used: Bechtel (1971) treats actor and partner factors separately and allows them to be positively correlated; Lev and Kinder (1957) treat actor and partner as separate factors and require that they be independent; Bechtel (1967) and Gleason and Halperin (1975) treat actor and partner as equivalent and combine them into a single subject factor.

This nonindependence between the row and column factor is just one of the special considerations that necessitate the development of a special analysis of variance model; another special problem is the nonindependence of partner behaviors that comes about because social behaviors are mutually contingent. This mutual contingency means that the behavior of j to i is correlated with the behavior of i to j; or, to put it another way, in the layout in Table 3, the cells across the diagonal from each other (the i,j and j,i pairs of cells) are correlated, either positively or negatively. Since this violates the ordinary analysis of variance assumptions, it requires special adaptations in the specification of the model and the derivation of expected mean squares.

Many of the basic assumptions about expectations are unchanged: The expectations of row, column, and interaction effects and the expectation of the error are still zero. Formally, $E(\alpha_i) = 0$, $E(\beta_j) = 0$, $E(\gamma_{ij}) = 0$, and $E(\epsilon_{ijk}) = 0$, for all i, j, and k $(j \neq i)$, where E is the expectation operator. The usual variance specifications are also made:

$$E(\alpha_i^2) = \sigma_\alpha^2; \quad E(\beta_j^2) = \sigma_\beta^2; \quad E(\gamma_{ij}^2) = \sigma_\gamma^2;$$

and

$$E(\epsilon_{ijk}^2) = \sigma_\epsilon^2, \quad \text{for all } i, j, \text{ and } k, \ (j \neq i).$$

The sources of nonindependence outlined earlier (the correlation between row and column factors and between pairs of cross-diagonal cells) must also be incorporated into the model specifications. The covariance between Person i's effect as an actor (α_i) and Person i's effect as a partner (β_i) is formally

Table 4
Summary of Round Robin Model Specifications

$$E(\alpha_i) = 0, \quad E(\alpha_i^2) = \sigma_\alpha^2 \quad \text{for all } i$$
$$E(\beta_j) = 0, \quad E(\beta_j^2) = \sigma_\beta^2 \quad \text{for all } j$$
$$E(\gamma_{ij}) = 0 \quad \text{for all } i \text{ and } j,\ j \neq i$$
$$E(\epsilon_{ijk}) = 0 \quad \text{for all } i,\ j, \text{ and } k$$

$$E(\alpha_i, \beta_j) = \begin{cases} \rho_1 \sigma_\alpha \sigma_\beta & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

$$E(\gamma_{ij}, \gamma_{i'j'}) = \begin{cases} \sigma_\gamma^2 & \text{if } i = i',\ j = j' \\ \rho_2 \sigma_\gamma^2 & \text{if } i' = j,\ j' = i \\ 0 & \text{otherwise} \end{cases}$$

$$E(\epsilon_{ijk}, \epsilon_{i'j'k'}) = \begin{cases} \sigma_\epsilon^2 & \text{if } i = i',\ j = j',\ k = k' \\ \rho_3 \sigma_\epsilon^2 & \text{if } i' = j,\ j' = i,\ k = k' \\ 0 & \text{otherwise} \end{cases}$$

stated as $E(\alpha_i, \beta_i) = \rho_1 \sigma_\alpha \sigma_\beta$, hereafter referred to as the row and column covariance.

The covariance between pairs of cross-diagonal cells is more complex, since it consists of two parts. First, it is reasonable to assume that there is some correlation between γ_{ij} and γ_{ji} (that is, between i's special adjustment to j and j's special adjustment to i). This nonzero covariance between the interaction effects within dyads is represented as $\rho_2 \sigma_\gamma^2$; formally, the nonindependence assumption requires that we specify $E(\gamma_{ij}, \gamma_{ji}) = \rho_2 \sigma_\gamma^2$. This term will be referred to as the cross-diagonal interaction covariance, and it will be interpreted as evidence for some kind of lasting reciprocity of behaviors within dyads. Furthermore, it is likely that the error terms for the behavior of social interaction partners on a particular day or time will also be correlated; formally, $E(\epsilon_{ijk}, \epsilon_{jik}) = \rho_3 \sigma_\epsilon^2$; this will be interpreted as evidence for a situation-specific reciprocity in partner behaviors. All other covariances are assumed to be zero; the model specifications outlined in this section are summarized in Table 4.

For some types of social interaction data, there may be a further source of nonindependence among observations; in person perception data, if Persons h and j discuss their perceptions of Person i, then the h, i and j, i ratings may be correlated. In the vocal activity data presented as an illustration in this paper, this additional source of nonindependence is assumed not to occur, and some care should be taken to minimize this type of nonindependence in designing studies that are to be analyzed using round robin designs.

Before outlining the calculations and parameter estimation procedures, we should note the alternative model specifications briefly described in Table 5; depending on the nature of the data, one of the previously developed models may provide more powerful means of analyzing some data. This new model has the least restrictive assumptions of any of these designs.

Calculation of Means and Mean Squares

The strategy adopted for the purpose of developing the new model was as follows: First, we set up formulae for the sample means and mean squares; next, taking into account the empty diagonal cells and the nonindependence assumptions included in the model specifications, we worked out the expected mean squares; finally, we used the expected mean square equations to solve for the variance components of the model (e.g., σ_α^2, σ_β^2, σ_γ^2) in terms of the sample mean squares. Significance testing on these variance components was done by jackknifing, since ordinary F ratios are not easily set up. After these procedures have been outlined in detail, the model will be demonstrated using vocal activity data.

The computation of sample means is straightforward; the only special consideration is the presence of empty cells in the diagonal.

Grand mean:

$$M_{...} = \frac{1}{rn(n-1)} \sum_i \sum_{j \neq i} \sum_k X_{ijk}.$$

Row mean:

$$M_{i..} = \frac{1}{r(n-1)} \sum_{j \neq i} \sum_k X_{ijk}.$$

Column mean:

$$M_{.j.} = \frac{1}{r(n-1)} \sum_{i \neq j} \sum_k X_{ijk}.$$

Cell mean:

$$M_{ij.} = \frac{1}{r} \sum_k X_{ijk}.$$

As usual, the estimate of the grand mean is $M_{...}$. Because of the empty diagonal cell, $M_{i..} - M_{...}$ is not an unbiased estimate of the row effect (α_i); the following argument will explain why the empty diagonal cells produce bias in the row means of this layout.

When you sum across row i, for instance, you get an average of Person i's behavior with $n - 1$ partners, not including self. This means that the estimate of Person i's talkativeness relative to the talkativeness of the other persons in the sample is biased, since Person i is not observed with the same set of partners that all the other persons had. Since we can get an estimate of Person i's effect as a partner (β_i), however, it is possible to correct for this "missing partner" bias. For instance, in the data on proportion of time spent speaking in conversations, any particular observation (X_{ij}) depends upon the talkativeness of i, the person being observed, and the amount of talk elicited by j, the partner. Typically, the amount of talk elicited by j depends on j's talkativeness. If Person i is the most talkative person in the sample, the the fact that i lacks himself or herself as a partner means the estimate of his/her talkativeness is slightly inflated, since he/she had the last talkative set of partners of any individual in the sample. An analogous argument holds for bias in the column means. The estimates for the row effects and the column effects are as follows:

$$\hat{\alpha}_i = \frac{(n-1)^2}{n(n-2)} M_{i\cdot\cdot} + \frac{(n-1)}{n(n-2)} M_{\cdot i\cdot} - \frac{(n-1)}{(n-2)} M_{\cdot\cdot\cdot},$$

$$\hat{\beta}_j = \frac{(n-1)^2}{n(n-2)} M_{\cdot j\cdot} + \frac{(n-1)}{n(n-2)} M_{j\cdot\cdot} - \frac{(n-1)}{(n-2)} M_{\cdot\cdot\cdot}.$$

These equations make sense in light of the previous discussion; to get an estimate of Person i's true row effect, it is necessary to correct for the "missing partner" bias by adjusting by some fraction of Person i's column mean. As usual, once the row and column effects are known, estimates of the interaction effects can be obtained by subtraction:

$$\hat{\gamma}_{ij} = M_{ij\cdot} - M_{\cdot\cdot\cdot} - \hat{\alpha}_i - \hat{\beta}_j.$$

The calculations for mean square row and mean square column are straightforward, but the coefficients are slightly different; note that the total number of observations in the design is $rn(n-1)$ and the number of observations in a row or column is $r(n-1)$.

$$MS_{\text{row}} = \frac{r(n-1)}{(n-1)} \sum_i (M_{i\cdot\cdot} - M_{\cdot\cdot\cdot})^2.$$

$$MS_{\text{column}} = \frac{r(n-1)}{(n-1)} \sum_j (M_{\cdot j\cdot} - M_{\cdot\cdot\cdot})^2.$$

Before going further, notice how the empty cells and the special nonindependence assumptions affect the variance components that are estimated by the row mean square. For $M_{i\cdot\cdot}$ and $M_{\cdot\cdot\cdot}$, we have:

$$M_{i\cdot\cdot} = \mu + \alpha_i - \frac{\beta_i}{(n-1)} + \bar{\gamma}_{i\cdot} + \bar{\epsilon}_{i\cdot\cdot}$$

$$M_{\cdot\cdot\cdot} = \mu + \bar{\alpha}_{\cdot} + \bar{\beta}_{\cdot} + \bar{\gamma}_{\cdot\cdot} + \bar{\epsilon}_{\cdot\cdot\cdot}$$

When these expressions are used to derive the expected mean square for row effect, the special assumptions incorporated into the model specifications mean that many cross-multiplication terms whose expected values would be zero in an ordinary analysis of variance now have nonzero expectations; for instance, since α_i is correlated with $\bar{\beta}_{\cdot}$, there will be a nonzero covariance term picked up when the α_i term in the equation for $M_{i\cdot\cdot}$ is multiplied by the $\bar{\beta}_{\cdot}$ term in the equation for $M_{\cdot\cdot\cdot}$ and their expectation is taken. The expected mean square for row in the round robin model includes the following terms.

$$EMS_{\text{row}} = r(n-1)\sigma_\alpha^2 + \frac{r}{(n-1)}\sigma_\beta^2 + r\sigma_\gamma^2 - 2r\rho_1\sigma_\alpha\sigma_\beta - \frac{r}{(n-1)}\rho_2\sigma_\gamma^2 + \sigma_\epsilon^2 - \frac{1}{n-1}\rho_3\sigma_\epsilon^2,$$

where $\rho_1\sigma_\alpha\sigma_\beta$ is the Row × Column covariance, $\rho_2\sigma_\gamma^2$ is the cross-diagonal interaction covariance, and $\rho_3\sigma_\epsilon^2$ is the cross-diagonal error covariance.

The complete derivation for the expected mean squares will not be presented here since it is rather lengthy; a copy of the derivation is available on request (Stoto, Kenny, & Warner, Note 1).

The number of untraditional variance components included in this expected mean square suggests that it will not be easy to find an appropriate error term that will make it

Table 5
Summary of Round Robin Design Specifications

Study	Actor/ partner factors	Relationship of cross diagonal cells	Relationship of cells within same row or column	Significance testing	Special features
Lev & Kinder (1957)	Separate and independent	Independent	Independent	F	—
Bechtel (1971)	Separate and positively correlated	Positively correlated	Independent	F	Tests for identity of row and column factors, and symmetry of interaction effects
Gleason & Halperin (1975)	Combined into single factor	Positively correlated	Positively correlated	Pseudo F	Split plot, with all subjects within each condition paired round robin style; main emphasis on treatment rather than subject variables
Warner, Kenny & Stoto (1979)	Separate, either positively or negatively correlated	Positively or negatively correlated	Independent	t (jackknifing)	—

possible to set up an F ratio to evaluate the significance of the row variance. Instead, we decided to solve directly for estimates of the variance components themselves; this requires seven equations in seven unknowns. Some additional sample mean squares are therefore needed, and the nature of these mean squares can be guessed from the set of unknowns appearing in EMS_{row}. The new mean squares are a mean square for row and column covariance (denoted $MS_{\text{Row}\times\text{Col}}$); a mean square for the covariance of cross-diagonal interaction effects (denoted $MS_{\text{Cell}\times\text{Cell}}$); and a mean square for the covariance of cross-diagonal errors (denoted $MS_{\text{Error}\times\text{Error}}$). Also, in setting up the sample mean squares, we have chosen to calculate a mean square for cells that is not the usual mean square for interaction; this was done because it is easier to evaluate the expected mean square for cells. The set of mean squares to be computed for the round robin layout are as follows:

$$MS_{\text{row}} = r\sum_i (M_{i..} - M...)^2,$$

$$MS_{\text{column}} = r\sum_j (M_{.j.} - M...)^2,$$

$$MS_{\text{cell}} = \frac{r}{n(n-1)-1}\sum_i\sum_{j\neq i}(M_{ij.} - M...)^2,$$

$$MS_{\text{Row}\times\text{Col}} = r\sum_i (M_{i..} - M...)(M_{.i.} - M...),$$

$$MS_{\text{Cell}\times\text{Cell}} = \frac{r}{n(n-1)-1}\sum_i\sum_{j\neq i}(M_{ij.} - M...) \times (M_{ji.} - M...),$$

$$MS_{\text{error}} = \frac{1}{n(n-1)(r-1)}\sum_i\sum_{j\neq i}\sum_k (X_{ijk} - M_{ij.})^2,$$

and

$$MS_{\text{Error}\times\text{Error}} = \frac{1}{n(n-1)(r-1)}\sum_i\sum_{j\neq i}\sum_k (X_{ijk} - M_{ij.}) \times (X_{jik} - M_{ji.}).$$

The coefficients for the variance components in the seven expected mean square equations corresponding to these sample mean squares are summarized in Table 6.

We now have the information necessary to solve for the estimates of the seven variance components. Let **M** be the (7×1) vector of

sample mean squares: (MS_{row}, MS_{col}, MS_{cell}, $MS_{Row\times Col}$, $MS_{Cell\times Cell}$, MS_{error}, $MS_{Error\times Error}$). Let $\mathbf{V}$ be the (7×1) vector of variance components to be estimated:

$$(\sigma_\alpha^2, \sigma_\beta^2, \sigma_\gamma^2, \rho_1\sigma_\alpha\sigma_\beta, \rho_2\sigma_\gamma^2, \sigma_\epsilon^2, \rho_3\sigma_\epsilon^2).$$

Let $\mathbf{C}$ be the (7×7) matrix of coefficients from the expected mean squares equations, as in Table 6.

By definition, $\mathbf{M} = \mathbf{CV}$. Therefore we can solve for estimates of the variance components by multiplying both sides by the inverse of the coefficient matrix, to get $\mathbf{V} = \mathbf{C}^{-1}\mathbf{M}$.

One can of course consider various special cases of the model as $\sigma_\alpha^2 = \sigma_\beta^2$ and $\rho_1 = 1$; $\rho_2 = 0$; $\rho_2 = 1$; $\rho_2 = -1$; $\rho_3 = 1$; or $\rho_3 = -1$. For any of the above assumptions one can simply drop the relevant mean square and change the coefficient matrix accordingly. In the case of $\sigma_\alpha^2 = \sigma_\beta^2$ and $\rho_1 = 1$, a sensible strategy would be to combine MS_{row} and MS_{column}, and drop $MS_{Row\times Column}$.

In order to ensure that all the parameters in the model are identified, it is necessary to have a minimum of four subjects. The model may be applied where $r = 1$ (each dyad is observed only once), but in this case it will not be possible to distinguish the interaction variance (σ_γ^2) from the error variance (σ_ϵ^2), or the cross-diagonal interaction covariance ($\rho_2\sigma_\gamma^2$) from the cross-diagonal error covariance ($\rho_3\sigma_\epsilon^2$).

Significance Testing

Once estimates of the variance components have been obtained, it is necessary to have some means of evaluating whether they differ significantly from zero.[2] A useful strategy is jackknifing (Mosteller & Tukey, 1977). The idea is to estimate the variance of a parameter by examining the empirical distribution of estimates of that parameter (such as σ_α^2), since there is no simple way to determine the variance of its theoretical distribution. For σ_α^2, for instance, this can be done as follows: First, create n different subsets of the data, each subset consisting of the full data matrix, omitting the data for one subject (i.e., one row and the corresponding column). For each of these leave-out-one subsets, we can estimate a value for σ_α^2. Now, these n different estimates of σ_α^2 are not independent, since the subsets overlap in membership; but Mosteller and Tukey have devised a way around this problem. A set of pseudoestimates is generated from the leave-out-one estimates, and the pseudoestimates can be treated as if they were independent. Let Y_{all} be the value of σ_α^2 for the whole group of n subjects; let $Y_{(j)}$ be the value of σ_α^2 for the subset of $n - 1$ subjects created by omitting person j. Then the pseudoestimates are given by $Y_{*j} = n(Y_{all}) - (n - 1)Y_{(j)}$, for $j = 1, 2, \ldots, n$. The Y_{*j} pseudoestimate is a "fake" estimate of σ_α^2, as if we had been able to estimate σ_α^2 based on only subject j; using this set of pseudoestimates, we can estimate the standard error

[2] To see why F ratios are difficult to set up, consider the expected mean squares associated with the ratio of mean square rows and mean square error. It is clear that this ratio would not be an unbiased estimate of the row effect variance.

Table 6
Expected Mean Square Coefficients

MS	df	σ_α^2	σ_β^2	σ_γ^2	$\rho_1\sigma_\alpha\sigma_\beta$	$\rho_2\sigma_\gamma^2$	σ_ϵ^2	$\rho_3\sigma_\epsilon^2$
Row	$n-1$	$r(n-1)$	$\frac{r}{n-1}$	r	$-2r$	$-\frac{r}{n-1}$	1	$-\frac{1}{n-1}$
Column	$n-1$	$\frac{r}{n-1}$	$r(n-1)$	r	$-2r$	$-\frac{r}{n-1}$	1	$-\frac{1}{n-1}$
Cell	$n(n-1)-1$	$\frac{r(n-1)^2}{n(n-1)-1}$	$\frac{r(n-1)^2}{n(n-1)-1}$	r	$\frac{-2r(n-1)}{n(n-1)-1}$	$\frac{-r}{n(n-1)-1}$	1	$-\frac{1}{n(n-1)-1}$
Row × Column	$n-1$	$-r$	$-r$	$-\frac{r}{n-1}$	$\frac{r[(n-1)^2+1]}{n-1}$	r	$-\frac{1}{n-1}$	1
Cell × Cell	$n(n-1)-1$	$-\frac{r(n-1)}{n(n-1)-1}$	$-\frac{r(n-1)}{n(n-1)-1}$	$\frac{-r}{n(n-1)-1}$	$\frac{2r(n-1)^2}{n(n-1)-1}$	r	$-\frac{1}{n(n-1)-1}$	1
Error	$n(n-1)(r-1)$	0	0	0	0	0	1	0
Error × Error	$n(n-1)(r-1)$	0	0	0	0	0	0	1

of σ_α^2 by the variance of the Y_{*j} values and use this to set up a t test (with $n - 1$ degrees of freedom) to evaluate whether $\sigma_\alpha^2 = 0$. Let y_* be the mean of the Y_{*j} pseudovalues, and let s^2 be the variance of the Y_{*j} pseudovalues. Then $(n^{\frac{1}{2}}y_*)/s$ is distributed as t with $n - 1$ degrees of freedom if $\sigma_\alpha^2 = 0$.

This jackknifing procedure can be used to test the significance of each of the variance and covariance components estimated for the model. For justification of the procedure and further examples of its applications, see Mosteller and Tukey, 1977 (chap. 8).

Interpretation of the Variance and Covariance Components

Since the components of this model differ from those of an ordinary analysis of variance, some comments on the meaning that the variance components have for various types of social interaction data may be helpful. Notice first that three variance components (actor, partner, and interaction) correspond to the components of an ordinary two-way analysis of variance and that three covariance components (row and column, cross-diagonal interaction, and cross-diagonal error) provide information somewhat similar to that of an intraclass correlation design.

Actor effect. This factor is related to individual differences among persons as sources of social behaviors; depending on the dependent variable, this factor can be renamed speaker, rater, perceiver, and so forth. Since each individual's behavior is observed over a number of different partners, the round robin analysis of variance provides a fairly strict test for stability; the actor effect cannot be large unless individuals are fairly consistent in their behavior, regardless of the person with whom they are paired.

Partner effect. This factor indicates whether there is a strong tendency for individuals to elicit particular types of behavior from other persons—for instance, high or low trait ratings, high or low level of attraction, high or low amounts of speech activity, and so forth.

Notice that it is possible to have a strong actor effect and a weak partner effect (this occurs in the vocal activity data presented in this paper as an illustration of the model) or a strong partner effect and a weak actor effect (as we have seen in preliminary work with person perception data, where the subjects are being trained to apply an objective rating system; there are small differences among ratings given by different raters and large differences in the ratings received by different individuals). Thus, although they are often highly correlated, it is important to stress again the logical distinction between these factors, and in some situations, particularly in person perception research, it is interesting to compare the magnitude of these two main effects.

Interaction effect. This component relates to the particular adjustment that each person makes to each particular partner. In vocal activity research, for instance, interaction effect indicates whether Person i paired with Person j consistently talks a little more (or a little less) than would be expected on the basis of i's talkativeness and j's tendency to elicit talk. This interaction effect also has meaning in person perception and attraction research—does Person i tend to like Person j more than would be expected on the basis of i's average liking and j's average tendency to inspire liking? The interaction effect picks up any tendency to make adjustments that are unique to each partner and hold up across repeated pairings with that partner. In a sense the interaction term measures what is unique to the interaction between partners.

Row and column covariance. If this covariance is very large relative to the row and column covariances, and if it is positive, it suggests that the row and column factors are nearly indistinguishable, in which case a single factor design (one that combines actor and partner into a single subject factor) may be more appropriate. In some situations, the row and column covariance may have a substantive interpretation; if the dependent variable is related to utility (for instance, the favorableness of a rating or the amount of some resource given by one person to another), then the relationship between the row factor (amount Person i gives out, on the average) and the column factor (amount Person i receives, on the average) may be viewed as a kind of "equity" factor. Another theoretical construct that comes close to describing this

factor in person perception is *projection*; a high row and column covariance could be taken to mean that the ratings a person gives to others are closely tied to his or her own traits (as perceived by others). For the vocal activity data, there is no simple interpretation of this covariance.

Cross-diagonal interaction covariance. This is one of the two dyadic reciprocity factors in the round robin design; it should be distinguished from the cross-diagonal error covariance, which picks up a different type of reciprocity. A large cross-diagonal interaction covariance means that for each of the possible (i, j) dyads, the systematic and enduring adjustment that i makes to j is correlated with the enduring adjustment that j makes to i. For interpersonal attraction data, it seems likely that this could be a strong relationship; that is, if i always tends to like j more than expected, it seems probable that in return j always tends to like i more than expected. The reciprocity picked up by this covariance is an enduring reciprocity, that is, a relationship between partner behaviors that holds up across different times or situations.

Cross-diagonal error covariance. This is a situation-specific reciprocity effect. If this covariance is large, then the behaviors of the partners are highly correlated in any particular situation, although they may not necessarily be related in the same way across different situations. For instance, in the vocal activity data, we will find a large cross-diagonal error covariance—since the proportion of time available for speaking in any particular conversation is approximately zero sum, the vocal activity of partners in specific conversations is highly negatively correlated. However, the reciprocity turns out not to be enduring—that is, although the sum of activity is approximately 1 in any particular conversation, the allocation of time between partners may be worked out differently in different conversations—in one, Person i may do more than his or her share of the talking; in the next, Person j may do more than his or her share. A fanciful example illustrating the independence of the situation-specific and the enduring reciprocity factors might be found in a soap opera, where the attractions and antipathies among the characters may be highly correlated and mutual in any one episode, but the pattern of attractions may shift from one week to the next—indicating strong situation-specific reciprocity and a lack of enduring reciprocity.

Summary

Clearly, the interpretation of the variance and covariance components in the round robin model depends upon the nature of the dependent variable and the type of substantive or theoretical questions that arise about the reciprocity of behaviors. The goal of this section is to show the flexibility of the new round robin model; there are some readily understandable parallels between well-known theoretical constructs and the factors in the round robin design. The following section will demonstrate the application of the round robin design to vocal activity data.

Empirical Illustration

Various aspects of speech activity in conversations and interviews have been singled out for study, for example, mean duration of vocalizations or pauses, probability of initiating or maintaining speech, and proportion of time spent speaking; these vocal activity parameters are interrelated, and certain aspects of vocal activity are highly consistent for individual speakers (Jaffe & Feldstein, 1970). Of these parameters, the simplest one to study is the proportion of time spent speaking, or speaker activity level. Past research on individual speaker consistency has generally involved at most two or three different partners; the round robin design provides a natural method of examining individual differences across many different partners in order to evaluate parametrically whether the reliable individual speaker differences claimed on the basis of this earlier research actually exist.

A round robin study of proportion of time spent speaking was conducted by Warner (Note 2). Eight participants (four male, four female) were enlisted for a study of conversation. Each of the 28 possible pairs conversed privately on 3 separate days for about 12 to 15 minutes each time. Each speaker's voice was recorded onto a separate channel of a

Table 7
Proportion of Time Spent Speaking in Conversations

Subject	Subject 1	2	3	4	5	6	7	8	Row *M*
1		55	65	59	62	87	69	81	67
		58	58	54	68	68	76	76	
		45	85	50	79	76	56	71	
2	22		21	34	25	35	39	58	38
	50		42	27	38	35	36	53	
	36		35	38	34	49	53	28	
3	31	56		34	31	24	34	36	35
	45	45		40	33	25	48	16	
	12	54		24	37	42	37	30	
4	26	62	68		46	31	58	72	55
	30	62	55		25	89	43	83	
	45	60	81		65	33	59	65	
5	66	56	70	67		36	58	73	61
	70	56	89	69		54	53	56	
	52	58	64	44		37	71	71	
6	27	30	74	45	49		70	40	48
	49	53	81	11	35		78	69	
	28	41	47	48	37		46	41	
7	52	40	77	32	52	49		67	50
	45	66	54	40	48	22		65	
	67	34	60	48	23	49		61	
8	39	47	75	47	37	45	37		55
	59	80	90	47	51	46	67		
	45	70	73	49	37	51	51		
Column *M*	43	54	65	43	43	47	54	58	51

Note. X_{ijk} = the proportion of time spent speaking by Person i to Person j on Day k (in percent).

stereo tape. The participants wore headsets with noise-canceling microphones close to the mouth; this type of microphone arrangement, which was somewhat intrusive, was necessary because ordinary microphones allow too much spillage of voices between channels. The proportion of time spent speaking by each person was determined by using a computer voice-operated relay to detect the presence or absence of speech in each of the two channels of the tape-recorded conversations. The voice-operated relay was calibrated by means of an indicator light that showed when the system was detecting speech activity; thus a human listener could manipulate the threshold and other settings on the voice-operated relay until the pattern shown by the light matched the perceived on–off pattern of speech activity.

The on–off vocal activity judgment was made twice per second, and this information was used to determine the proportion of time spent speaking by each person over the time of the whole conversation.

Subjects were instructed that the study was about the process of becoming acquainted and were told that they could talk about whatever they liked. Pilot tests had indicated that this was probably much more natural than imposing some kind of task; most dyads had little difficulty in carrying on a conversation, and a number of conversations were quite animated and contained information of a rather personal nature. Subjects spent their free time between recording sessions in a lounge where they were free to talk.

An effort was made to promote a feeling of ease in the situation, at the cost of experimental control; for example, several persons in the study were previously acquainted, some of the conversations that took place for the tape recorder were continuations of conversations that had begun in the lounge, and many conversations continued after the tape recorder was turned off. Partly because of the marathon nature of the scheduling (the eight subjects spent three 8-hour days participating in the study), a certain amount of group cohesiveness developed. For all these reasons, it seems likely that the conversations reported in this study are more natural than those that have been elicited from persons who come into the laboratory "cold" for just one session or who converse via intercom without visual contact.

The proportion of time spent speaking was tabulated for all 84 conversations, and the results are displayed in the $8 \times 8 \times 3$ data matrix shown in Table 7.[3]

The O-Bird round robin analysis of variance program, which carries out all the calculations, parameter estimation, and significance testing as outlined in this paper, was run on these data.[4] The results are displayed in a modified source table, with t tests in place of the usual F ratios. Since the true variance components

[3] No transformation for proportional data (such as arcsine) was used, since none of the proportions were very near zero or one.

[4] The O-Bird program (in Fortran) is available from the second author.

(row, column, interaction) must be nonnegative, a one-tailed t was used for these three tests; the covariances (row and column, cross-diagonal interaction, and cross-diagonal error) can be either positive or negative, and so a two-tailed t is needed for these three tests.[5]

Discussion

There were two significant effects: (a) clear-cut individual speaker differences in activity level, as shown by the significant t value for row effect and (b) strong situational reciprocity between the speech of partners in particular conversations, as indicated by the significant t value for the cross-diagonal error covariance. No other effects were statistically significant (although the sizes and signs of all other effects were generally consistent with our expectations, which lends some additional plausibility to the model). Our results confirm earlier findings of individual differences; we also have information about the cross-diagonal error convariance, which can be interpreted as evidence of a situation-specific reciprocity of speech activity level. This was anticipated, since it is a common sense observation that the time allocation within any particular conversation is approximately zero sum, that is, the more time taken up by one speaker, the less time available to the other speaker. Thus, the proportions of time spent speaking by the two persons in a particular conversation (X_{ijk}, X_{jik}) are negatively correlated.

The partner effect was smaller than the speaker effect, which would suggest that the impact of partner activity level on an individual's speech production is not as great as the effect of his or her own "preferred" activity level. As anticipated, there was a negative covariance between the activity level of a person and that person's effect on other prople's activity level when he or she was their partner; however, this covariance was not significant, so it seems appropriate to conclude that the relationship between an individual's activity level and his or her tendency to elicit a high or low activity level from others is not so strong that these two factors should be considered equivalent.

There was a relatively modest and nonsignificant interaction effect. This can be rationalized as follows: although it is true that in any particular conversation between persons i and j they adjust their activity levels to each other to achieve a total proportion of time active of about 1.0, this adjustment can be made in a number of different ways, and the adjustment that they make in any one particular conversation is unique to that conversation and does not carry over to later conversations (at least, not in our study, which considers only three conversations per dyad). That is, in one conversation, i may talk a little more than would be expected (based on i's talkativeness and j's tendency to elicit talk); in another conversation, i may talk less than would be expected. The small interaction variance indicates that there is not a fixed adjustment for each dyad, so that when i talks to j, he or she always talks a little more (or less) than usual.

The cross-diagonal interaction covariance for these data is virtually zero. This is reasonable; if i's adjustment to j is not stable or fixed across different situations, it is not reasonable to expect a strong relationship between i's average adjustment to j, and j's average adjustment to i. Furthermore, even if these partner adjustments were consistent, as would be indicated if there were a large interaction effect, the adjustments in activity level might not necessarily be correlated for partners. Any combination is possible —depending upon the "preferred" activity levels of i and j, both partners could talk more, both partners could talk less, i could talk more and j could talk less, or j could talk more and i could talk less; any of these types of adjustment could conceivably be needed in order to work out the time allocation between partners. Since all these combinations can and do occur, on the average the covariance between i's adjustment to j and j's adjustment

[5] When testing whether a variance component is zero in the population, a one-tailed t is appropriate, since the true variance component in the population cannot be negative. Since the pseudovalues can be negative (and will be negative about half the time if the true variance component being estimated is zero), there is no reason why these pseudoestimates could not be t distributed. Of course, the covariance terms can be either positive or negative, so two-tailed ts should be used for the covariance terms.

Table 8
Source Table for Round Robin Analysis of Variance

Source	*MS*	Variance or covariance estimate	*t*(7)
Row (speaker)	2438.3167	91.9680	2.0144*
Column (partner)	1388.1423	40.9178	1.2669
Interaction	622.8201	29.7534	1.0262
Row and column	−1380.7346	−40.3749	−1.1547
Cell × Cell	−354.1238	4.0811	0.0997
Error × Error	−95.5282	−95.5282	−2.4500**
Error	146.0462	142.0462	—

*$p < .05$, one-tailed $t(7)$. **$p < .05$, two-tailed $t(7)$.

to i will be nearly zero. The small (virtually zero) covariance obtained for this cross-diagonal interaction covariance indicates a lack of enduring reciprocity in speech activity level—there is no adjustment that i always makes to j that is correlated with an adjustment that j always makes to i. It seems likely that certain other vocal activity parameters (such as mean switching pause duration) and other variables such as attraction, might show enduring reciprocity; but there seems to be no such effect for activity level of speakers.

The significant individual speaker differences, taken together with the nonsignificant interaction effect, suggest that the appropriate unit of analysis for proportion of time speaking is the individual person rather than the dyad. The strong situational reciprocity, together with the lack of enduring reciprocity, suggests that the reciprocity of speech activity levels depends heavily on situational factors—perhaps topic of conversation, time of day, moods of the participants, and so forth, rather than on some lasting "agreement" between partners as to their relative dominance or right to claim speaking time. At present the model is not set up to handle order effects, but it could be extended to account for predictable shifts in the time allocation between partners over time (if these occur).

Although the main focus of this discussion is interpretation of the significant effects, there are additional aspects of the results in Table 8 that bear further examination. It is possible to estimate correlations from the mean squares:

$$\hat{r}_1 = \frac{MS_{\text{Row}\times\text{Column}}}{(MS_{\text{row}} * MS_{\text{column}})^{\frac{1}{2}}} = -.7504,$$

and

$$\hat{r}_2 = \frac{MS_{\text{Cell}\times\text{Cell}}}{MS_{\text{cell}}} = -.5686.$$

The $\hat{r}_1$ is the correlation estimate for the relationship between the row and column factors that would be obtained directly from the raw data. It is a biased estimate of the true relationship between row and column factors; this can be seen by inspecting the set of variance components belonging to $EMS_{\text{Row}\times\text{Column}}$, EMS_{row}, and EMS_{column}. The ratio of these raw mean squares clearly does not provide an estimate of the true correlation ρ_1. The same problem applies to the estimate for r_2, which is also biased; however, no such problem arises in estimating ρ_3, which is just $(MS_{\text{Error}\times\text{Error}})/MS_{\text{error}}$.

An alternative means of estimating the correlations is to take ratios of the variance and covariance components derived from the round robin model, as follows:

$$\hat{\rho}_1 = \frac{\widehat{\rho_1\sigma_\alpha\sigma_\beta}}{(\hat{\sigma}_\alpha{}^2\hat{\sigma}_\beta{}^2)^{\frac{1}{2}}} = -.6582,$$

$$\hat{\rho}_2 = \frac{\widehat{\rho_2\sigma_\gamma{}^2}}{\hat{\sigma}_\gamma{}^2} = .1372.$$

Although the components of these new correlation estimates are unbiased, our initial experiences with them indicate that they may be unreliable. Also, just as estimates for the variance components of the round robin model can be less than zero due to sampling error, correlation estimates derived from the variance and covariance components may fall outside the range of plus or minus one, especially if the denominators are small. Thus some caution is in order when using or interpreting these correlation estimates. It should be kept in mind that correlations based on the raw data in a round robin–type layout do not estimate what they seem to be estimating. Such raw correlations are commonly reported in person perception literature; their interpretation is highly problematic.

Another useful strategy is comparing the magnitude of various effects. A simple instance where this may be substantively interesting is in person perception research, where it may be useful to know whether there is greater variability among the ratings that people give (rater or actor effect) or the ratings that people receive (partner effect); that is, is the trait in question "in the eye of the beholder" (influenced mainly by the person doing the rating)? Or is it a trait that different observers can agree upon, influenced mainly by characteristics of the person being rated? This comparison can be made by looking at the ratio (or the difference) between the row and column variance components, and the significance of the ratio or difference can be evaluated by jackknifing. This comparison of actor and partner effects was not significant for the vocal activity data, $t(7) = .635$. That is, even though the row variance was significantly greater than zero and the column variance was not significantly greater than zero, the row variance was not significantly greater than the column variance.

More elaborate comparisons among the variances can also be constructed. For example, for some applications it might be instructive to ask to what extent the variability of social behavior is accounted for by characteristics of the individuals (actor and partner effects) as opposed to characteristics of the dyad or the situation (interaction and error). To the extent that actor and partner factors predominate, social behavior can be viewed as additive; social interaction can be conceived of as a linear system, that is, a system with behavior that can be predicted from the behavior of its components. Such a test of additivity might take the following form:

$$A = \frac{\sigma_\alpha^2 + \sigma_\beta^2}{\sigma_\gamma^2 + \sigma_\epsilon^2}.$$

For our data, this ratio was significantly greater than zero, $t(7) = 2.44$, $p < .025$ one-tailed, based on the jackknifing of this ratio over different subsets of the data; that is to say, some nonnegligible amount of the variability in vocal activity level is accounted for by the components of the conversational system, the speakers themselves. Depending upon the nature of the research problem, other comparisons among variances can be set up.

Special Applications

As with any statistical model, the decision to use the round robin model depends on two considerations—first, whether the assumptions of the model are satisfied, and second, whether the model provides the desired information. It is possible to use the round robin model to analyze data on unconstrained social interactions in natural settings or to make group comparisons; however, it is important to realize that these data may violate certain assumptions of the round robin model.

Consider the type of study in which social interactions are observed in natural settings such as classrooms. One might count the number of aggressive or attention-getting acts by each child. Since such data violate the independence assumptions of ordinary chi-square and analysis of variance, one should not use these tests. Clearly each act within a given classroom is dependent to some extent on the other acts that occur within that classroom. This might suggest that the round robin model is suitable for these data; in addition to reciprocity effects, however, there are likely to be other forms of social influence such as modeling or shared attitudes toward particular class members. In postulating the round robin model, we assume that (except for the reciprocity covariances) the observations are independent, for example, h's behavior toward Person i is assumed to be uncorrelated with j's behavior toward i. Clearly if Person j imitates h's behavior or shares h's attitudes toward other individuals, the behaviors of h and j will be correlated. This means that the independence assumptions of the round robin model are violated. It may be possible to minimize these other forms of social influence by preventing the participants from observing the interactions of other dyads and preventing participants from discussing their perceptions and attitudes. In many natural social environments, however, this kind of control is not feasible, and the use of such stringent controls may sacrifice too much external validity.

We know of no statistical test of the covariance structure of the data to evaluate whether the independence assumptions of the round robin model are satisfied. Such a test could conceivably be developed along the lines of

work done by Huynh and Feldt (1970). In the absence of statistical criteria, the investigator must realize that modeling and shared expectations will bias the estimates of the round robin model parameters.

Another difficulty in the analysis of interaction data from unconstrained social situations is that participants self-select their partners and may not interact with everyone else in the group. This results in large numbers of missing observations or zero frequencies. At present the analysis we propose cannot handle missing data. The presence or absence of ties between persons reveals network structure, however, so block models or network analysis are viable alternatives (White & Breiger, 1975).

Another possible application involves group comparisons: either comparisons of two or more round robin layouts or comparisons of subsets of subjects within a round robin layout. For instance, the vocal activity study reported earlier included four male and four female participants. One might wish to examine differences in talkativeness between males and females or between same sex and cross-sex dyads. This might be done by taking the mean difference in the row effect estimates (α_i) for males versus females in such of the leave-out-one subsets of data and jackknifing to get a significance test. One might also wish to correlate the actor effect (α_i), talkativeness, with a personality scale such as dominance.

Another consideration in applications of the round robin model is sample size. The round robin analysis requires a minimum of four persons, and the jackknife procedure raises this to five. In fact, we believe that the analysis will provide unstable estimates for sample sizes less than eight. Recall that if there are only seven members of the group, there are only six degrees of freedom in the t test of the actor and partner main effects. To increase the efficiency in estimation, one can replicate the round robin design on different sets of persons. The parameter estimates could then be pooled across the different groups, resulting in more stable estimates.

Summary

The new round robin analysis of variance provides a tool for dealing with social interaction data that allows assessment of both individual differences and reciprocity in social behaviors. It readily lends itself to the substantive problems encountered in social psychological research on person perception, attraction, vocal activity and other mutually contingent behaviors. This means that instead of regarding the mutual contingency of social behaviors as a problematic deviation from the independence assumptions of statistical models traditionally used in the nonsocial sciences, it is possible to explicitly incorporate mutual contingency into the design and to treat it as an interesting effect in its own right. By tailoring statistical models to the questions that arise in social psychological research, we can leave behind some of the limitations that are inherent in the use of statistical models that were created to study nonsocial phenomena.

The round robin data layout is not new; this type of design has been employed in a number of classic social psychological studies (Campbell, Miller, Lubetsky, & O'Connell, 1964; Cronbach, 1955; Newcomb, 1961). It continues to be used in small group and person perception research (e.g., Bales, 1970). What is new is the statistical model specifically tailored to this design, which makes it possible to test the significance of individual differences, interaction effects, and various types of reciprocity. This new round robin model is ideally suited to the kind of research Tagiuri called for in his 1969 article on person perception in the *Handbook of Social Psychology*. He recommended more study of person perception in the context of ordinary transactions that occur in the natural environment, where the persons are interacting with each other and each person is simultaneously judge and object.

The round robin design deserves serious consideration as an alternate research strategy that provides information complementary to that from traditional experimental designs. We hope that the availability of techniques that facilitate analysis of data from naturally occurring social interactions will stimulate interest in the interactive, mutually contingent aspects of social behavior and the relationship between individual differences and social interaction.

Reference Notes

1. Stoto, M., Kenny, D. & Warner, R. *Derivation of the expected mean squares for the round robin analysis of variance*. Unpublished manuscript, Harvard University, 1978.
2. Warner, R. *Temporal patterns in dialogue*. Unpublished doctoral dissertation, Harvard University, 1978.

References

Bales, R. F. *Personality and interpersonal behavior*. New York: Holt, Rinehart & Winston, 1970.

Bechtel, G. G. The analysis of variance and pairwise scaling. *Psychometrika*, 1967, *32*, 47–65.

Bechtel, G. G. The analysis of variance for paired comparisons. *Psychometrika*, 1971, *36*, 135–154.

Campbell, D. T., Miller, N., Lubetsky, J., & O'Connell, E. J. Varieties of projection in trait attribution. *Psychological Monographs*, 1964, *78* (15, Whole No. 592).

Cronbach, L. J. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychological Bulletin*, 1955, *52*, 177–193.

Gleason, J. R., & Halperin, S. A paired comparisons model for round robin experiments. *Psychometrika*, 1975, *40*, 43–54.

Goldman-Eisler, F. Individual differences between interviewers and their effect on interviewee's speech behavior. *Journal of Mental Science*, 1952, *98*, 660–671.

Huynh, H., & Feldt, L. S. Conditions under which mean square ratios in repeated measurements designs have exact *F*-distributions. *Journal of the American Statistical Association*, 1970, *65*, 1582–1589.

Jaffe, J., & Feldstein, S. *Rhythms of dialogue*. New York: Academic Press, 1970.

Lev, J., & Kinder, E. New analysis of variance formulas for treating data from mutually paired subjects. *Psychometrika*, 1957, *22*, 1–15.

Mosteller, F., & Tukey, J. *Data analysis and regression: a second course in statistics*. Reading, Mass.; Addison-Wesley, 1977.

Newcomb, T. M. *The acquaintance process*. New York: Holt, Rinehart & Winston, 1961.

Snedecor, G. E., & Cochran, W. G. *Statistical methods*. Ames: Iowa University Press, 1967.

Tagiuri, R. Person perception. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (Vol. 3, 2nd ed.). Reading, Mass.: Addison-Wesley, 1969.

Welkowitz, J., Cariffe, G., & Feldstein, S. Conversational congruence as a criterion of socialization in children. *Child Development*, 1976, *47*, 269–272.

White, H., & Breiger, R. Pattern across networks. *Society*, 1975, *12*(5), 68–73.

Received November 2, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 10, 1758-1768

The Independence of Evaluative and Item Information: Impression and Recall Order Effects in Behavior-Based Impression Formation

Elizabeth K. Dreben, Susan T. Fiske, and Reid Hastie
Harvard University

Impression and recall order effects were examined in the context of Anderson's information integration model of impression formation. Subjects read four items of information about each stimulus person and participated in one of five between-subjects conditions: (a) rated the stimulus person after the presentation of each item and, after the fourth rating, recalled the four items; (b) did the same, except that a trait-rating filler task was interpolated between each rating and the next presentation; (c) did the same, except that the filler task was math problems; (d) made only one rating, after the presentation of the fourth item, did no filler task, and did recall the four items; (e) did the same, without recall. In the first three conditions, each impression rating was affected most by the most recently presented item, whereas in the two final response conditions, there was a slight primacy effect. In contrast, task conditions had little effect on recall; serial position curves for all four recall conditions were U-shaped, exhibiting both primacy and recency, as commonly found in free recall research. Low and nonsignificant correlations between impression weights and recall, as well as the markedly different serial position curves, were discussed as evidence for distinct processes of recall and impression formation.

Recent theoretical analyses of human cognition have depended heavily on the concept of abstraction processes to explain performance in perception, comprehension, memory, and judgment tasks. For example, psychological models for visual perception (Minsky, 1975; Posner, 1969), sentence comprehension (Bransford & Franks, 1972; Schank & Abelson, 1977), verbal memory (Bartlett, 1932; Mandler & Johnson, 1977), visual memory (Mandler & Parker, 1976), and decision making (Kahneman & Tversky, 1972) have emphasized the abstraction of prototypes, data structures, semantic products, deep structures, schemata, and themes as the central component of cognitive information processing. An important empirical question concerns the relationship between mental representations of the concrete surface structure of the stimulus and the schematic structure that is abstracted from the original stimulus (e.g., Sachs, 1967).

The present research is concerned with the relationship between memory for concrete factual information about a person's behavior and an abstract evaluative impression based on that information. In particular, we will be concerned with the relationship between the importance or impact of an item of information on the abstract impression and the likelihood that that specific item is recalled in a subsequent memory test. Before we discuss our method for studying the impression–memory relationship, we will review relevant work on impression and memory processes considered in isolation from one another.

This article is based on the first author's undergraduate honors thesis. The authors acknowledge the helpful comments of Norman Anderson, David Green, and David Kenny.

Elizabeth K. Dreben is now at Boston University and Susan T. Fiske is now at Carnegie-Mellon University.

Requests for reprints should be sent to Reid Hastie, Department of Psychology and Social Relations, Harvard University, Cambridge, Massachusetts 02138.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3710-1758$00.75

Impression Formation Processes

In 1946 Asch introduced an impression formation task in which experimental subjects are shown a sequence of trait adjectives attributed to a single fictional person and are asked to judge the person's personality. This task has dominated research on social judgment during the past 30 years. Anderson's information integration theory (e.g., 1974) averaging model provides the most popular and most thorough account for the impression abstraction process in the Asch task. Anderson's model prescribes that information from stimulus elements or cues will be integrated according to an averaging rule to produce a response. Quantitatively:

$$R = w_1 s_1 + w_2 s_2 + \cdots = \sum_{i=1}^{n} w_i s_i$$

where w_i = weight and s_i = scale value.[1] The subscripts refer to the serial positions of the piece of information in the sequence (i.e., $w_1 s_1$ denotes the weight and scale value of the first trait in an Asch task). The restriction that the weights must sum to one makes the model an averaging model.

One major finding obtained in the Asch paradigm concerns *order* effects. In general, early information seems to have a greater effect on the response than does subsequent information, creating a primacy effect. However, given certain task conditions, such as continuous responding, final recall, and others, recency effects occur.

According to Anderson's averaging model, primacy occurs when the relative weights of the first items are larger than the relative weights of the last items. Recency occurs in the reverse situation, when the last items are weighted more heavily than the first items. The averaging model is used to estimate the effects of each serial position on the final rating. (These serial position effects are proportional to the weights in the averaging model.)

Anderson suggests both an inconsistency discounting explanation and an attention decrement explanation to account for the changing weights that cause the primacy effect. According to the discounting explanation, later adjectives that are inconsistent with earlier adjectives are discounted and weighted less. The attention decrement hypothesis posits that the earlier information counts more because the subjects attend less to each subsequent piece of information.

Several experiments have been designed to test between these explanations by varying certain requirements of the standard procedure. Stewart (1965) found the primacy effect when the subjects made only one rating after all the adjectives were presented, but he found a recency effect when subjects made a rating after each adjective. Anderson (1968) obtained recency when subjects pronounced aloud each adjective after it was presented. This recency occurred when final, continuous, and intermittent (rating once in the middle as well as at the end) response conditions were used. Hendrick and Constantini (1970) varied the degree of inconsistency between the traits and found that intertrait inconsistency did not change the primacy effect. They observed a recency effect when pronunciation of each adjective was required, and primacy without the pronunciation requirement.

The results we have just reviewed are all supportive of the attentional account for serial position effects in impression formation. Only one result in the impression formation literature appears to contradict such an account. Hendrick, Constantini, McGarry, and McBride (1973) found that variations in the time interval between trait presentations (1 to 5.8 sec) did not affect primacy. However, even this result is not unequivocally contrary to attentional accounts. Thus, we have emphasized attentional mechanisms in our theoretical analysis of the present experimental task.

To look at order effects in greater detail, Anderson and Farkas (1973) used an experimental design that permitted estimation of serial position weights. They used a continuous response task and found recency in the

[1] For the purposes of this discussion the scale value and weight of the initial impression (designated $w_0 s_0$) are ignored.

last serial position (for each rating), with the early portions of the curve being flat. Their explanation, seen as a specification of the attention decrement hypothesis, postulates a *surface* component that refers to the short-term recency effect, reflecting the immediate reaction to the stimulus just presented and a *basal* component that represents the stable component of the impression reflected in the flat portion of the curves. Fluctuations in the content of the surface component correspond to variations in attention.

One experiment in this tradition was concerned with the impression–memory relationship that is central in the present investigation. In 1963 Anderson and Hubert conducted an impression formation experiment that replicated the standard primacy effect obtained in earlier research. When subjects were told they would be asked to recall the adjectives, however, they found reduced primacy and, in one condition, a slight recency effect. Anderson and Hubert also found a strong recency effect for recall in all conditions, even when there was a primacy effect on the impression. They cited the impression results as evidence for their attention decrement hypothesis. The reduced primacy was interpreted as evidence that subjects anticipating recall equalized attention to all items as they were presented. More importantly, Anderson and Hubert concluded that the difference between impression and recall serial position effects implied "that the formation of the impression involves a memory process which is distinct from, and not dependent on, the immediate verbal memory for the adjectives just heard" (Anderson & Hubert, 1963, p. 386).

Free-Recall Memory Processes

The free-recall paradigm in memory research is analogous to the Asch impression formation task. In free recall the serial position effect is usually in the form of a bow-shaped curve (Klatzky, 1975). There is a primacy effect, followed by an asymptote in the middle section and then a recency effect. Variables such as rate of presentation, word frequency of occurrence in English, and list length variations affect primacy and asymptote portions of the curve (Glanzer, 1973). Variations in delay between study and recall affect the recency portion. These results support the two-component memory hypothesis, that is, that there is a limited short-term store and a very large long-term store.

It is important to note that the simple two-store model is not universally popular. For example, one result, apparently discrepant from the two-component theory, is that of Bjork and Whitten (1974). They used distractor-delay tasks to interfere with the short-term component and thereby eliminate recency effects. These variations did not affect recency, however, and so Bjork and Whitten concluded that long-term storage must underlie the recency effect.

Overview

The present experiment used the Anderson and Farkas (1973) design with continuous responding and includes the Anderson and Hubert (1963) free recall measure. In addition, the Bjork and Whitten (1974) interpolated distractor task, or delay between the serial positions, was also employed. The logic of our examination of the impression–recall relationship is an extension of the Anderson and Hubert (1963) reasoning. The focus is on the correlation between the forms of impression weight and recall probability serial position curves. However, we carry the analysis forward in two respects: first, by examining complete position-by-position serial curves, and second, by varying interpolated response tasks that previous research has shown will affect impression and memory processes. Two interpolated response tasks were chosen, math problems and trait ratings, because we predicted that they would have differential effects. Since traits are relatively similar to behavior sentences, we hypothesized that the trait task would interfere more on both the impression and recall tasks than would the math task.

Two recent papers (Riskey, 1979; Rywick & Schaye, 1974) report related extensions of the Anderson and Hubert approach. For the present, we will only note that these investigations reached contradictory conclusions on

the issue of the impression–memory relationship. Riskey concurred with Anderson and Hubert that abstract impression and item specific representations are independent while Rywick and Schaye concluded that the impression was dependent on specific item information in long-term memory. We will discuss these results more thoroughly after we have described our own empirical investigation.

Method

The subjects completed ratings of eight sequences, each of which was composed of four sentence pairs describing a person. We used sentences describing behaviors, instead of traits or paragraphs, because people often use sentences involving behaviors in person descriptions (e.g., Fiske & Cox, 1979). There were five experimental conditions: three continuous response groups who made a cumulative likability rating after each sentence pair and were asked to recall as much as possible after the fourth or final rating for each stimulus person, and two final response groups who made only one rating after the fourth sentence pair for each stimulus person.

Subjects

Eighty Harvard-Radcliffe undergraduates recruited as paid subjects were tested in groups of 1 to 5 people, with each subject in a separate cubicle. Subjects tested at the same time were always in the same experimental condition. The continuous response groups were tested as one phase of the experiment, and the final response groups were tested 5 months later. The five experimental conditions can be compared, as experimental procedures were uniform for all groups.

Procedure

There were five cover sheets, three for continuous and two for final responding. In the first continuous responding group (no-task condition), the subjects were instructed to read two sentences about a person and then to make a rating below the sentences on an unmarked 200-mm scale with endpoints labeled *most likable* and *least likable*. They were to turn the page when told and then were to read another two sentences. They then made a rating based on all the previous information. They were told that after four ratings they would be asked to recall as much as possible about the stimulus person.

The two other continuous responding groups included the same tasks described above, but added a task in between each page for sentence pairs. The alternating task was either a page of multiplication problems (math condition) or a page of trait ratings (trait condition). Instructions were to read and rate a person and, upon signal, to turn the page and work on a separate task. The alternating tasks would be repeated four times (after each sentence pair) and would be followed by the recall task.

In the fourth condition (final-only condition), subjects were instructed to read two sentences about a person, and on signal they were to turn the page and read another two sentences. After four sentence pairs a final likability rating was made. In the fifth condition (final with recall condition), instructions were the same as those in the fourth condition, except that the likability rating was followed by the recall task.

During the experiment, subjects were given 10 sec to read and rate each sentence pair, 30 sec for the second task (in the alternating conditions), and 90 sec for the free-recall measure.

Stimuli

The stimuli were pairs of sentences, with each pair composed of either a positive sentence and a neutral sentence or a negative sentence and a neutral sentence. An example of a positive–neutral pair is "Alan bought groceries for an elderly lady next door who was ill" and "Alan pressed the button and waited for the elevator to come." For generalizability of stimuli, and so that each sentence appeared only once for each subject, there were 64 sentences (8 stimulus people each composed of 8 sentences): 16 positive, 16 negative, and 32 neutral. To insure that the sentences within each category (i.e., positive, negative, or neutral) would be considered roughly equivalent in terms of likability, the sentences were pilot tested before the experiment was conducted. The mean and spread of ratings for a larger group of sentences was examined, and 64 sentences were chosen (with the lowest positive mean at 139.40, the highest negative mean at 49.75, and the neutral means ranging between 93.05 and 105.15 on a 200 mm scale).

Design

The design is a $5 \times (4 \times 2)$ factorial design, with one between-subjects variable and two within-subjects variables. The five-level variable is the subject's task during the experiment session that was introduced above. The four-level repeated measure is serial position (1 through 4) and the two-level repeated measure is (high or low) valence.

In order to reduce the number of observations per subject, the research used a half-replication of the 2^4 design [2] (e.g., Cochran & Cox, 1957) with 8 sequences

[2] The half-replicationof the 2^4 design was analyzed as a 4×2 within-subjects design. The four-level factor corresponds to input serial position and the two-level factor corresponds to item valence at each position (evaluating positive or negative).

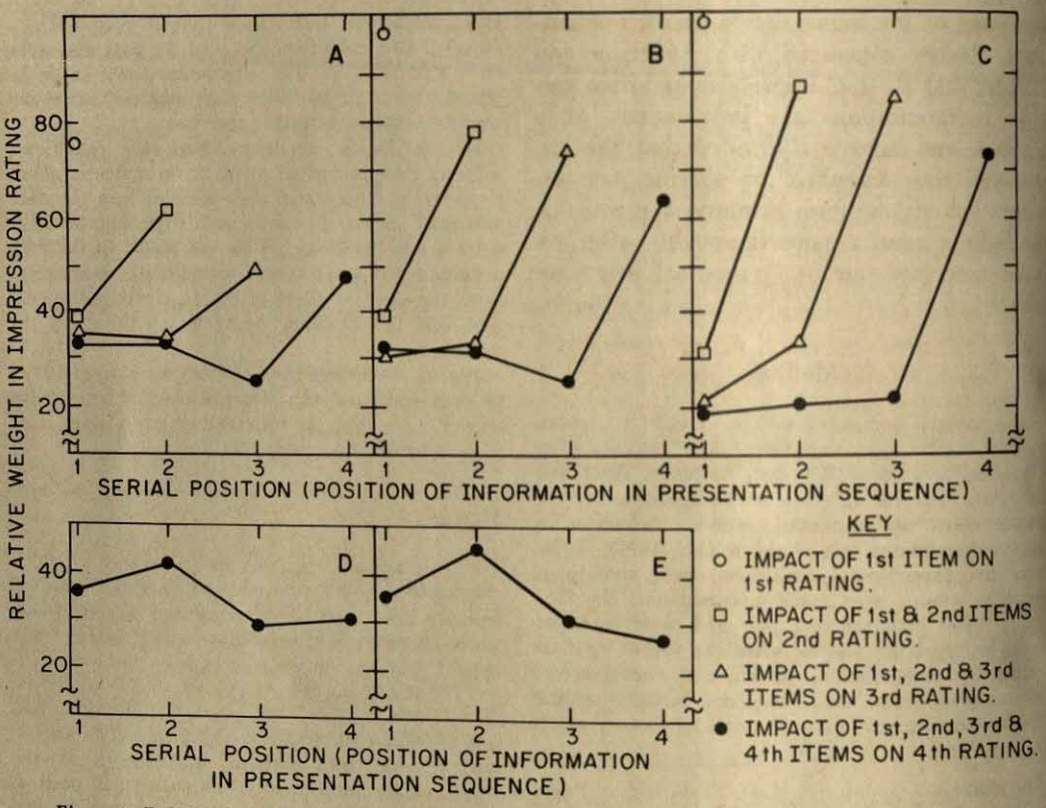


Figure 1. Relative impact of information presented at first through fourth positions on impression ratings made after each position. (Note that each subject in Panels A–C made four ratings, and the impact of the most recently presented item was always the greatest. Conditions in which subjects made only one rating, Panels D and E, show no recency and some primacy. The five panels represent the between-subjects task conditions: [a] four impression ratings, one after information presented at each serial position, and a recall task after fourth rating [no-task condition]; [b] four ratings, final recall, and irrelevant trait-rating filler after each rating [trait condition]; [c] four interpolated ratings, final recall, and irrelevant math filler after each rating [math condition]; [d] final rating, final recall, and no filler [final with recall]; [e] final rating only, no recall, and no filler [final only condition].)

(each containing 4 sentence pairs) instead of the full 16. Each half-replication was counterbalanced in an 8×8 Greco-Latin square so that both stimulus people and sentence pairs occur equally often in all positions. This ensured that order of presentation was not confounded with the order effects under study. A complete replication of this design requires 8 subjects in each condition. Two replications of the design were run with 16 subjects for each between-subjects condition, making 80 subjects in all.

Results

Impression Order Effects

The main analysis was done on the final impression rating, Rating 4. There was one between-subjects variable, task condition (math/trait/no task/final with recall/final only) and two within-subjects variables, position (four levels) and valence (high/low, or H/L).

Overall, the continuous responding conditions showed recency, and the final responding conditions did not. The serial position curves (serial position weights) calculated for each of the four ratings for no task, trait, and the math conditions are displayed in Figure 1 and clearly show a recency effect. In contrast, subjects in the final response conditions showed a primacy effect. The differential effects of serial position are illustrated by the main effect for position, $F(3, 225) = 10.88$, $p < .001$. The serial position weights in Table

1 show that Position 4 has the greatest impact on the final rating for the continuous response conditions but not for the final response conditions.

The most sensitive tests of the effects of the experimental tasks are the planned contrasts differentiating among the shapes of the five task condition serial position curves. These contrasts partition the variance from the significant Task × Serial Position interaction effect, $F(12, 225) = 4.36$, $p < .001$. The task main effect was nonsignificant ($p > .5$). The striking effect in this analysis was the comparison between the final and continuous responding conditions on the linear effect, $F(1, 225) = 21.29$, $p < .001$.

As in previous research using similar tasks and the serial position counterbalancing plan (Anderson & Farkas, 1973), several higher order interactions, concentrated in the Serial Position × Task × Replication effects, were significant.

Recall Order Effects

The dependent measure for the recall task was the number of valenced (H or L) sentences recalled from each of the four serial positions. The amount of recall at each position was analyzed, using analysis of variance for the four-level within-subjects variable of serial position, the two-level within-subjects variable valence (H or L), and the between-subjects variable of task (trait/math/no task/final with recall). The recall task showed the standard U-shaped curve (both primacy and recency) for all the conditions, although there were differences in the sharpness of the curves depending on the condition. Figure 2 shows the serial position effects in which the no-task condition exhibits the largest primacy and recency effects. The math and final with recall conditions show intermediate curvature across serial positions, whereas the trait condition exhibits the flattest curve. These results are reflected by the significant Task × Serial Position interaction, $F(9, 180) = 2.10$, $p < .05$, which shows the differential position effects as a consequence of experimental condition.

There was a main effect for serial position, $F(3, 180) = 19.23$, $p < .001$, which indicates that the location of a valenced sentence influenced how likely the sentence was to be recalled. The task main effect was nonsignificant. There was also a main effect for valence (H or L), $F(1, 160) = 30.84$, $p < .001$, such that negative information was recalled more frequently than was positive information across all serial positions.

The number of errors in recall (either intrusions of previous stimuli presented or recall made faulty by the inclusion of stimuli) were analyzed, and there were no patterns for either serial position or task effects. (The error rates were very low; for intrusions, $M = .15$, and for faulty recall, $M = .07$.)

A separate analysis of the recall of neutral sentences revealed only a significant main effect for serial position, $F(3, 180) = 7.33$,

Table 1
Relative Influence of Each Serial Position on Final Impression Rating

Serial position	Task condition: Four ratings, math filler, with recall	Four ratings, trait filler, with recall	Four ratings, no filler, with recall	Final-only, no filler with recall	Final-only, no filler, without recall	Row *M*
First	17.64	32.31	33.34	34.99	35.02	30.66
Second	22.89	36.06	35.56	41.78	45.92	36.44
Third	25.11	24.34	25.78	29.70	30.83	27.15
Fourth	71.30	65.50	46.31	30.82	27.23	48.23
Column *M*	34.23	39.56	35.25	34.32	34.75	35.62

Note. Table entries are proportional to information integration weights (for derivation, see Anderson & Farkas, 1973).

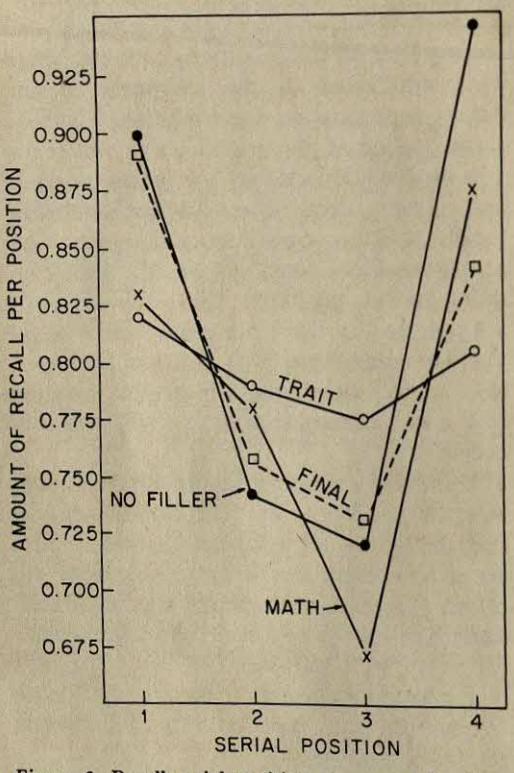


Figure 2. Recall serial position curves for the four between-subjects conditions in which recall was tested: no task (labeled *no filler*), math interpolated response task, trait rating interpolated response task, and final rating with recall task.

$p < .001$. Neutral sentence recall exhibited the typical U-shaped serial position curves but no other systematic patterns.

Relationships Between Impression and Recall

Overall, recall curves differ markedly from the impression serial position curves. In particular, there is a clear primacy effect for recall in all task conditions—although there is no such effect for impressions in continuous responding task conditions. It seems that earlier information is not forgotten, but it is given less weight in impressions.

At a finer level of analysis, correlations were calculated to measure the degree to which total valenced recall at each position corresponded to differential serial position weights. For the correlations, the impression difference scores (positive minus negative information *for each position*) were calculated on the final (fourth) likability rating.[3] Examination of the mean correlations pooled across the four serial positions for each of the four relevant between-subjects conditions (trait/math/no task/and final with recall) shows that none of these four pooled mean correlations is significantly different from zero. But the no-task and math conditions ($r = .28$ and $r = .38$ respectively) show a higher correlation between recall and differential impression rating than do the trait and final with recall conditions ($r = -.15$ and $r = .14$). Thus there is a low but nonsignificant degree of correlation between recall and differential likability ratings.

Discussion

The focus of the present experiment is on the relationship between an abstract personality impression and the specific items of information on which that impression is based. Anderson and Hubert (1963) and Riskey (1979) have conducted research that uses a method similar to the present experimental paradigm and have concluded that the impression, once formed, is independent from the representation of specific item information in memory. Our conclusion is in agreement with these investigators but is based on research with rich behavioral stimulus materials, involves an extensive variation in subjects' processing tasks, and includes a more detailed analysis of impression and recall results than

---

[3] For the Rating 4 analysis only, because half-replications were used in this design, the higher order interaction effects are confounded with replication. The difference score at Position 4 is equal to the four-way interaction of Replication × Difference Score at Position 1 × Difference Score at Position 2 × Difference Score at Position 3. The pattern of the first three positions in any sequence is identical for the two replications (i.e., HHH). It is the fourth sentence pair (H or L) that differs for the two replications.

The valence of the information at Position 4 is determined by the levels of Positions 1–3 and replication (1 or 2). Thus, the alias for Position 4 is the interaction of Position 1 × Position 2 × Position 3 × Replication (Cochran & Cox, 1957, p. 276). Since the other three ratings were not affected by the information at Position 4, that factor could be dropped from the design.

earlier experimental work. In this discussion we will briefly summarize our analyses of impression and recall results and then end with a theoretical note on the significance of these results in the context of the conclusions of Anderson and Hubert and of Riskey.

Impression Order Effects

First, the present results using sentences as stimuli reproduce the impression order effects obtained in previous research using traits, paragraphs, or photographs (Anderson & Farkas, 1973; Anderson & Lampel, 1965; Asch, 1946). This suggests that information embodied in the sentence descriptions of behaviors is used in the same way that information contained in the other stimulus materials is used. Thus, order effects do not seem sensitive to the stimuli used (traits, sentences, or paragraphs). This insensitivity is in marked contrast to the dependency of order effects on the method of presentation of the stimuli. As mentioned earlier, continuous responding (these data; Anderson & Farkas, 1973; Stewart, 1965), pronouncing aloud (Anderson, 1968), and warned recall (Anderson & Hubert, 1963) all serve to change the primacy effect into a recency effect. Thus, although the order effects are fragile in the sense that they can be changed when task requirements change, these same order effects are quite stable across variations in the stimuli employed.

Second, the final versus continuous responding variation in impression rating task requirements clearly affected the shape of the serial position curve, as shown by recency effects with continuous responding and no recency with final responding (Panels A, B, and C vs. Panels D and E in Figure 1). The distractor conditions (math and trait tasks) showed more recency than did the no-task condition (Panels B and C vs. Panel A in Figure 1). The impression-irrelevant math problems and trait ratings may interfere with the process of impression formation, with the interruptions leading subjects to weight immediate information the most in their ratings.

Third, the current no-task condition (Panel A in Figure 1) replicated the Anderson and Farkas (1973) recency effect, and our addition of a warned recall task did not change the impression serial position curves. This differs from the Anderson and Hubert (1963) finding that warned recall decreased the primacy effect and sometimes produced recency.

Finally, the results for all conditions suggest that negative information has more weight than positive, which accords with the negativity effects in the impression formation literature (Kanouse & Hanson, 1972).

Recall Order Effects

The recall curves for the no-task, math, and final with recall conditions (see Figure 2) are the usual U-shaped curves obtained in the free-recall task. The trait condition curve is flattest but also shows slight primacy and recency effects. Although the time elapsed between the start of the no-task and final with recall conditions and the recall task was 2 minutes shorter than in the distractor conditions, the math recall curve is quite similar to the curves from the nondistraction conditions. The reduced recency in the math and trait conditions could be due to either the four distractor tasks themselves or to the final distractor task, which delayed the recall task.

The trait recall curve, considered in isolation, is interesting because the Bjork & Whitten (1974) finding of recency in the presence of a distractor task was not replicated. In one experiment, word pairs were presented, with simple multiplication problems between each pair and further multiplication before the recall task. They obtained a U-shaped curve and suggested that this was evidence for long-term store effects on recency. In another experiment, they found similar recency effects, whether or not there was a 30-sec final delay before recall, as long as the interpolated task between each word pair was presented. This is further evidence, according to Bjork & Whitten, that recency results from a long-term memory process.

Bjork and Whitten suggested that the discriminability between items to be remembered caused the recency effect and is produced by using distractor tasks. These interruptions temporally separate the stimuli to make each

piece of information distinct. The current results cast doubt on this explanation. Although the math condition, which is similar to their interpolated task, replicated Bjork and Whitten's recency effect, the trait condition did not. It seems likely that the nature of the distractor task is important and that these tasks do not simply temporally distinguish between the stimuli. The distractor task may make the items more distinct by means of a figure–ground relationship. Since the math is clearly different from the verbal stimuli, it may separate each item of information more than do traits, which are more similar to sentences.

Relationships Between Impression and Recall

Under continuous rating conditions, subjects clearly weight the earliest information the least in their final impression rating, but this is not because they no longer remember it. All of the relevant recall serial position curves show clear primacy and recency effects. A more complete summary of the relationship between mean recall and mean impression weight would be obtained by plotting the mean recall at each position (points plotted in Figure 2) against the impression serial position weights on Rating 4 (solid points in Figure 1). Such a summary reveals that there is no simple relationship between recall and impressions, at least as indicated by their respective mean serial position curves. The nonsignificant correlations between recall and differential impression ratings based on the data from each individual character were reported above as further evidence that no simple relationship between memory and the impression exists.

The final responding conditions provide a replication of the Anderson and Hubert (1963) and the Riskey (1979) findings. Impression serial position curves show slight primacy and no recency, whereas recall serial position curves show dramatic recency and some primacy. Our interpretation of these results is in complete agreement with the Anderson and Hubert and Riskey accounts. Attention decreases across the sequence of stimulus items, producing the primacy effect in the impression serial position curve. Memory for individual stimulus items is not relevant to the impression, once their evaluative content has been integrated into the current impression. Thus, the U-shaped memory serial position curve is unrelated to the monotonically decreasing impression serial position curve.

A somewhat different picture emerges from an examination of results from continuous responding conditions. Here a large recency effect is consistently obtained for the *most recent item* on each of the four likability ratings. Anderson and Farkas (1973) have proposed a two-component model for a similar task, distinguishing between *surface* and *basal* components of an impression. Our interpretation is that each impression rating is dependent on information immediately available in short-term memory (surface component) as well as on the enduring impression that has accumulated across all previous stimulus events (basal impression). Thus, under continuous responding conditions the impression rating *is* dependent on immediately available item information as well as on the more abstract, item-independent, cumulative impression. The increase in recency effect magnitude with increasingly demanding tasks interpolated between ratings (no task *to* trait task *to* math task) adds support to this interpretation. (Note that this ordering of tasks by difficulty or cognitive demand is a post hoc interpretation.)

The present results from both final responding and continuous responding task conditions are quite consistent with an account based on attentional mechanisms (Anderson & Hubert, 1963; Riskey, 1979) and on a surface–basal component distinction (Anderson & Farkas, 1973). The emphasis on attentional mechanisms is strengthened even further by a detailed comparison of response requirements in continuous and final responding rating tasks. Under continuous responding conditions, the impression rating scale was printed on each page *containing* behavior descriptions, located below those descriptions. These are the conditions under which large recency effects were obtained. Thus, the information that has a large impact on the rating is not only recently avail-

able to comprehension and short-term memory mechanisms, it is directly available in the visual (attentional) field. On the other hand, under final responding conditions, the final (single) rating scale was printed on the page *following* the last page of behavior description sentences. Thus, under final responding conditions where no recency and slight primacy effects were obtained, immediate attention was not dominated by the final most recent information.

We would like to close this report by noting that our conclusion is an argument for the relative independence of a general abstract impression and the specific information on which the abstraction was based. Furthermore, we have cited a simple attentional mechanism as an explanatory principle to account for the differential impact of recent events on the evaluative impression under continuous and final responding conditions. This explanation presents a picture of the subject as a person whose behavior is stimulus driven and whose thoughts are dominated by the immediate environment (Kahneman & Tversky, 1972; Taylor & Fiske, 1978). Although we believe our subjects' judgments were dominated by the given stimulus information, we also attribute considerable cognitive complexity to comprehension, valuation, and integration operations that are assumed to underlie the impression formation process.

References

Anderson, N. H. Application of linear-serial model to a personality-impression task using serial presentation. *Journal of Personality and Social Psychology,* 1968, *10,* 354–362.

Anderson, N. H. Information integration theory: A brief survey. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology.* San Francisco: Freeman, 1974.

Anderson, N. H., & Farkas, A. J. New light on order effects in attitude change. *Journal of Personality and Social Psychology,* 1973, *28,* 88–93.

Anderson, N. H., & Hubert, S. Effects of concomitant verbal recall on order effects in personality impression formation. *Journal of Verbal Learning and Verbal Behavior,* 1963, *2,* 379–391.

Anderson, N. H., & Lampel, A. K. The effect of context on the ratings of personality traits. *Psychonomic Science,* 1965, *3,* 433–434.

Asch, S. E. Forming impressions of personality. *Journal of Abnormal and Social Psychology,* 1946, *41,* 258–290.

Bartlett, F. C. *Remembering.* Cambridge: Cambridge University Press, 1932.

Bjork, R. A., & Whitten, W. B. Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology,* 1974, *6,* 173–189.

Bransford, J. D., & Franks, J. J. The abstraction of linguistic ideas: A review. *Cognition,* 1972, *1,* 211–240.

Cochran, W. B., & Cox, G. M. *Experimental Designs* (2nd ed.). New York: Wiley, 1957.

Fiske, S. T., & Cox, M. G. Person concepts. The effect of target familiarity and descriptive purpose on the process of describing others. *Journal of Personality,* 1979, *47,* 136–161.

Glanzer, M. Storage mechanisms in recall. In G. H. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 5). New York: Academic Press, 1973.

Hendrick, C., & Constantini, A. F. Effects of varying trait inconsistency and response requirements on the primacy effect in impression formation. *Journal of Personality and Social Psychology,* 1970, *15,* 158–164.

Hendrick, C., Constantini, A. F., McGarry, J., & McBride, K. Attention decrement, temporal variation, and the primacy effect in impression formation. *Memory and Cognition,* 1973, *1,* 193–195.

Kahneman, D., & Tversky, A. Subjective probability: A judgment of representativeness. *Cognitive Psychology,* 1972, *3,* 430–454.

Kanouse, D. E., & Hanson, L. R. Negativity in evaluations. In E. E. Jones et al. (Eds.), *Attribution: Perceiving the causes of behavior.* Morristown, N.J.: General Learning Press, 1972.

Klatzky, R. L. *Human memory: Structures and processes.* San Francisco: Freeman, 1975.

Mandler, J. M., & Johnson, N. S. Remembrance of things parsed: Story structure and recall. *Cognitive Psychology,* 1977, *9,* 111–151.

Mandler, J. M., & Parker, R. E. Memory for descriptive and spatial information in complex pictures. *Journal of Experimental Psychology: Human Learning and Memory,* 1976, *2,* 38–48.

Minsky, M. A framework for representing knowledge. In P. Winston (Ed.), *The psychology of computer vision.* New York: McGraw-Hill, 1975.

Posner, M. I. Abstraction and the process of recognition. In G. H. Bower and J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 3). New York: Academic Press, 1969.

Riskey, D. R. Verbal memory processes in impression formation. *Journal of Experimental Psychology; Human Learning and Memory,* 1979, *5,* 271–281.

Rywick, T., & Schaye, P. Use of long-term memory in impression formation. *Psychological Reports,* 1974, *34,* 939–945.

Sachs, J. D. S. Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics*, 1967, *2*, 437–442.

Schank, R. C., & Abelson, R. P. *Scripts, plans, goals, and understanding*. Hillsdale, N.J.: Erlbaum, 1977.

Stewart, R. H. Effect of continuous responding on the order effect in personality impression. *Journal of Personality and Social Psychology*, 1965, *1*, 161–165.

Taylor, S., & Fiske, S. T. Salience, attention, and attribution: Top of the head phenomena. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 11). New York: Academic Press, 1978.

Received November 6, 1978 ■

Instructions for Submission of Manuscripts

Papers submitted to the *Journal of Personality and Social Psychology* after April 1, 1979 will be considered under new policies approved by the Publications and Communications Board. Effective with the 1980 volume of the journal, the following policies will apply:

The *Journal of Personality and Social Psychology* publishes original papers in all areas of personality and social psychology. It emphasizes empirical reports but may include specialized theoretical, methodological, and review papers. Beginning in April 1980, the journal will be divided into 3 independently edited sections: ATTITUDES AND SOCIAL COGNITION, INTERPERSONAL RELATIONS AND GROUP PROCESSES, and PERSONALITY PROCESSES AND INDIVIDUAL DIFFERENCES. Authors should submit manuscripts to the appropriate section editor according to the section definitions and manuscript preparation instructions provided on page 468 of the March 1979 issue of the journal.

Section editors' addresses are as follows: ATTITUDES AND SOCIAL COGNITION, Melvin Manis, Editor, Research Center for Group Dynamics, Institute for Social Research, University of Michigan, Ann Arbor, Michigan 48106; INTERPERSONAL RELATIONS AND GROUP PROCESSES, Ivan D. Steiner, Editor, Department of Psychology, University of Massachusetts, Amherst, Massachusetts 01003; PERSONALITY PROCESSES AND INDIVIDUAL DIFFERENCES, Robert Hogan, Editor, Johns Hopkins University, Department of Psychology, 33rd and Charles Streets, Baltimore, Maryland 21218.

Journal of Personality and Social Psychology
1979, Vol. 37, No. 10, 1769-1778

Self-Focus, Felt Responsibility, and Helping Behavior

Shelley Duval, Virginia Hensley Duval, and Robert Neely
University of Southern California

Two experiments were designed to determine if juxtaposing self-focus and salient distressed others would (a) increase self-attribution of responsibility for those needy others and (b) increase willingness to help those others. In Experiment 1, these hypotheses were tested by exposing subjects to their images on a TV screen 4 minutes before, immediately before, immediately after, or 4 minutes after seeing a videotape on victims of the venereal disease epidemic. As predicted, results indicated that subjects who saw their images immediately before or after seeing the videotape on victims of venereal disease felt more responsibility for and were more willing to help that group than did subjects in other conditions. In Experiment 2, subjects filled out a biographical questionnaire either 4 minutes before or immediately before seeing a videotape on poverty-stricken Latin Americans. Results from this experiment also confirmed predictions. Subjects who filled out a biographical questionnaire immediately before seeing a videotape on poverty in Latin America felt more responsibility for and were more willing to help the distressed group than did subjects who completed the questionnaire 4 minutes prior to exposure to the tape on Latin American poverty. Additional evidence indicated that this effect is probably not mediated by the sole operation of the self-evaluative mechanism posited by Duval and Wicklund or by any change in attitudes or norms regarding the distress of and/or necessity of helping distressed others. Implications of the self-focus/felt responsibility/helping behavior relationship for the general study of helping behavior are discussed.

Research clearly demonstrates that helping behavior is related to feelings of responsibility for the welfare of others in distress (Geer & Jarmecky, 1973; Schwartz, 1973; Schwartz & Clausen, 1970; Tilker, 1970). This relationship between feelings of responsibility for the welfare of others and helping behavior can be stated within the theoretical framework of attribution theory. Factors that increase self-attribution of responsibility for a distressed person will increase helping behavior. Factors that decrease self-attribution of responsibility for a distressed person will decrease helping behavior.

In retrospect, research based on an attributional approach to helping behavior (Ickes, Kidd, & Berkowitz, 1976; Piliavin, Rodin, & Piliavin, 1969; Schopler & Matthews, 1965) appears (with one exception, Wegner & Schaefer, 1978) to have been guided by one general question: What factors affect a person's attribution of responsibility, given that the attributional processes are governed by rational rules or schema such as discounting? Recent developments in attribution theory (Duval & Duval, in press; Duval & Hensley, 1976; Jones, 1976) indicate that attribution does not always conform to the logical patterns dictated by reason or common sense. This suggests that an attributional approach to the study of helping behavior might also attempt to determine what conditions affect a person's attribution of responsibility, given that the attribution processes are not strictly governed by rational rules. The present research is addressed to this question. Specifically, two studies were designed to determine whether conditions that affect the focus of attention also affect self-attribution of re-

Requests for reprints should be sent to Shelley Duval, Department of Psychology, University of Southern California, Los Angeles, California 90007.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3710-1769$00.75

sponsibility for the welfare of distressed persons and thus helping behavior.

Heider (1958) suggests that attribution of responsibility reflects the processes of unit formation. To the extent that conditions produce a unit relationship between some agent and some effect, that is, some event or situation, responsibility for the effect will be attributed to the agent. Duval and Duval (in press) and Duval and Hensley (1976) propose that the focus of attention is one variable that affects this process. To the extent that attention focuses on an agent, B, rather than on A, C, or others, the tendency to link B and an effect, X, in a unit relationship and thus to attribute responsibility for X to B should increase. Data from several studies support this hypothesis (Arkin & Duval, 1975; Duval & Wicklund, 1973; Pryor & Kriss, 1977; Storms, 1973; Taylor & Fiske, 1975).

Applied to a situation in which the distress of other persons is or becomes salient, the focus of attention–attribution hypothesis suggests that increasing self-focus will increase self-attribution of responsibility for the welfare of these distressed persons. Given the relationship between felt responsibility and helping behavior, increasing attribution of responsibility for distressed persons to self should in turn increase willingness to help those distressed persons. In order to test these two hypotheses, subjects were exposed to a stimulus designed to increase self-focus for 1 minute either 4 minutes before, immediately before, immediately after, or 4 minutes after they saw a documentary videotape designed to increase awareness of the plight of victims of the venereal disease (VD) epidemic. Attribution of responsibility for the welfare of VD victims and willingness to help VD victims were then measured. It was predicted that subjects in the immediately before and after conditions would feel more responsible for and would be more willing to help victims of VD than would subjects in the 4-minute before or after conditions.

Before turning to the experimental procedure, we might briefly discuss the basis for the predictions regarding the temporal interval between focusing on self, presentation of the distressed group, and attribution of responsibility to self, as well as the rationale for choosing victims of VD as the potential recipients of help.

The prediction regarding time interval, focus of attention and self-attribution is derived from the unit formation principle of temporal proximity. This principle states that decreasing the time interval that separates the presence or occurrence of an agent and an event increases the tendency to attribute responsibility to the agent and vice versa (Heider, 1958). Data from Michotte (1963) clearly support this prediction. We propose that the principle of temporal proximity applies not only to the time interval that actually separates the presence or occurrence of an agent and an event but also to the time interval that separates periods of increased awareness of the agent and the event. From this perspective, decreasing the time interval that separates the periods of increased focal awareness of self and increased awareness of a situation, such as the distress of VD victims, should increase the tendency to link self and the situation in a unit formation. According to Heider (1958), this increased tendency should lead to greater attribution of responsibility for the welfare of VD victims to self. Thus, subjects who experience increased self-focus immediately before seeing the videotape about VD victims should attribute greater responsibility for the welfare of that group to self and should be more willing to help VD victims than should subjects who experience increased self-focus 4 minutes before seeing the videotape. In addition, data from Michotte (1963) suggest that decreasing the time interval between the periods of increased self-focus and increased awareness of the distress of VD victims will increase self-attribution regardless of the order in which these periods of increased awareness occur.

With regard to the choice of VD victims, we should first note that studies on felt responsibility and helping behavior vary widely in terms of the number of help recipients, the proximity of the subjects to those people, and the extent to which situational cues provide justification for feeling personal responsibility. In the present experiment, we are particularly interested in determining

whether the attribution processes associated with the focus of attention affect felt responsibility and thus helping behavior when the subject is essentially removed from a fairly large group of persons in distress and has few situational cues that justify attribution of responsibility for that group to self. Selection of victims of VD as the target population represents a major part of the attempt to create these conditions.

Experiment 1

Method

Subjects. Subjects were 55 female undergraduates who volunteered for the study in partial fulfillment of a course requirement. Subjects were randomly assigned to four experimental conditions and one control condition and were run individually.

Procedure. Subjects were greeted by a female undergraduate experimenter and were seated in front of a television set. They were told that they were participating in an experiment designed to survey students' attitudes concerning certain social issues. The experimenter further explained that a short documentary videotape had been prepared to familiarize them with some of the facts concerning the particular social issue being investigated. Subjects were told that they would be asked to indicate their attitudes concerning the social issue by filling out a short questionnaire following presentation of the videotape. At this point, the experimenter explained that the study was also concerned with determining whether a simple audio or audio and visual presentation of factual material was more effective. Supposedly, some subjects would only hear the audio portion of the videotape, whereas others would be exposed to both the audio and the visual portions of the videotape. The rationale for interest in the effectiveness of various types of presentations was that previous experiments had failed to produce conclusive evidence on this point. All subjects were told that they were in the audio and visual condition.

After explaining the purpose of the experiment, the experimenter warned subjects not to be surprised if a live picture of themselves appeared on the TV screen at some time during the experiment. She indicated that the experiment would be submitted to the American Science Foundation for a grant and that the foundation wanted all grant proposals to include videotape recordings of all subjects and their exact experimental situations. If there were no questions, the experimenter told subjects she had to warm up the videotape machine for approximately 5 minutes, after which the documentary videotape would be shown and the questionnaire administered. The experimenter then moved behind a partition in the room in order to avoid contact with subjects prior to and after presentation of the videotape. After approximately 5 minutes had elapsed, a 2-minute videotape was shown. The videotape showed an adult male talking about victims of VD. He first presented statistics concerning the epidemic spread of venereal disease to a significant percentage of the population. He then discussed the various harmful physical and mental consequences that victims of venereal disease typically suffer. His presentation did not discuss the causes of VD or responsibility for the welfare of victims of VD, nor did it request any form of help for VD victims.

After viewing the videotape, subjects remained seated for an additional 5 minutes while the experimenter supposedly rewound the videotape and shut down the equipment. After 5 minutes had elapsed, the experimenter reentered the subject's portion of the cubicle and administered the dependent measures.

Time interval and before–after manipulations. In all conditions the experimenter first went behind the partition and activated some equipment that made noises audible to the subject. In the 4 minutes before condition, the experimenter then activated a TV camera positioned so that the subject's face appeared on the TV monitor. The subject was exposed to her image for 1 minute. After turning the camera off, the experimenter waited 4 minutes and then played the videotape. After showing the videotape, the experimenter activated other equipment that made noises audible to the subject, waited behind the partition for 5 minutes, and then administered the dependent measures. In the brief time interval before condition, the experimenter activated the equipment, waited for approximately 4 minutes, and then activated the TV camera focused on the subject. The subject was exposed to her image on the TV monitor for 1 minute. After the 1-minute exposure, the experimenter turned the camera off, delayed for 1 to 2 sec, and then presented the videotape. After exposure to the videotape, the experimenter waited 5 minutes and then administered the dependent measures. In the brief time interval after condition, the experimenter activated the equipment, waited for approximately 5 minutes, and then presented the videotape. The end of the videotape was followed by a 1- to 2-second delay. The experimenter then turned on the live TV camera, exposed the subject to her image on the TV monitor for 1 minute, waited for approximately 4 minutes, and administered the dependent measures. In the 4 minutes-after condition, the experimenter turned on the equipment, waited for 5 minutes, played the videotape, waited for 4 minutes, and then exposed the subject to her image on the TV monitor for 1 minute. The dependent measures were then administered. A control condition using similar instructions was also run. Subjects waited 5 minutes before and after seeing the videotape. However, no TV camera was visible to the subjects, and the subjects were never exposed to a live image of their faces on the TV monitor.

Table 1
Mean Attribution of Responsibility for Victims of VD to Self

Condition	Immediate	4-minute interval
Before	9.0	6.9
After	10.27	7.9
Control	6.5	

Note. The larger the number, the greater the attribution to responsibility to self. For each condition, $n = 11$. VD = venereal disease.

Dependent measures. Five minutes after the videotape ended, subjects were given a questionnaire to fill out. The questionnaire included items concerning the content and effectiveness of the videotaped presentation. It also included questions designed to assess subjects' feelings of responsibility for the welfare of VD victims and willingness to help victims of VD. These questions were the following:

1. To what extent would you say that the failure to control the current VD epidemic is due to the public's, including your own, lack of concern rather than to the VD victim himself?
2. To what extent would you be willing to take a 1-hour course dealing with the prevention and treatment of VD and then conduct two half-hour educational sessions with groups of college freshmen?
3. A VD clinic exists in the college community but is underfinanced and greatly understaffed. How much time per week would you be willing to spend doing volunteer work at the clinic?
4. How large a financial contribution would you be willing to make to the National Center for the Prevention of VD?

All questions were accompanied by 15-point scales anchored at both ends with appropriate positive and negative labels. After subjects had completed the questionnaire, the experimenter checked for suspicion and then revealed the true nature of the experiment and experimental procedures.

Results

An analysis of data obtained from questions concerning the content and effectiveness of the videotaped presentation was carried out to determine if the experimental manipulations affected the subject's memory of the information presented on the tape or her impression of the effectiveness of the presentation. No significant differences between conditions were found for content or for effectiveness ($F < 1$ in all cases). In addition, no differences between the control condition and any of the experimental conditions were found on these measures ($F < 1$ in all cases).

The first experimental hypothesis suggests that a person should tend to see self as responsible for the welfare of people in distress to the extent that a brief time interval separates the periods of increased self-focus and increased awareness of distressed persons. In addition, this effect should occur regardless of the order in which increased self-focus and increased awareness of the distressed others occur. An analysis of the attribution of responsibility to self data (Table 1) is consistent with this prediction. Subjects in the brief time interval conditions felt more personal responsibility for the welfare of VD victims than did subjects in the 4-minute time interval conditions, $F(1, 41) = 5.01$, $p < .05$. There was no effect for before–after or any interaction between before–after and time interval ($F < 1$) in both cases. A comparison between the brief time interval conditions and the control condition evidenced the presence of a reliable difference, $F(1, 50) = 7.01$, $p < .05$. A similar comparison between the 4-minute time interval conditions and the control did not reveal the presence of a significant difference ($F < 1$).

The second experimental hypothesis suggests that focusing on self immediately before or after awareness of the distress of others is increased should result in increased willingness to help those persons. The mean responses to each of the questions designed to measure willingness to help victims of VD are presented in Table 2. Responses to these questions were highly correlated (rs ranging from .59 to .76) and were summed for purposes of analysis. The results of this analysis indicated that subjects in the brief time interval conditions expressed greater overall willingness to help victims of VD than did subjects in the 4-minute time interval conditions, $F(1, 41) = 6.41$, $p < .05$. No other significant main effects or interactions were found. Additional comparisons indicated that subjects in the brief time interval conditions were more willing to help than were subjects in the control condition, $F(1, 50) = 4.20$, $p < .05$, whereas subjects in the 4-minute interval conditions were not ($F < 1$).

Table 2
Mean Willingness to Help Victims of VD

Item	No interval before	No interval after	4-minute interval before	4-minute interval after	Control
Teach class	11.2	11.2	8.5	7.9	7.0
Volunteer work at clinic	7.6	8.9	7.6	6.0	8.2
Personal contribution	6.9	7.7	3.3	6.6	5.2
Combined	25.7	27.8	19.4	20.5	20.4

Note. Numbers represent means for each condition. The higher the number, the greater the willingness to help.

Discussion

The results from Experiment 1 clearly support both experimental hypotheses. Subjects who focused on self immediately before or after the plight of VD victims was made salient attributed greater responsibility for the welfare of that group to self and indicated greater willingness to expend time, effort, and money to help VD victims than did subjects who focused on self 4 minutes before or after seeing the videotape on VD victims. Comparisons between the experimental conditions and control condition indicate that these main effects were due to increases in attribution of responsibility to self and willingness to help in the brief time interval conditions rather than to any decrease in the magnitude of those responses in the 4-minute time interval conditions.

Alternate explanations. It might be argued that putting the subject's image on the screen immediately before or after the videotape had been shown caused subjects to suspect that the experimenter wanted them to feel personally responsible for victims of the VD epidemic. This argument is challenged by the results in the 4-minute time interval/after condition. Subjects in this condition saw their image immediately before they filled out the dependent measures. If subjects were susceptible to the belief, "The experimenter is trying to get me to feel personally responsible for the welfare of victims of VD by putting my face on TV," it seems reasonable that introducing the subjects' images immediately before the dependent measures were administered would be just as likely to make them suspicious about the motives of the experimenter as would introducing their images before or after the videotape. Since subjects in the 4-minute time interval/after condition did not differ from subjects in the control condition, it would be difficult to argue that the experimental results were due to experimental demand.

Aside from an experimental demand explanation, it might be argued that increasing subjects' self-focus in the brief time interval conditions affected helping behavior by increasing their efforts to conform to a generalized norm or standard of helping others (Wicklund, 1975). However, if increasing self-focus affected helping behavior via a direct self-evaluative mechanism (Duval & Wicklund, 1972), it seems reasonable to expect increased willingness to help when self-focus was increased immediately before measures of helping were taken, that is, in the 4-minute time interval/after condition. Since there is no evidence of increased willingness to help in this condition, we must conclude that the self-attribution interpretation of the obtained effects, at least under these experimental conditions, must be given substantial consideration.

Experiment 2

The data from Experiment 1 are consistent with the two experimental hypotheses but do raise several issues. Since the self-attribution measure separated VD victims and subjects as members of the general public into two discrete groups that might be responsible for victims of the VD epidemic, it is conceivable that subjects actually considered themselves as separate from the general public when responding to this question. To resolve this interpretational ambiguity,

the attribution-of-responsibility-to-self measure in Experiment 2 was more direct and specifically asked subjects to determine the extent to which they were responsible for the people in question.

Second, to provide evidence concerning the generality of the proposed relationship between self-focus, felt responsibility, and helping behavior and to demonstrate the conceptual replicability of the results obtained in Experiment 1, Experiment 2 used a different technique to increase self-focus and presented a different population as potential help recipients. To increase self-focus, subjects were asked to fill out a biographical questionnaire, a procedure that should increase self-focus (Duval & Wicklund, 1972) as effectively as a TV camera. The target population consisted of poverty-stricken Latin Americans rather than VD victims.

Third, we have assumed that increasing self-focus affects felt responsibility for a specific group only if awareness of that group's distress is increased immediately before or after self-focus is increased. This may not be the case. Increasing self-focus a few minutes before measuring felt responsibility for a particular group of distressed persons may increase felt responsibility for that group even when awareness of the group's distress has not been increased. To test this possibility, Experiment 2 included a control condition in which subjects filled out the biographical questionnaire but were not exposed to any material designed to increase awareness of poverty-stricken Latin Americans prior to administration of the dependent measures.

Finally, the inclusion of the 4-minute time interval/after condition in Experiment 1 renders an experimental demand or a direct self-evaluation interpretation of the obtained results unlikely. On the other hand, increasing self-focus typically causes attitudes and norms to change in the direction of any relevant and positive reference group's majority position (Duval, 1976; Duval & Wicklund, 1972; Wicklund & Duval, 1971). The existence of this effect suggests that juxtaposing the subject's image and material related to victims of VD may have caused the subject's attitudes and/or norms concerning victims of VD to change in a direction dictating more helping behavior. From this point of view, the increased willingness to help observed in the brief time interval conditions of Experiment 1 might have been due to changes in attitudes and/or norms rather than to the differences in felt responsibility. We view this as an unlikely possibility, since any mention of other people's attitudes and norms regarding victims of VD was carefully avoided. Nevertheless, it is a plausible alternative explanation for the results in Experiment 1 that cannot be ruled out. In fact, no study on felt responsibility and helping behavior has determined if the manipulation of felt responsibility also affects attitudes and norms. Thus, in Experiment 2, the dependent measures included two items designed to measure subjects' attitudes and norms toward the target population, poverty-stricken Latin Americans.

The hypotheses of Experiment 2 were essentially the same as those of Experiment 1. To test these hypotheses, subjects filled out a biographical questionnaire either 4 minutes before or immediately before seeing a videotape on poverty-stricken Latin Americans. It was predicted that subjects in the brief time interval condition would feel more responsibility for and would be more willing to help poverty-stricken Latin Americans than would subjects in the 4-minute time interval condition.

Method

Subjects. Subjects were 55 female undergraduates who volunteered for the study in partial fulfillment of a course requirement. Subjects were randomly assigned to two experimental and three control conditions and were run individually.

Procedure. Subjects were greeted by a male experimenter and were seated in front of a television set. From this point on, the experimenter's explanations of the purpose of the study and the subject's role in the study were similar to the explanations given to subjects in Experiment 1. After the purpose of the experiment had been explained, subjects were told that they would also be asked to fill out a short biographical questionnaire. Subjects were told that this information was needed in order to determine whether participants in the survey were all relatively homogeneous with regard to personality and background. After these instructions, the experimenter's actions were essentially the same as those in Experiment 1.

The 10-minute documentary videotape was composed of segments taken from a film called *Highland Indians of Peru*. These segments consisted of fairly impactful scenes depicting the urban and rural poor of Peru and were accompanied by a spoken commentary. However, the videotape and commentary did not suggest that the film referred to a particular people or country and treated the scenes as though they were characteristic of large portions of the Latin American population. In addition, the videotape did not discuss the causes of poverty in Latin America or responsibility for this situation, nor did it request any form of help for the Latin American poor.

After viewing the videotape, subjects remained seated for an additional 5 minutes while the experimenter supposedly rewound the videotape and shut down the equipment. After the 5-minute waiting period had elapsed, the experimenter reentered the subject's portion of the cubicle and administered the dependent measures.

Time interval manipulation. In the 4-minute time interval condition, the experimenter went behind the partition, reentered the cubicle, and handed the subject a biographical questionnaire saying: "Please take about a minute to fill this out, and let me know when you've finished." After the subject had finished the questionnaire and had returned it to the experimenter, the experimenter waited for 4 minutes. He then played the videotape about poverty-stricken Latin Americans on the subject's TV monitor. In the brief time interval condition, the experimenter turned on the equipment, waited for 4 minutes, and then asked the subject to fill out the biographical questionnaire. After the subject had finished the questionnaire and had returned it, the experimenter played the documentary videotape on the subject's TV monitor.

Biographical questionnaire. The biographical questionnaire was designed to require subjects to think about themselves as objects but purposely avoided asking any questions that might sensitize them to responsibility for the poor in Latin America. The items included on the questionnaire were the following: (a) age; (b) sex; (c) scholastic major; (d) intended area of employment after completing college; (e) hobbies; (f) briefly describe your physical characteristics (size, weight, looks, etc.); (g) briefly describe your most unique characteristic. The item concerning uniqueness was considered the critical question in terms of increasing self-focus, since data (Duval, 1976; Duval & Duval, Note 1; Duval & Siegel, Note 2) indicate that increasing the salience of a person's novelty as an object in the world increases the tendency for attention to focus on self.

Control conditions. In the no-biographical-questionnaire control condition, the experimenter followed the same procedures used in the experimental conditions, except that the biographical questionnaire was not mentioned and the subject did not, in fact, fill out such a questionnaire. The videotape on poverty in Latin America was shown approximately 5 minutes after the experimenter's initial instructions were completed. In the no-videotape control condition, the subject filled out a biographical questionnaire but did not see the videotape concerning poverty-stricken Latin Americans. The biographical questionnaire was administered approximately 4 minutes after the experimenter had completed his initial instructions. In the no-biographical-questionnaire/no-videotape control condition, the subjects neither filled out the biographical questionnaire nor saw the videotape on poverty in Latin America. In all control conditions, the time interval between the end of the experimenter's initial instructions and administration of the dependent measures was approximately the same as the corresponding time interval in the experimental conditions.

Dependent measures. The dependent measures included the following items, listed in order of presentation:

1. The level of poverty suffered by many Latin Americans is totally undesirable.
2. The level of poverty suffered by many Latin Americans ought to be substantially reduced.
3. To what extent do you feel any personal responsibility for the poverty-stricken people of Latin America?
4. To what extent would you be willing to take a 1-hour course dealing with poverty in Latin America and then conduct two half-hour educational sessions with groups of college freshman?
5. Several agencies dealing with poverty in Latin America exist in Los Angeles but are understaffed. How much time per week would you be willing to spend doing volunteer work at one of these agencies?
6. How large a personal financial contribution would you be willing to make to the Agency for Latin American Development?

All questions were accompanied by 15-point scales anchored with the appropriate positive and negative levels. After subjects had completed the questionnaire, the experimenter checked for suspicion and informed the subject of the actual purpose of the experiment and experimental procedures.

Results

Data from Experiment 1 indicated that subjects in the brief time interval conditions attributed more responsibility for victims of VD to the general public including self. This was interpreted as reflecting greater attribution of responsibility to self. As was mentioned, however, this measure could have reflected the subject's attribution of responsibility to the general public but not to self. Experiment 2 attempted to resolve this question by using a more directly phrased question concerning the subject's attribution of

Table 3
Mean Attributed Responsibility and Willingness to Help

Item	Brief time interval	4-minute time interval	No bio control	No video control	No bio control/no video control
Felt responsibility	5.21_a	3.06_b	2.81_b	2.54_b	3.19_b
Teach class	8.31	5.10	4.4	3.46	3.77
Volunteer work at clinic	7.79	3.57	4.32	4.42	3.58
Personal contribution	6.24	3.29	4.00	3.59	4.20
Combined	22.34_a	11.96_b	12.72_b	11.45_b	11.65_b

Note. For each condition, $n = 11$. Numbers represent means for each condition. The higher the number, the greater the responsibility attributed to self and willingness to help. Means with similar subscripts do not differ.

responsibility to self. As predicted (see Table 3), subjects who completed the biographical questionnaire immediately before seeing the documentary videotape on poverty-stricken Latin Americans attributed more responsibility to self than did subjects who completed the questionnaire 4 minutes before viewing the videotape, $t(20) = 2.57$, $p < .02$. An additional analysis using the Newman–Keuls procedure indicated greater attributed self-responsibility in the brief time interval condition than in any other condition ($p < .05$).

Data from Experiment 2 also replicate the findings in Experiment 1 with regard to willingness to help the target population (Table 3). Subjects in the brief time interval condition exhibited greater willingness to expend time, effort, and money (*r*s ranged between .76 and .82 and were summed for analysis) to help the poor of Latin America than did subjects in the 4-minute time interval condition, $t(20) = 2.66$, $p < .02$. An additional analysis using the Newman–Keuls procedure indicated greater willingness to help in the brief time interval condition than in any other condition ($p < .05$).

Experiment 2 was also designed to determine whether the experimental manipulation of self-focus affected subjects' attitudes and/or norms concerning poverty-stricken Latin Americans. As is evident in Table 4, expressed attitudes and norms did not differ between conditions ($p > .20$).

Conclusions

A number of conclusions regarding the relationships between self-focus, attribution of responsibility to self, and helping behavior can be drawn from the results obtained in Experiments 1 and 2. First, it appears that increasing self-focus increases self-attribution for a situation or event. However, this relationship obtains only to the extent that the periods of increased self-focus and awareness of the situation or event are separated by a relatively short time interval. These results suggest that the unit formation principle of temporal proximity does apply to the time interval that separates periods of increased self-focus and increased awareness of situations and events, as well as to the time interval that separates the actual presence

Table 4
Mean Attitude and Norm Regarding Poverty in Latin America

Level of poverty	Immediately before	4 minutes before	No bio control	No video control	No bio control/no video control
Totally undesirable	13.34	12.86	13.14	10.40	9.82
Ought to be substantially reduced	12.85	13.22	14.67	12.21	13.74

Note. Numbers represent means for each condition. The higher the number, the stronger the attitude or norm.

or occurrence of agents, situations, and events. In and of itself, this conclusion has obvious implications for dealing with both the theoretical and empirical relationship between self-focus and self-attribution.

Second, juxtaposing self-focus and a salient group of distressed persons increases willingness to expend time, effort, and money to help that group. The failure to find increased self-attribution and willingness to help in the 4-minute interval/after condition of Experiment 1 suggests that an experimental demand or a direct self-evaluation interpretation (Duval & Wicklund, 1972) of this relationship is not highly plausible. In Experiment 2, increasing self-focus produced differences in felt responsibility for and willingness to help the poor in Latin America without differentially affecting subjects' attitudes and norms toward the target population and their difficulties. These data suggest that the effect of self-focus on helping behavior is not caused by change in attitudes and/or norms. Having ruled out these alternative explanations, we conclude that juxtaposing self-focus and a salient group of distressed persons affects helping behavior by increasing felt responsibility for the welfare of that group. However, this conclusion must be tempered by the results obtained by Federoff and Harvey (1976), which suggest that some as yet unidentified variable may affect the relationship between self-focus and internal versus external attribution for a negative event.

The obtained relationship between self-focus and helping behavior has other implications for the study of helping behavior. Specifically, any conditions that affect self-focus when distressed others are salient should affect felt responsibility and thus helping behavior. Since the focus of attention is generally attracted to "important" or "novel" stimuli, conditions that increase or decrease the relative situational importance or novelty of an individual when distressed others are salient should increase or decrease that person's level of self-focus, felt responsibility, and tendency to help the needy others. The relevance of this proposition for the general study of helping behavior has been demonstrated in a recent experiment. Using the notion that the novelty of two interacting groups is affected by the relative number of people in each group (Duval, 1976; Duval & Wicklund, 1972), Wegner and Schaefer (1978) hypothesized that increasing the number of subjects who could help a single victim should decrease subjects' levels of situational novelty, self-focus, and helping behavior. Similarly, increasing the number of victims confronting a single subject should increase the subject's level of situational novelty, self-focus, and helping behavior. These hypotheses were confirmed. Although felt responsibility was not measured, subjects run in groups of three tended to focus on self and to help a needy person less than did subjects run singly; individual subjects confronted with three victims focused on self and helped those victims more than did individual subjects faced with a single victim.

The novelty/self-focus/felt responsibility/helping behavior relationship may also apply to the relationship between distinctive suitability (Schwartz, 1970), dependency (Berkowitz, 1972), and helping. A person who is one of a "limited and uniquely qualified" group of potential helpers (high distinctive suitability) would be more novel in that situation than a person who is merely one of many people who could help (low distinctive suitability, Schwartz, 1970). To the extent that a person is depended on by others, that person is defined as uniquely important with respect to helping others. Thus, increasing distinctive suitability or dependency may increase a person's tendency to help others in distress because those manipulations affect the individual's level of situational novelty or importance, self-focus, felt responsibility and, thus, helping behavior.

In summary, research has found a strong empirical relationship between felt responsibility and helping behavior. The present research demonstrates that juxtaposing self-focus with a salient group of needy others increases felt responsibility for and willingness to help those others. We have suggested that the effects of distinctive suitability, dependency, and relative number of potential helpers and victims on helping behavior may be mediated by self-focus and attribution of responsibility to self. Thus, the self-focus/felt responsibility relationship may have sub-

stantial significance for the general study of helping behavior.

Reference Notes

1. Duval, S., & Duval, V. H. *Novelty and causal attribution.* Paper presented at the meeting of the Western Psychological Association, Seattle, April 1977.
2. Duval, S., & Siegal, K. N. *Some determinants of objective self-awareness.* Paper presented at the meeting of the American Psychological Association, Toronto, August–September 1978.

References

Arkin, R., & Duval, S. Focus of attention and causal attributions of actors and observers. *Journal of Experimental Social Psychology,* 1975, *11,* 427–438.

Berkowitz, L. Social norms, feelings, and other factors affecting helping and altruism. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6). New York: Academic Press, 1972.

Duval, S. Conformity on a visual task as a function of personal novelty on attitudinal dimensions and being reminded of the object status of self. *Journal of Experimental Social Psychology,* 1976, *12,* 87–98.

Duval, S., & Duval, V. H. *Maximizing similarities and minimizing differences: A cognitive consistency theory of causal attribution.* Hillsdale, N.J.: Erlbaum, in press.

Duval, S., & Hensley, V. Extensions of objective self-awareness theory: The focus of attention-causal attributoin hypothesis. In J. H. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research.* Hillsdale, N.J.: Erlbaum, 1976.

Duval, S., & Wicklund, R. A. *A theory of objective self-awareness.* New York: Academic Press, 1972.

Duval, S., & Wicklund, R. A. Effects of objective self-awareness on attribution of causality. *Journal of Experimental Social Psychology,* 1973, *9,* 17–31.

Federoff, N. A., & Harvey, J. H. Focus of attention, self-esteem, and attribution of causality. *Journal of Research in Personality,* 1976, *10,* 336–345.

Geer, J. H., & Jarmecky, L. The effect of being responsible for reducing another's pain on subject's response and arousal. *Journal of Personality and Social Psychology,* 1973, *26,* 232–237.

Heider, F. *The psychology of interpersonal relationships.* New York: Wiley, 1958.

Ickes, W. J., Kidd, R. F., & Berkowitz, L. Attributional determinants of monetary help giving. *Journal of Personality,* 1976, *44,* 163–178.

Jones, E. E. How do people perceive the causes of behavior. *American Scientist,* 1976, *64,* 300–305.

Michotte, A. *The perception of causality.* New York: Basic Books, 1963.

Piliavin, I. M., Rodin, J., & Piliavin, J. A. Good Samaritanism: An underground phenomenon? *Journal of Personality and Social Psychology,* 1969, *13,* 289–299.

Pryor, J. B., & Kriss, M. The cognitive dynamics of salience in the attribution process. *Journal of Personality and Social Psychology,* 1977, *35,* 49–55.

Schopler, J., & Matthews, M. The influence of perceived causal locus of partner's dependence on the use of interpersonal power. *Journal of Personality and Social Psychology,* 1965, *2,* 609–612.

Schwartz, S. H. Elicitation of moral obligation and self-sacrificing behavior: An experimental study of volunteering to be a bone marrow donor. *Journal of Personality and Social Psychology,* 1970, *15,* 283–293.

Schwartz, S. H. Normative explanations of helping behavior: A critique, proposal and empirical test. *Journal of Experimental Social Psychology,* 1973, *9,* 349–364.

Schwartz, S. H., & Clausen, G. T. Responsibility, norms, and helping in an emergency. *Journal of Experimental Social Psychology,* 1970, *16,* 299–310.

Storms, M. D. Videotape and the attribution process: Reversing actor's and observer's points of view. *Journal of Personality and Social Psychology,* 1973, *27,* 165–175.

Taylor, S., & Fiske, S. T. Point of view and perceptions of causality. *Journal of Personality and Social Psychology,* 1975, *32,* 439–445.

Tilker, H. A. Socially responsible behavior as a function of observer responsibility and victim feedback. *Journal of Personality and Social Psychology,* 1970, *14,* 95–100.

Wegner, D. M., & Schaefer, D. The concentration of responsibility: An objective self-awareness of group size effects in helping situations. *Journal of Personality and Social Psychology,* 1978, *36,* 147–155.

Wicklund, R. A. Objective self-awareness. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 8). New York: Academic Press, 1975.

Wicklund, R. A., & Duval, S. Opinion change and performance facilitation as a result of objective self-awareness. *Journal of Experimental Social Psychology,* 1971, *7,* 319–342.

Received November 6, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 10, 1779–1789

Sex Role Identity and Its Relationships to Sex Role Attributes and Sex Role Stereotypes

Michael D. Storms
University of Kansas

Kagan's and Kohlberg's theories of sex role identity are examined in this paper. Both theories premise relationships among an individual's sex role identity, his/her perception of sex role stereotypes, and his/her own masculine and feminine sex role attributes, but the theories specify different patterns of correlations among these variables. Kagan's theory views sex role identity as the product of the mutual influence of stereotypes and attributes, whereas Kohlberg's theory posits sex role attributes as the product of stereotypes and identity. College men and women were administered measures of each variable including the Personal Attributes Questionnaire (PAQ; Spence and Helmreich), modified versions of the PAQ to measure stereotypes of men and women, and a measure of sex role identity developed by the present author. The results indicated an overall relationship between sex role identity and the other two variables. Beyond that, the results generally supported Kohlberg's theory over Kagan's. Significant correlations were found where Kohlberg's theory would predict a relationship—between sex role stereotypes and sex role attributes. Nonsignificant correlations were found where Kagan's theory would predict a relationship—between sex role stereotypes and sex role identity. These results are compared to earlier findings by Spence, Helmreich, and Stapp, and questions are raised about the psychological constructs underlying measures of sex role attributes.

Over the past decade, researchers have amassed considerable data on sex role stereotypes (e.g., Rosenkrantz et al., 1968) and masculine and feminine sex role attributes (e.g., Bem, 1974; Spence & Helmreich, 1978). Relatively less attention has been paid to self-concepts of masculinity and femininity—that is, to sex role identity. Although sex role identity has existed as a theoretical construct for some time (Kagan, 1964; Kohlberg, 1966), there has been no widely accepted measure of sex role identity and little research on its relationships to sex role attributes and sex role stereotypes.

Kagan (1964) and Kohlberg (1966) have both defined sex role identity as an acquired self-concept of being masculine or feminine, and they have proposed that sex role identity is closely related to sex role stereotypes and sex role attributes. The two theorists differ sharply, however, about the causal structure of these relationships. Kagan postulates that sex role identity is the product of differences between individuals' sex role attributes and their perceptions of sex role stereotypes; Kohlberg believes that sex role identity is the cause of those differences. Figure 1 illustrates the causal structure of these two theories.

Specifically, Kagan (1964) theorizes that individuals compare their own attributes against sex role stereotypes to arrive at a relativistic self-concept of their sex role identity. In Kagan's (1964) words:

Sex-role identity represents the degree to which an

This research was supported in part by University of Kansas General Research Fund Grants 3800 and 3313. The author is indebted to Robert Helmreich, Kevin McCaul, and Shelley Taylor for their comments on an earlier draft and to Scott Lambers for his invaluable research assistance.

Requests for reprints should be addressed to Michael D. Storms, Department of Psychology, Fraser Hall, University of Kansas, Lawrence, Kansas 66045.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3710-1779$00.75

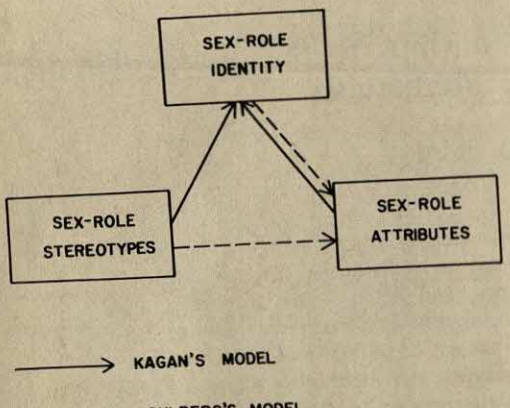


Figure 1. Kagan's (1964) and Kohlberg's (1966) models of the relationships among sex role identity, attributes, and stereotypes.

individual regards himself as masculine or feminine. . . . The degree of match or mismatch between the sex-role standards of the culture and an individual's assessment of his own overt and covert attributes provides him with a partial answer to the question, "How masculine (feminine) am I?" (p. 144)

In short, Kagan views sex role identity as the product of sex role stereotypes and sex role attributes.

In contrast, Kohlberg (1966) asserts that sex role identity is firmly established early in life and thereafter serves to mediate the influence of sex role stereotypes on the individual's development of sex role attributes. Sex role identity guides the individual's attachment to, evaluation of, and desire to emulate sex role prototypes (specific adult role models) and sex role stereotypes (abstract role models). Sex role attributes, according to Kohlberg's theory, are the product of the individual's sex role identity and perception of sex role stereotypes.

The theories described above lead to many similar hypotheses about the relationships among sex role identity, sex role stereotypes, and sex role attributes. Both theories predict a multiple correlation among the three sex role variables—knowing any two of the variables should enhance prediction of the third. Both theories also predict a correlation between sex role attributes and sex role identity, although for opposite reasons—either because attributes influence identity (Kagan) or because identity influences attributes (Kohlberg).

The two theories differ most noticeably in their hypotheses about the relationships between sex role stereotypes and the other two variables. According to Kagan's theory, sex role stereotypes and sex role attributes interact to determine sex role identity. Thus, Kagan's theory predicts a correlation between stereotypes and identity but not (necessarily) a correlation between stereotypes and attributes. Conversely, according to Kohlberg's theory, sex role stereotypes and sex role identity interact to determine sex role attributes. Kohlberg's theory predicts a correlation between stereotypes and attributes but not (necessarily) between stereotypes and identity.

The relationship between sex role stereotypes and sex role attributes has already been investigated in a study by Spence, Helmreich, and Stapp (1975). Briefly, they found no correlation between these two variables—a finding that would support Kagan's theory and contradict Kohlberg's. Unfortunately, there are several methodological problems with the Spence et al. study that cast some doubt on the relevance of their data to the present discussion. Examination of these problems will require a detailed review of the procedures used.

The main purpose of Spence, Helmreich, and Stapp's (1975) study was to introduce a new measure of sex role attributes, the Personal Attributes Questionnaire (PAQ). Spence et al. created the PAQ to measure stable, underlying personality traits associated with masculinity and femininity, but they were also aware of a possible alternative interpretation of their scale. Spence et al. noted that items on masculinity–femininity scales (including the PAQ) are usually adopted from lists of sex role stereotypes. Thus, an individual's endorsement of a particular item could merely indicate his/her knowledge of and identification with the corresponding stereotype and not an accurate reflection of his/her actual traits.

Spence, Helmreich, and Stapp (1975) proposed a test of the competing stereotype interpretation of the PAQ. They reasoned that according to the stereotype hypothesis, people's self-ratings on masculinity and feminin-

ity scales should be correlated with their perceptions of sex role stereotypes. For example, people who believe that a particular trait is very characteristic of their sex in general should rate themselves higher on that trait than will people who think the trait is less characteristic of their sex. In contrast, according to the trait hypothesis, self-ratings reflect genuine attributes of individuals independent of their beliefs about sex roles, and the two should not be correlated. People who have more of a trait should rate themselves higher on that trait independently of their beliefs about how much of the trait others of their sex have.

Spence, Helmreich, and Stapp (1975) selected 55 traits that pretest subjects had rated as more typical of one sex than the other. These items were sorted into three scales—two independent masculinity (M) and femininity (F) scales and a third, bipolar masculinity–femininity (M–F) scale. Male and female college students first rated themselves on each of the 55 items, then rated each item for the typical male or female. Eighteen correlations were generated by comparing male and female subjects' self-ratings on each scale to their stereotype ratings on all three scales. Finding only 5 of these 18 correlations statistically significant, Spence et al. rejected the stereotype hypothesis and concluded that the PAQ measures genuine traits. In the present context, these results could also be interpreted as evidence against Kohlberg's theory of sex role identity, which also predicts a correlation between sex role stereotypes and sex role attributes.

Unfortunately, there are at least three methodological problems with the Spence, Helmreich, and Stapp (1975) study that cast doubt on their conclusions. First, Spence et al. used different item formats for subjects' stereotype ratings than for their self-ratings. For self-ratings, subjects were presented with bipolar items (e.g., *very passive* to *very active*), whereas for stereotype ratings subjects were presented with single labels (e.g., *active*). These two formats could have created different psychological dimensions for the subjects; a subject who saw only the word *active* may have assumed that the intended opposite was *inactive* rather than *passive*. Thus, subjects may have rated stereotypes along dimensions somewhat different from those they used in rating themselves. That would have greatly reduced any possibility of obtaining correlations between the attribute and stereotype ratings.

A second problem was created by the particular type of rating scale that followed each stereotype item. Each single label was followed by a bipolar, unidimensional scale ranging from *much more characteristic of the male* to *much more characteristic of the female*. The problem with using unidimensional scales to measure stereotypes is that ratings in the middle of such scales are ambiguous. A middle rating could mean that the trait was seen as either highly characteristic of both sexes or not characteristic of either sex. That ambiguity in turn could have reduced correlations between attribute and stereotype ratings. According to the stereotype hypothesis, subjects who thought a particular trait highly characteristic of both sexes would have rated the trait in the middle of the stereotype scale, but they would have rated themselves high on the trait. On the other hand, subjects who saw the trait as uncharacteristic of either sex would also have rated the trait in the middle of the stereotype scale, but they would have rated themselves low on the trait. These two possibilities would cancel each other out and artifactually lower the correlation between attribute and stereotype ratings.

A final problem with Spence, Helmreich, and Stapp's (1975) study concerns the 18 correlations they presented as tests of the stereotype hypothesis. The 18 correlations were generated by comparing attribute ratings on each of three different scales to stereotype ratings on all three scales separately for each sex. But two of the three attribute scales (the M scale and the F scale) were theoretically designed and empirically shown to be uncorrelated with each other. If attribute ratings alone were not expected to correlate between scales, it was inappropriate to expect attribute ratings and stereotype ratings to correlate between scales. Eliminating these cross-scale comparisons, only 6 of the 18 possible correlations actually provide a valid test of the

stereotype hypothesis—namely, self-ratings on each scale compared to stereotype ratings only on the *same scale*, calculated separately for each sex on each scale. Of these 6 correlations, 3 were significant and in the predicted direction. Viewed thus, Spence, Helmreich, and Stapp's (1975) results provide highly equivocal evidence about the stereotype hypothesis and not definitive evidence against it. It would also be unfair to rule against Kohlberg's theory of sex role identity on the basis of those results.

The present study was designed to investigate Kagan's (1964) and Kohlberg's (1966) models of sex role identity, using a paradigm similar to Spence, Helmreich and Stapp's (1975) but with several methodological improvements. First, subjects were asked to make attribute ratings and stereotype ratings on identical sets of bipolar items. Second, subjects' stereotypes were measured by taking separate ratings of the typical male and the typical female on all items. Third, attribute ratings and stereotype ratings were compared only within each scale and not between scales. Finally, a measure of sex role identity was included.

Method

Subjects and Procedure

One hundred four male and 110 female students in introductory social psychology at a midwestern state university completed a battery of four instruments described below. The batteries were completed during one class period at the beginning of the semester. Subjects' responses were voluntary and anonymous. The instructor in the course debriefed the class and presented a lecture based on the results later in the semester.

Instruments

Because no widely accepted measure of sex role identity presently exists, a new instrument was designed for the present study.[1] In keeping with the definition of sex role identity as "a global self-concept of one's masculinity and femininity," the Sex Role Identity Scale (SRIS) was written to include extremely simple, face-valid queries about self-concept. Subjects were asked to rate themselves on three masculine and three feminine items: "How masculine (feminine) is your personality?" "How masculine (feminine) do you act, appear, and come across to others?" and "In general, how masculine (feminine) do you think you are?" Each item was followed by a 31-point scale with the endpoints labeled *not at all masculine (feminine)* (scored as 0) to *extremely masculine (feminine)* (scored as 30).

Subjects' sex role attributes were measured by Spence and Helmreich's (1978) Personal Attributes Questionnaire, a shortened version of the Spence, Helmreich, and Stapp (1975) measure described earlier. It contains 24 of the original 55 items, still categorized into three subscales—an M scale, an F scale, and a unidimensional M–F scale. Each scale contains 8 items; each item is scored from 0 to 4; and total scores on each scale can range from 0 to 32. Higher values indicate greater endorsement of M items, F items, or the masculine pole of M–F items.

Subjects' stereotypes were measured by two modified versions of the PAQ. One questionnaire asked subjects to rate the "typical male," and one asked subjects to rate the "typical female." Thus, stereotype ratings for each sex were gathered separately using exactly the same scales as the self-ratings. The same scoring system was used for each stereotype scale as was used for the corresponding self-rating scale.

The order of the four questionnaires in the battery was randomized across subjects, except that the two stereotype questionnaires always appeared together.

Results

The Sex Role Identity Scale

Since a new measure of sex role identity was designed for this study, its psychometric properties were examined first. The scale showed strong internal consistency: the three masculine identity items intercorrelated positively better than .66 for men and .68 for women (all $ps < .001$), and the three feminine identity items intercorrelated positively better than .80 for men and .70 for women (all $ps < .001$). More important, despite all the evidence of multidimensionality in people's sex role attributes (Bem, 1974; Constantinople,

[1] Two researchers have designed instruments that may appear related to sex role identity. First, Stein (1971) has measured *sex role preference*, which may share some conceptual meaning with sex role identity. The two concepts are not identical, however. Logically, one can prefer to have sex role characteristics that differ from one's present sex role self-concept. Second, Heilbrun (1976) has published a measure called "sex role identification"; however, the composition and use of that instrument suggest that it measures specific sex role attributes rather than a global sex role self-concept.

Table 1
Means, Standard Deviations, and p *Values for Male–Female Differences on all Measures*

Scale	Men		Women		
	M	*SD*	*M*	*SD*	*p*
Sex role identity	24.43	4.25	7.86	5.26	.001
Attribute self-ratings					
M	22.58	4.78	20.15	4.30	.001
F	21.56	4.41	23.92	3.38	.001
M–F	16.77	4.54	12.87	4.07	.001
Stereotype ratings of men					
M	24.68	3.96	24.08	3.80	*ns*
F	17.10	5.00	17.22	4.76	*ns*
M–F	20.53	4.32	19.83	4.17	*ns*
Stereotype ratings of women					
M	15.75	4.13	16.94	4.99	*ns*
F	24.70	4.10	24.18	3.69	*ns*
M–F	9.09	3.79	10.84	3.95	.001

Note. Higher values indicate greater endorsement of masculine items on M scales, of feminine items on F scales, and of the masculine pole of items on the M–F scales and the Sex Role Identity Scale. For men, $n = 104$. For women, $n = 110$. M = masculine. F = feminine.

1973; Spence & Helmreich, 1978), subjects in this study conceptualized their sex role identities along a single dimension. Masculine identity items correlated negatively with feminine identity items better than −.64 for men and −.74 for women (both $ps < .001$). Therefore, the six items were combined by scoring the three feminine items in the opposite direction from the three masculine items, summing, and dividing by six—thus yielding a 31-point, bipolar sex role identity score ranging from 30 (most masculine/least feminine) to 0 (least masculine/most feminine).

Sex Differences on All Measures

All of the hypotheses in the present study presume the existence of sex-of-subject differences in sex role identity and attributes and sex-of-target differences in sex role stereotypes. Table 1 illustrates these differences. Male and female subjects reported not only significant but extremely large differences on the measure of sex role identity, with means of 24.4 and 7.9, respectively, $p < .001$. Males and females also reported significant differences on all three subscales of the sex role attributes measure, the PAQ. Men reported higher self-ratings on the M scale and the bipolar M–F scale, whereas women reported higher self-ratings on the F scale.

Although men and women differed in their attribute and identity self-ratings, they agreed in their stereotypes about the typical man and the typical woman. Men and women differed in their stereotype ratings in only one instance: Women rated the typical woman higher on the M–F scale. Men and women also agreed on the differences between the typical man and the typical woman, as can be seen by making vertical comparisons between the two stereotype scales in Table 1. Both sexes rated the typical man higher on the M scale and the M–F scale and rated the typical woman higher on the F scale (all $ps < .001$). In short, all of the instruments produced results consistent with past research and consistent with the assumptions underlying this study.

Correlations Among Measures

Correlations among identity, attribute, and stereotype ratings are presented in Table 2. Before discussing these correlations, I should explain the structure of Table 2. As argued earlier, it is appropriate to consider only those correlations occurring within each subscale of

Table 2
Correlations Among Sex Role Identity, Sex Role Stereotypes, and Sex Role Attributes for Men and Women

Correlation	Attribute/stereotype scales					
	Men			Women		
	M	F	M–F	M	F	M–F
Multiple						
$I = AS$	.43***	.20*	.34***	.23**	.42***	.27**
$A = IS$	.53***	.19*	.39***	.31***	.43***	.41***
$S = AI$	.39***	.26**	.21*	.35***	.22*	.33***
Zero-order and partial						
AI	.43***	−.05	.34***	.07	−.41***	.27**
$AI–S$	.39***	−.08	.34***	.13	−.39***	.27**
SI	.19*	.19*	.06	−.19*	−.17*	.05
$SI–A$	.04	.19*	−.02	−.22*	−.10	−.16
SA	.39***	.17*	.21*	.28***	.20*	.33***
$SA–I$	.34***	.19*	.20*	.30***	.14	.33***

Note. Variables are abbreviated: A = attributes, S = stereotypes, and I identity. In the multiple correlations the first variable is the dependent variable. In the rest of the table, two variable initials indicate a zero-order correlation and three variable initials indicate a partial correlation between the first two variables controlling for the third variable. For men, $n = 104$. For women, $n = 110$. An equals sign indicates multiple correlations, whereas a dash indicates partial correlations. M = masculine. F = feminine.
* $p < .05$. ** $p < .01$. *** $p < .001$.

the PAQ and not between scales. Since every correlation in this study involved one or both of the subscaled instruments (i.e., the self-rating version of the PAQ and/or the stereotype-rating version), each correlation was calculated six times, once for each subscale for each sex—thus yielding the six columns in Table 2. The hypotheses were tested by examining the three multiple correlations among the three variables, the three zero-order correlations between each pair of variables, and the three partial correlations between each pair of variables controlling for the third variable—thus yielding the nine rows in Table 2.[2]

The multiple correlations among sex role identity, sex role attributes, and sex role stereotypes are presented first in Table 2. Both sex role identity theories presume that sex role identity is involved in a casual structure with sex role attributes and sex role stereotypes. For Kagan, sex role identity is the product of this structure; for Kohlberg, sex role attributes are the product. Kagan's theory thus leads one to examine the multiple correlation with sex role identity as the dependent variable. Kohlberg's theory is more appropriately tested with sex role attributes as the dependent variable. In fact, both of those multiple correlations, and even the correlation with stereotypes as the dependent variable, were significant for both sexes on each of the three attribute/stereotype subscales, as seen in the top three rows of Table 2. One might notice that Kohlberg's correlations (second row) accounted for somewhat more variance than Kagan's correlations (first row) in four of the six comparisons—10% more in two instances (men on the M scale and women on the M–F scale) and about 5% more in two instances (men on the M–F scale and women on the M scale). This is far from a crucial test between the two theories, however; both theories are essentially supported by these results.

[2] Many other analyses were performed on these data, including all possible part correlations, correlations of the differences between two variables and the third, and correlations of the differences between male and female stereotypes and the other variables. All of these auxiliary analyses produced results entirely consistent with the main analyses presented in Table 2 and so are not reported here.

The two theories also agree that significant correlations should obtain between sex role attributes and sex role identity (although they disagree about the causal mechanisms that should produce these correlations). Given the direction in which the various scales are coded, these correlations should be positive for both sexes on the M and M–F subscales and negative on the F subscale. The correlations presented in the fourth and fifth rows of Table 2 strongly supported both theories in four of the six comparisons. For both the zero-order and the partial correlations, men reported a significant correspondence between their sex role identity and their sex role attributes on the M and M–F scales of the PAQ. Women reported a similar correlation between their sex role identity and sex role attributes on the F scale and the M–F scale. Thus it appears that sex role identity was related to sex role attributes, but only on those scales that were most appropriate for the subject's own sex—that is, the masculinity scale for men, the femininity scale for women, and the masculinity–femininity scale for both sexes.

A relationship between sex role identity and sex role attributes was also evident in an analysis of subjects classified according to sex role type. Table 3 presents the distribution of subjects across the four categories of traditional masculine, traditional feminine, androgynous, and undifferentiated. The percentage of subjects classified into each category corresponded closely to the normative distribution of college men and women reported by Spence and Helmreich (1978). More important, one-way analyses of variance (least squares method) indicated strong sex role identity differences among the groups, $F(3, 100) = 6.59$, $p < .001$ for men and $F(3, 106) = 6.55$, $p < .001$ for women.

Further analysis of sex-typed groups confirmed that sex role identity was related only to masculine attributes for men and only to feminine attributes for women. Men with higher M scores (androgynous and masculine men) reported more masculine sex role identities than men with lower M scores (feminine and undifferentiated men). In contrast, scores on the F scale were irrelevant. Although androgynous men have higher F scores than masculine men, they did not report a significantly more feminine sex role identity. Nor did feminine men report a more feminine identity than undifferentiated men. Similarly, women who scored higher on the F scale (androgynous and feminine women) reported more feminine sex role identities than women who scored lower on the F scale (masculine and undifferentiated women). But women who scored higher on the M scale (androgynous vs. feminine women and masculine vs. undifferentiated women) did not report more masculine sex role identities.

Table 3
Number, Percentage, and Sex Role Identity Means of Men and Women by PAQ Category

PAQ category	Men			Women		
	n	%	*M*	*n*	%	*M*
Androgynous	32	31	24.82[a]	36	33	6.80[a]
Masculine	41	39	26.02[a]	13	11	12.46[b]
Feminine	11	10	22.14[b]	38	34	6.28[a]
Undifferentiated	20	19	21.78[b]	23	21	9.50[b]

Note. Higher mean values indicate a more masculine sex role identity. Within sex, means *not* sharing the same superscript are significantly different at $p < .05$ by weighted contrasts. PAQ = Personal Attributes Questionnaire.

The analyses presented so far can be interpreted as supporting both Kagan's (1964) and Kohlberg's (1966) models of sex role identity, with the exception of subjects' attribute ratings on cross-sex scales, as described above. There are two other correlational analyses, however, on which Kagan's and Kohlberg's theories make distinctly different predictions. First, Kagan's theory predicts a causal link, and therefore a correlation, between sex role stereotypes and sex role identity. Furthermore, these correlations should have the opposite sign of the correlations between attributes and identity (row 4 of Table 2). To illustrate this logic by example: The more masculine *attributes* a man has the more likely he is to compare "favorably" with male stereotypes and infer a highly masculine identity for himself, but the more masculine he perceives the *stereotypic* male to

be, the more likely he is to compare unfavorably and infer a less masculine identity. Kohlberg's theory does not predict any of these correlations because it views sex role stereotypes and sex role identity as two separate, independent variables.

The correlations between sex role stereotypes and sex role identity are presented in the sixth and seventh rows of Table 2. Examining the zero-order correlations first, the predictions derived from Kagan's theory were confirmed in only two of the six comparisons (men on the F scale and women on the M scale). In fact, two of the six correlations were significant ($p < .05$) in the opposite direction from that predicted by Kagan's theory. In case these correlations were contaminated by the influence of sex role attributes, the partial correlations between stereotypes and identity were also calculated. Both of the significantly reversed correlations then disappeared, but no new correlations reached significance in the predicted direction.

Kagan's and Kohlberg's models also lead to opposing predictions about the relationship between sex role attributes and sex role stereotypes. Kohlberg's theory posits a causal link, and therefore a correlation, between sex role stereotypes and sex role attributes. Furthermore, all of these correlations should be positive across the three attribute subscales and for both sexes. The more one sees one's own stereotypic sex as having *any* attribute, the more one should develop that attribute. Kagan's theory predicts no correlations between sex role attributes and sex role stereotypes because it views these as separate, independent variables. The correlations between sex role stereotypes and sex role attributes are presented in the last two rows of Table 2. All six zero-order correlations and five of the six partial correlations reached significance in the direction predicted from Kohlberg's theory.

Discussion

The Nature of Sex Role Identity

This study examined a relatively overlooked concept in current sex role research—sex role identity. A very simple six-item measure of sex role identity was introduced and was shown to have solid internal reliability. Although the instrument contained separate masculine and feminine identity items, thus allowing for multidimensional responses, subjects conceptualized their sex role identity along a single, bipolar dimension. This is not surprising, given that until very recently everyone, including psychologists, viewed masculinity and femininity as diametric. Despite research evidence that masculine and feminine sex role *attributes* form orthogonal dimensions, and despite the increased currency of the term *androgyny*, people are likely to persist in thinking of masculinity and femininity along a single continuum.

This measure of sex role identity was then compared to measures of sex role stereotypes and sex role attributes taken from Spence and Helmreich's (1978) instrument, the PAQ. Two theories of sex role identity were used to derive hypotheses about the relationships among these three sex role variables. Kagan's (1964) theory posits that individuals compare self-perceptions of their own sex role attributes with their perceptions of sex role stereotypes and then infer their relative sex role identity. Kohlberg's (1966) theory posits that people already have firmly established and relatively unchanging sex role identities that guide the development of sex role attributes by means of the modeling or emulating of sex role stereotypes. Both theories predict some shared correlational patterns—namely, that the three sex role variables will form a significant multivariate relationship and, more specifically, that sex role identity and sex role attributes will correlate. On the whole, these predictions were strongly confirmed in the present study, with one exception to be discussed later.

The two theories differ, however, in their predictions about the relationship between sex role stereotypes and the other two variables; Kagan's theory predicts a relationship between stereotypes and identity, whereas Kohlberg's theory predicts a relationship between stereotypes and attributes. Virtually no support was found for the predictions derived from Kagan's theory. Extremely strong support was found for the predictions derived from Kohlberg's theory.

Obviously, the results of a single study may be open to a number of refutations and alternative explanations. Proponents of Kagan's theory could argue, for example, that the instruments used in this study are invalid or measure the wrong sex role attributes and stereotypes and that people base their sex role identities on some other set of traits and stereotypes not tapped by the PAQ. That argument seems implausible, however, for the following reasons. First, there is a wealth of evidence supporting the PAQ as a valid measure (Spence & Helmreich, 1978, 1979). More important, the results of this study demonstrated that sex role identity, attributes, and stereotypes formed a meaningful, coherent, and consistent pattern in nearly every analysis in this study except those analyses derived from Kagan's theory.

Kohlberg's theory of sex role identity was overwhelmingly supported by the present results, with one qualifying detail. It seems that sex role identity influences the development of same-sex-typed attributes (masculine attributes for men and feminine attributes for women) but has little influence on the development of opposite-sex-typed attributes. In retrospect, this makes logical sense. Given that sex role identity is a unidimensional construct, if it were to influence the development of *both* same- and opposite-sex-typed attributes, those attributes would also be unidimensional. Thus, the orthogonality of masculine and feminine attributes and the development of androgynous attributes in some individuals are possible only because the influence of sex role identity is limited to same-sex-typed attributes.

Kohlberg's theory has many important ramifications for future research in the area of sex roles. Conceptually, it suggests that sex role identity may be one of the most powerful and central variables in guiding the development of sex role attributes, in filtering the impact of sex role stereotypes on the individual, and in moderating the influence of situational variables on sex role behavior. Practically, it suggests that sex role identity is a fundamental variable to assess in research on any sex role topic.

The Nature of Sex Role Attributes

Researchers have long debated the meaning of masculinity and femininity or, more precisely, the psychological constructs measured by masculine and feminine attribute scales (Constantinople, 1973; Locksley & Colten, 1979; Pedhazur & Tetenbaum, 1979; Rosenkrantz et al., 1968). Do such scales reflect true masculine and feminine traits (or self-perceptions of traits)? Or do sex role attribute scales merely reflect what respondents think are the socially acceptable, stereotypic attributes for their sex? Spence, Helmreich, and Stapp (1975) argued against the stereotype interpretation of their instrument, the PAQ. They reasoned that the stereotype interpretation would predict a correlation between peoples' perceptions of stereotypes and their PAQ scores. Finding no such correlation in their data, Spence, Helmreich, and Stapp dismissed the stereotype explanation.

The present study included important methodological improvements on Spence, Helmreich, and Stapp's (1975) design and found a significant relationship between sex role stereotypes and sex role attributes. As discussed above, these results support Kohlberg's model of sex role identity. Do these results also present a serious challenge to Spence, Helmreich, and Stapp's trait interpretation of the PAQ? We would argue that these results do not, necessarily, discount the PAQ. In the first place, the absolute size of the correlations between attributes and stereotypes in this study was small. A great deal of variance remained in PAQ attribute scores that could not be accounted for by stereotype scores. Second, a correlation between sex role attributes and stereotypes is not really inconsistent with a trait interpretation of the PAQ. According to Kohlberg's model, stereotypes *should* correlate with attributes precisely because those stereotypes influence the early development of sex role traits. In short, Spence, Helmreich and Stapp's trait interpretation of the PAQ is not damaged by the present results, and their concern about a correlation between the PAQ and sex role stereotypes may be unfounded in light of Kohlberg's theory.

We are cautious, however, about certain

surplus meanings that may accompany the trait view of sex role attributes. At issue is the notion that traits are immutable and intransient aspects of the personality, to be assessed but never changed. An important implication of Kohlberg's theory is that people can and will change, modify, and develop new sex role attributes as their stereotypes change. This notion suggests a line of future research in which subjects' perceptions of stereotypes are manipulated and instruments like the PAQ are used as dependent variables. Such research is not likely to be undertaken, however, if investigators view sex role attributes as fixed and immalleable *personality* traits.

The trait validity of Bem's (1974) sex role attribute scale, the Bem Sex Role Inventory (BSRI), has recently faced more serious challenges than Spence and Helmreich's PAQ. Pedhazur and Tetenbaum (1979) subjected the BSRI to factor analysis and discovered that the two items on the instrument—the items Masculine and Feminine—accounted for nearly 80% of the variance. Furthermore, despite Bem's contention that masculinity and femininity are independent trait dimensions, females who rated themselves high on the Feminine item tended to rate themselves low on the Masculine item, and vice versa for males.

Given the high loadings on the Masculine and Feminine items in the BSRI and the inverse relationship between those items, it seems possible to us that the BSRI actually measures sex role identity rather than sex role attributes. This possibility is given further credence by the results of one of Bem's own studies. Bem and Lenney (1976) gave subjects an opportunity to engage in a variety of sex-typed masculine behaviors (oiling the squeaky lid of a tool box, baiting a fishing hook, etc.) or feminine behaviors (ironing napkins, filling a baby bottle, etc.). They found that masculine and feminine subjects (as determined by BSRI scores) chose sex role appropriate tasks and avoided opposite-sex-typed tasks, even if offered more money to perform the latter. Androgynous subjects, in contrast, freely engaged in whichever sex-typed behavior paid more.

Although Bem and Lenney predicted the above result, it actually provides negative evidence regarding the trait validity of the BSRI. All of the masculine and feminine sex-typed behaviors used in Bem and Lenney's study were in fact highly agentic/instrumental tasks. Baiting a fishhook and ironing a napkin obviously differ in their sex role appropriateness, but they both require agentic manipulation of the physical environment. Thus, in terms of the *traits* that the BSRI supposedly reflects, masculine subjects should have freely engaged in any of the experimental tasks, whereas feminine subjects should have actively avoided all the experimental tasks. If the BSRI measures sex role identity, however, Bem and Lenney's results make perfect sense. Kohlberg's theory postulates that sex role identity guides the individual's modeling of sex role stereotypes. Thus, if the BSRI in fact classifies people by sex role identity, Bem and Lenney's subjects chose exactly the stereotypic tasks one might expect.[3]

The Nature of Sex Role Stereotypes

There has been recent controversy over the apparently tenuous nature of sex role stereotypes (Locksley & Colten, 1979; Pedhazur & Tetenbaum, 1979). Stereotypes of men and women vary markedly as a function of measurement techniques (Ehrlich & Rinehart, 1965) and the inclusion or exclusion of role-specific information about the target (Clifton, McGrath, & Wick, 1976). Some writers have argued that research should focus on sex role schemata rather than on sex role stereotypes (Locksley & Colten, 1979; Spence & Helmreich, 1979). Schemata are more personalistic, situation-specific expectations held by individuals about their own and others' sex role attributes and behaviors, and thus may have more predictive validity than abstract stereotypes.

---

[3] Helmreich, Spence, and Holahan (1979) recently attempted to replicate Bem and Lenney's (1976) results using the PAQ instead of the BSRI. The PAQ failed to predict subjects' choice of sex role behaviors. Ironically, that failure to replicate can be interpreted as positive evidence that the PAQ measures sex role attributes and not sex role identity.

It is interesting to note that Kohlberg frequently used the term *schema* in place of the term *stereotype* in his 1966 theoretical paper. According to him, individuals are exposed to both abstract stereotypes and specific schemata from peers, parents, and other adults. The relevance of sex role information to the individual, or the extent to which the individual models after the stereotypes and schemata provided by others, is guided by his or her sex role identity. From that interaction each individual builds a catalog of very specific personal schemata, intentions, preferences, and behavioral propensities. In short, Kohlberg's theory, with its emphasis on the development of specific, individualistic sex role schemata, presaged many of the current trends in sex role research and could provide an integrative model for future research.

References

Bem, S. L. The measure of psychological androgyny. *Journal of Clinical and Consulting Psychology,* 1974, *42,* 155–162.

Bem, S. L., & Lenney, E. Sex typing and the avoidance of cross-sex behavior. *Journal of Personality and Social Psychology,* 1976, *33,* 48–54.

Clifton, A. K., McGrath, D., & Wick, B. Stereotypes of women: A single category? *Sex Roles,* 1976, *2,* 135–148.

Constantinople, A. Masculinity–femininity: An exception to a famous dictum? *Psychological Bulletin,* 1973, *5,* 389–407.

Ehrlich, H. J., & Rinehart, J. W. A brief report on the methodology of stereotype research. *Social Forces,* 1965, *43,* 564–575.

Heilbrun, A. B. Measurement of masculine and feminine sex role identities as independent dimensions. *Journal of Consulting and Clinical Psychology,* 1976, *44,* 183–190.

Helmreich, R. L., Spence, J. T., & Holahan, C. K. Psychological androgyny and sex role flexibility: A test of two hypotheses. *Journal of Personality and Social Psychology,* 1979, *37,* 1631–1644.

Kagan, J. Acquisition and significance of sex-typing and sex-role identity. In M. L. Hoffman, & L. W. Hoffman (Eds.), *Review of child development research* (Vol. 1). New York: Russell Sage, 1964.

Kohlberg, L. A. A cognitive-development analysis of children's sex-role concepts and attitudes. In E. E. Maccoby (Ed.), *The development of sex differences.* Stanford, Calif.: Stanford University Press, 1966.

Locksley, A., & Colten, M. E. Psychological androgyny: A case of mistaken identity? *Journal of Personality and Social Psychology,* 1979, *37,* 1017–1031.

Pedhazur, E. J., & Tetenbaum, T. J. Bem Sex Role Inventory: A theoretical and methodological critique. *Journal of Personality and Social Psychology,* 1979, *37,* 996–1016.

Rosenkrantz, P. S., Vogel, S. R., Bee, H., Broverman, I. K., & Broverman, D. M. Sex role stereotypes and self-concepts in college students. *Journal of Consulting and Clinical Psychology,* 1968, *32,* 287–295.

Spence, J. T., & Helmreich, R. *Masculinity and femininity: Their psychological dimensions, correlates, and antecedents.* Austin: University of Texas Press, 1978.

Spence, J. T., & Helmreich, R. L. The many faces of androgyny: A reply to Locksley and Colten. *Journal of Personality and Social Psychology,* 1979, *37,* 1032–1046.

Spence, J. T., Helmreich, R., & Stapp, J. Ratings of self and peers on sex role attributes and their relation to self-esteem and concepts of masculinity and femininity. *Journal of Personality and Social Psychology,* 1975, *32,* 29–39.

Stein, A. H. The effects of sex role standards for achievement and sex role preference on three determinants of achievement motivation. *Developmental Psychology,* 1971, *4,* 219–231.

Received November 14, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 10, 1790-1797

Acknowledgment of Handicap as a Tactic in Social Interaction

Albert H. Hastorf, Jeffrey Wildfogel, and Ted Cassman
Stanford University

Nonhandicapped people often report discomfort and uncertainty when interacting with handicapped individuals. The three studies reported here investigated a possible tactic that handicapped people could use to reduce a fellow interactant's discomfort and uncertainty. Nonhandicapped subjects watched two videotapes of handicapped individuals being interviewed. Each subject then chose the handicapped person with whom he would prefer to work on a cooperative task. Results of all three studies supported the hypothesis that a handicapped person acknowledging his handicap will be preferred to a handicapped person who does not acknowledge his handicap. In Study 1, subjects significantly preferred a handicapped person who acknowledged his handicap to a handicapped person who did not disclose anything personal. In Study 2, subjects significantly preferred an acknowledging person over one who made a personal disclosure other than about his handicap. In Study 3, subjects preferred the individual acknowledging a handicap over one who disclosed something else personal even when the acknowledging individual was clearly nervous about doing so. These results suggest that acknowledging the handicap may be a promising tactic.

Experimental evidence has shown that physical handicaps can negatively affect personal encounters. Nonhandicapped individuals often report discomfort and uncertainty when interacting with handicapped individuals (Davis, 1961; Kleck, Ono, & Hastorf, 1966), and these subjective reports have been corroborated by an objective measure of emotional arousal (Kleck et al., 1966). Further, when interacting with the handicapped, nonhandicapped individuals exhibit less variability in their behavior, express opinions less representative of their actual beliefs, gesture less, and even end the interaction sooner than they do when interacting with a nonhandicapped individual (Kleck, 1968; Kleck et al., 1966). Finally, it has been shown that this tension and discomfort is felt by the handicapped individual on the "other side" of the interaction as well (Comer & Piliavin, 1972).

There are several plausible explanations for the discomforting effect physically handicapped individuals have upon the nonhandicapped. First, nonhandicapped individuals may feel discomfort because the mere presence of a handicapped person forces upon them the realization that they, too, are vulnerable to similar disabilities (Novak & Lerner, 1968). There may also be a fear that the stigma accorded to handicapped individuals is contagious, that being seen with the stigmatized is "discrediting" by association (Goffman, 1963).

Yet another source of the discomforting effect handicapped individuals have on the nonhandicapped may stem from the nonhandicapped individual's uncertainty as to what kind of behavior is expected and appropriate. There are strong societal norms to treat the handicapped kindly and carefully, but there are equally strong norms to treat

This study was partially supported by the Boys Town Center at Stanford University and by National Institutes of Health Grant MH29446-02.

The authors would like to thank William DeJong and Robert Kleck for the helpful comments on an earlier draft of this manuscript. Gregory Northcraft contributed substantially to the final form of the manuscript.

Requests for reprints should be sent to Albert H. Hastorf, The Boys Town Center, Stanford University, Stanford, California 94305.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3710-1790$00.75

them just like anyone else (Kleck et al., 1966) so as not to appear condescending. Similarly, nonhandicapped individuals may find themselves in conflict over a desire to explore the handicap because it constitutes a novel stimulus and a duty to adhere to strong norms against staring at another person (Langer, Fiske, Taylor, & Chanowitz, 1976). Finally, feelings of discomfort may arise in the presence of a handicapped individual because a deformity is unsightly. Closely related to this is the notion that discomfort arises because the deformity violates our expectations of what a "whole" person should look like (Richardson, 1976).

Whatever the cause of the discomfort felt by nonhandicapped individuals in the presence of the handicapped, it is clear that it makes the physically handicapped the socially handicapped as well. Since the handicapped do not receive accurate feedback concerning the appropriateness of their own behavior or experience the normal behaviors of others, (Hastorf, Northcraft, & Picciotto, 1979) it is likely that their learning of rules about social interactions and their development of sensitivity to others is impaired. In addition, Kleck (1969) has found that although the norms to be kind to the handicapped may result in positive impressions of them in initial encounters, this positive first impression usually attenuates with further interaction. This finding is hardly surprising in view of the impaired social skills development of the handicapped mentioned above. Thus, because of the way the nonhandicapped react to a physical handicap, the handicapped individual may be led to the attribution: "I am the type of person who causes others to feel uncomfortable, and am therefore avoided. When people *do* get to know me, they appear to like me less the better they get to know me." The negative implications for self-concept are not difficult to imagine.

Thus, where the handicapped can be successful in overcoming physical and legal barriers, psychosocial barriers confronting them may prevent their entry into the mainstream of society. The present series of investigations sought to determine if there are tactics the handicapped might employ to overcome these psychosocial barriers.

The discomfort of the nonhandicapped in interactions with the handicapped seems to be a major cause of the psychosocial barriers faced by the handicapped. It therefore seems reasonable that any tactic that reduces this discomfort would facilitate interactions between the handicapped and nonhandicapped. One such tactic may be acknowledging the handicap. Just as one may find it disarming when a relatively short individual introduces himself with the phrase "You may have noticed I'm a little tall for a leprechaun," so might it serve to reduce interactional tension when a handicapped individual conversationally acknowledges his or her handicap. This would seem especially true if the acknowledgment serves the purpose of signaling that the handicap is *not* an interactional obstacle, mention of which or involvement with which must be carefully circumnavigated.

The only study that has investigated the tactic of acknowledging the handicap found that it had no impact on the behavior of nonhandicapped individuals or their ratings of handicapped people (Farina, Sherman, & Allen, 1968). There are, however, circumstances that cast doubts on the validity of this conclusion. The study used male subjects, and its findings are therefore equally consistent with the hypothesis of no relationship between disclosure and liking for nonhandicapped males. Furthermore, the handicapped acknowledgment in the Farina et al. study may not have been effective because it did not reduce the significance of the handicap as a topic of suppressed concern. For acknowledgment of the handicap to be an effective tactic, it may be necessary for the handicapped person to mention it in a manner that conveys that he or she is not overly sensitive about it, that it is all right if the handicap comes up in conversation. The confederate in the Farina et al. study merely mentioned that his handicap caused inconveniences and was the result of a car accident. Such a statement, rather than implying that the handicap is an acceptable topic of conversation, may indicate that the topic has been opened *and closed.*

The present series of investigations explored the tactic of acknowledging the handicap in interaction between handicapped and nonhandicapped individuals. Three studies

were conducted to test the hypothesis that nonhandicapped people would prefer to interact with handicapped individuals who acknowledged their handicap in a manner that showed that they were not overly sensitive about it. In each study, nonhandicapped subjects watched two videotapes, each of a handicapped person in a wheelchair being interviewed. All subjects were told they would interact later with one of the handicapped persons they had viewed. After each interview, subjects gave their impressions of the person they had just seen, and after seeing both interviews, subjects chose the handicapped person with whom they wanted to interact.

Experiment 1

This first study investigated whether nonhandicapped subjects prefer a handicapped confederate who acknowledges his handicap over another who does not. It was hypothesized that nonhandicapped subjects would feel more at ease with the acknowledging confederate and would therefore have more favorable impressions of him and choose to interact with him.

Method

Subjects. Fifty-three male and female students at Stanford University served as subjects. Most were first-year students enrolled in the introductory psychology course. Subjects in the course participated for course credit; others participated for money. Only the data of 48 subjects were included in the analysis. Two subjects declined to state a preference between confederates, 2 subjects suspected that the interviews were staged, and 1 subject had met one of the confederates beforehand. The data of these 5 subjects were omitted from the sample.

Procedure. Subjects were told that the experiment was investigating the effects of seeing and hearing a paraplegic individual (as compared to only reading about him) on future interactions with him. Subjects were informed that the two paraplegics working on the project were extremely busy, and that their schedules were unpredictable. For these reasons, it was explained, they would be shown videotapes of the two paraplegics during this first session and would then return at a later date to perform a simple task with one of them. This cover story provided an opportunity to ask subjects their impressions of the confederates and preferences for a partner without arousing suspicion.

Two interviews of approximately 3½ minutes duration each were taped for four apparent paraplegic confederates. The four tapes of two confederates (Mathew and Peter) were selected for use in the experiment because pretests showed that they were most closely matched on attractiveness and were liked substantially more than the other two confederates. One of the confederates (Mathew) was not actually a paraplegic. His tapes were nevertheless used because pretests confirmed that subjects could not detect which of the four confederates was not a true paraplegic.

The interviews were carefully rehearsed to differ only in whether the interviewee acknowledged his handicap. In all interviews, the confederate told of his happy childhood, good relationship with his parents, part-time work, and plans to take a year off from school. Acknowledgment and nonacknowledgment interviews differed in how the confederate responded to the question, "How do you feel about yourself and your college experience up to this point?" In the acknowledgment interview, the confederate spoke of the problems of being in a wheelchair but said that he had learned to accept the inconveniences. He mentioned that he realized people were afraid to talk about his handicap, but he encouraged them to ask questions, anyway; only in this way could his handicap be gotten out of the way so that people could really get to know him. In the nonacknowledgment interview, the confederate responded that college life was agreeable, that school work kept him rather busy, and that he had recently begun to play clarinet in the school orchestra. Effects of confederate viewed, sex of subject, and order of presentation of the confederates and interview conditions were counterbalanced.

After watching the first interview, subjects rated the handicapped interviewee on an impression scale and answered four questions about him. The impression scale consisted of nine polar adjective pairs separated by a 7-point scale. The adjective pairs were pleasant–unpleasant, positive–negative, excitable–calm, active–passive, strong–weak, tough–fragile, likable–unlikable, well adjusted–poorly adjusted, and hardworking–lazy. Two of the four questions asked how the subject thought the handicapped interviewee would act in certain situations, and two asked how the subject would act in certain situations with the handicapped interviewee. After the second interview, subjects rated the second confederate in the same manner.

Following the second rating, subjects also chose the handicapped person with whom they would prefer to work in the second session. Subjects were told that an effort would be made to give them the partner they chose. Although the experimental situation of choosing between two handicapped people may at first glance seem a strange one, subjects did not seem to consider the situation odd. Now that handicapped people are more in the news, this type of situation is apparently viewed as quite plausible. When the second questionnaire was completed, the experimenter asked each subject to explain the rea-

sons for his or her choice. The experimenter then explained the true purpose of the experiment and answered any questions the subject had.

Results and Discussion

Significantly more subjects (71%) chose to work with the confederate who acknowledged his handicap than with the nonacknowledging confederate, $\chi^2(1) = 7.52$, $p < .01$. Analyses of the effects of confederate viewed, sex of subject, and order of presentation of the confederates and interview conditions did not reveal any significant differences in which confederate was preferred. Furthermore, 79% of the subjects who chose the acknowledging confederate reported that the acknowledgment was an important factor in their decision. Clearly the acknowledgment exerted a major influence on working partner preference.[1]

Sign tests revealed that of the four questions asked each subject about the confederates, only the question "Do you think you would feel uncomfortable during a conversation with the interviewee and your friends, if someone began talking about a touch football game?" successfully distinguished between acknowledging and nonacknowledging confederates, $p < .05$. This suggests that subjects chose to work with the acknowledging confederate because they anticipated feeling more comfortable in his presence.

Analysis of the polar adjectives revealed that the acknowledging confederate was rated more favorably on the evaluative and potency factors and was perceived as more likable and better adjusted, smallest $t(47) = 2.02$, $p < .05$. Subject ratings of the handicapped confederates thus also support the hypothesis that acknowledging the handicap is an effective tactic for increasing working-partner preference. Moreover, the subjects' ratings suggest that the acknowledging tactic may be successful because it introduces the handicapped person as someone who is not overly sensitive about his handicap.

Experiment 2

The results of Experiment 1 suggest that acknowledging the handicap is an effective tactic in handicapped/nonhandicapped interaction because acknowledgment reduces discomfort of the nonhandicapped interactant. However, the findings of the first study do not rule out the possibility that the effectiveness of acknowledgment arises because the acknowledgment reveals personal information, thereby increasing intimacy and liking (Jourard, 1959; Worthy, Gary, & Kahn, 1969). A second study therefore compared the preferences of nonhandicapped subjects who had watched a handicapped confederate who acknowledged his handicap and another who disclosed something else personal but did not mention his handicap. Since it was hypothesized that it is acknowledgment of the handicap that helps nonhandicapped individuals feel less uncomfortable and influences their preferences, it was expected that subjects would still have more favorable impressions of, and choose to interact with, the confederate who acknowledged his handicap.

Method

Subjects. Fifty-five male and female students from the same subject population as in Experiment 1 participated in the second study. The data of 48 subjects were included in the analysis. Three subjects were suspicious that the interviews were staged, and 4 subjects had met or heard about one of the confederates. The data of these 7 subjects were omitted from the analysis.

Procedure. The procedure was identical to that of Experiment 1, except that the two interviews that subjects watched featured a confederate who acknowledged his handicap and one who made a personal disclosure unrelated to his handicap. This personal disclosure interview replaced the nonacknowledgment interview of the first experiment.

In the personal disclosure interview, the confederate stated that school was going well and that he had several good friends but that recently he and his girlfriend of 8 months had been having problems. He then described their uncertainty about next year and the "weird religious trip" his girlfriend's mother was "laying on her." He finished the interview by saying that he hoped it all worked out. Although there are other types of disclosures that might have

[1] A similar finding has since been reported. An unpublished paper by Bazakas (Bazakas, Note 1) reports that acknowledgment of the handicap results in more favorable reactions toward a handicapped confederate only when he presents himself as a "coping" person.

been used, this type of disclosure had the advantage of being very revealing. Whether other types of disclosures might have different effects on preferences for the handicapped is a question that merits further exploration.

The dependent measures were essentially the same as those used in Experiment 1. One of the adjective pairs for the activity factor (excitable–calm) was replaced by fast–slow, and three more polar adjective pairs were added (warm–cold, considerate–inconsiderate, and friendly–unfriendly). *Fast–slow* replaced *excitable–calm* because the two adjective pairs comprising the activity factor were insignificantly correlated in the first experiment ($r = .04$).

Results and Discussion

As predicted, subjects preferred the confederate who acknowledged his handicap over the one who made a personal disclosure unrelated to his handicap, $\chi^2(1) = 13.02$, $p < .001$. Fully 77% of subjects chose to work with the confederate acknowledging a handicap.

As in Experiment 1 there were no preference effects due to confederate viewed, sex of subject, or order of presentation of the confederates and interview conditions.

As a manipulation check of personal disclosure, the personal disclosure interviews from Experiment 2 and the nondisclosure interviews from Experiment 1 were shown to 32 additional subjects who rated the interviews on 17 polar adjective pairs. Two of these adjective pairs were combined into a "disclosing" measure. The scales were open–closed and secretive–revealing. A t test of these ratings revealed that there was a tendency to see the personal disclosure confederate as more disclosing, $t(31) = 1.67$, $p < .11$. Although the personal disclosure interview may have differed from the nondisclosure interview on other dimensions, this notion cannot be properly tested. There is no way to distinguish whether other differences would be due to the disclosure itself or to the specific type of disclosure used.

As in Experiment 1, belief in the importance of the handicap acknowledgment in determining subjects' preferences for partner is substantiated by the reports of subjects who chose to work with the confederate who acknowledged his handicap. Eighty-two percent of the subjects choosing the acknowledging confederate stated that handicap acknowledgment was an important factor in the decision.

A sign test on the question "Would you be more comfortable" found that subjects thought they would be significantly more comfortable with the acknowledging confederate, $p < .05$. Again, this suggests that choice of partner may be related to how comfortable acknowledgment of the handicap makes the nonhandicapped individual feel. Analyses of the polar adjectives revealed that the confederate acknowledging the handicap was rated more favorably on the evaluative factor and was perceived as both better adjusted and more friendly, smallest $t(47) = 2.19$, $p < .05$.

The results of Experiment 2 again support the hypothesis that acknowledgment of the handicap (and not just the volunteering of something personal) is an effective tactic in increasing partner preference. Further, the findings of this second study are consistent with the hypothesis that the acknowledgment tactic is effective because it helps the handicapped individual appear less sensitive about his handicap, thereby making those who interact with him more comfortable.

Experiment 3

It may be difficult, at least for some handicapped individuals, to encourage the nonhandicapped to ask questions about their handicap. Handicapped individuals are, in fact, often sensitive about their handicaps and will probably not be able to use the tactic of acknowleging their handicap without showing nervousness and apprehension. Before recommending a tactic, it is important to know something of its limits. This third study explored the consequences for a handicapped individual of showing signs of tension and anxiety while acknowledging the handicap.

It was hypothesized that nonverbally communicated nervousness would cast a different interpretation upon the confederate's acknowledgment of the handicap. Specifically, it was hypothesized that this nervousness would lead the subject to feel that it was *not* acceptable to talk about the handicap. The

characteristic response to discrepant information is to rely more on the nonverbal than the verbal information (Mehrabian, 1971). Consequently, subjects might feel even more uncomfortable around a handicapped person who (nonverbally) showed discomfort about being handicapped while verbally insisting otherwise. It was therefore predicted that the findings of the first two experiments would be reversed: that subjects would have more favorable impressions of, and choose to interact with, a confederate who disclosed something personal but neither mentioned the handicap nor showed signs of nervousness while being personal.

Method

Subjects. Fifty-two male and female students from the same subject population as in Experiments 1 and 2 participated in the study. The data of 48 subjects were included in the analysis. Three subjects had met or heard about one of the confederates, and 1 subject declined to state a preference between confederates. The data from these 4 subjects were omitted from the sample.

Procedure. The procedure was virtually identical to that of Experiment 2, except that a "nervous" acknowledgment of handicap replaced the interview acknowledging the handicap of the second experiment. The verbal text of the interviews remained unchanged. However, during most of the handicap acknowledgment in the nervous condition, the confederate attempted to display tension by avoiding eye contact, running his hand through his hair, and clasping his hands tightly together. The other interview, as in Experiment 2, was of a confederate who made a personal disclosure. In this interview, the confederate neither mentioned his handicap nor showed signs of nervousness. The dependent measures were the same as in Experiment 2.

Results and Discussion

Contrary to predictions, subjects tended to prefer the confederate who acknowledged his handicap, $\chi^2(1) = 3.52$, $p < .07$, even when he was nervous about doing so. Sixty-five percent of subjects chose to work with this confederate.

An analysis of the effects of confederate viewed, sex of subject, and order of presentation of the confederates and interview conditions revealed that the order of presentation of the interview condition significantly affected confederate preferences, $\chi^2(1) = 5.83$, $p < .02$. That is, the confederate who was nervous while acknowledging his handicap was preferred significantly more often when he was seen first than when he was seen second. Since order of presentation of the interview condition did not affect confederate preferences in the first two studies, this finding is difficult to explain. Nevertheless, it remains striking that although the confederate who acknowledged his handicap was clearly nervous about doing so, he was still preferred.

As a manipulation check on perceived nervousness, the nervous acknowledment of handicap and personal disclosure interviews were shown to 32 additional subjects who rated each interviewee on 17 polar adjectives. Three of the adjective pairs were combined into a "nervousness" measure. The scales were at ease–nervous, relaxed–tense, and comfortable–uncomfortable. A t test of these ratings revealed that the nervous handicap acknowledgment interview was seen as significantly more nervous than the personal disclosure interview, $t(31) = 2.32$, $p < .05$. There were no differences in perceived nervousness of the confederate who acknowledged a handicap as a function of whether he was seen first or second.

Even though the confederate was perceived as nervous while acknowledging his handicap, 74% of subjects choosing the acknowledging confederate identified the acknowledgment of handicap as an important factor in their decision, and 19% of subjects who chose the nervous confederate did so despite acknowledging that the confederate clearly was not comfortable with his own handicap. Apparently, then, some subjects may have chosen the nervous confederate because he was *trying* to cope with his handicap.

Subjects realized that they would be just as uncomfortable around the confederate who acknowledged a handicap as around the confederate who only made a personal disclosure; more than half of the subjects reported that they would be uncomfortable around the confederate who acknowledged his handicap. Finally, nervousness while acknowledging did have a negative effect, as 76% of subjects who rejected the nervous, acknowledging confederate stated that they

rejected him because he did not seem adjusted to his handicap.

The nervous confederate was perceived to be more passive, $t(47) = 2.12$, $p < .05$. Comparison of no other polar adjective pair ratings attained significance. However, the variance of the ratings on 10 of the 12 adjective pairs was larger when subjects were rating the nervous, acknowledging confederate. A check on the variance of subject ratings in the first two experiments revealed that this pattern of variance in the third experiment was a significant reversal of the pattern in the first two studies, $\chi^2(2) = 38.6$, $p < .001$. One interpretation of these results is that in the first two experiments, subjects were more certain how to act with the confederate acknowledging a handicap because his communication in acknowledging made it clear what behavior was appropriate. Impressions of subjects were therefore relatively uniform. In contrast, the message of the acknowledgment of handicap in the third experiment was ambiguous. While verbally communicating an openness about his handicap, the nervous confederate was clearly betraying an anxiousness about the topic nonverbally. As a result, impressions of subjects were uncertain and thereby varied.

This interpretation is consistent with the ambivalence hypothesis suggested by Katz and his colleagues (Katz, Glass, & Cohen, 1973; Katz, Glass, Lucido, & Farber, 1977). They point out that ambivalence creates a tendency toward behavioral instability. As a result, any positive information amplifies favorable attitudes, and similarly, any negative information amplifies unfavorable attitudes. Our subjects, like most nonhandicapped individuals, probably have ambivalent (approach/avoidance conflicted) attitudes toward the handicapped. Acknowledgment of the handicap in the first two experiments was positive information because it reduced discomfort. Therefore, it amplified favorable impressions of the acknowledging confederates. Acknowledgment of the handicap in the third experiment contained both positive information (verbally) and negative information (nonverbally). Therefore, it amplified either favorable or unfavorable impressions, depending on which source of communication was more salient for each subject. This possibility of amplification of impressions in either direction accounts for the increased variance found in ratings of the acknowledging confederate in Experiment 3; the probable high salience of his nervousness accounts for the reduced effectiveness of the acknowledging tactic.

General Discussion

The findings of these three experiments have important implications for social policy; the tactic of acknowledging a handicap seems very promising. In all three studies, the confederate who acknowledged his handicap was preferred to the other handicapped confederate. This finding was true even if the other confederate made a personal disclosure.

In contrast to the results of investigations on personal disclosure (Ehrlich & Graeven, 1971; Jourard & Landsman, 1960), acknowledgment of the handicap was effective even though males were doing the acknowledging. Perhaps the macho stereotype of males, which views *disclosure* as weak and effeminate, is not a barrier to the acknowledging tactic.

It is noteworthy that the acknowledging confederate was still slightly more preferred even when he was clearly nervous about the acknowledgment. This suggests that handicapped people may use the tactic effectively even before they are well adjusted to their handicap. By acknowledging their handicap, handicapped individuals can reduce the discomfort and uncertainty of the nonhandicapped and thereby increase their opportunities for social interaction and its accompanying benefits. This increased opportunity for social interaction can then itself help promote smoother adjustment to the handicap.

Further research is needed, however, to explore the impact of the acknowledgment tactic. It would be useful to know, for instance, if the acknowledgment tactic attenuates the biased, inhibited behavior that the nonhandicapped usually demonstrate in the presence of handicapped individuals. The behavior of nonhandicapped individuals should be measured in interactions with handicapped individuals employing the acknowl-

edgment tactic. Studies of this kind will ultimately determine the utility of acknowledging the handicap.

Reference Note

1. Bazakas, R. *The interpersonal impact of coping, dependency, and denial: Self-presentations by the disabled.* Unpublished manuscript, Department of Educational Psychology, New York University, 1978.

References

Comer, R. J., & Piliavin, J. A. The effects of deviance upon face-to-face interaction: The other side. *Journal of Personality and Social Psychology,* 1972, *55,* 33–39.

Davis, F. Deviance disavowal: The management of strained interaction by the visibly handicapped. *Social Problems,* 1961, *9,* 120–132.

Ehrlich, H. J., and Graeven, D. B. Reciprocal self-disclosure in a dyad. *Journal of Experimental Social Psychology,* 1971, *7,* 389–400.

Farina, A., Sherman, M., & Allen, J. G. Role of physical abnormalities in interpersonal perception and behavior. *Journal of Abnormal Psychology,* 1968, *73,* 590–595.

Goffman, E. *Stigma: Notes on the management of spoiled identity.* Englewood Cliffs, N.J.: Prentice-Hall, 1963.

Hastorf, A. H., Northcraft, G. B., & Picciotto, S. R. Helping the handicapped: How realistic is the performance feedback? *Personality and Social Psychology Bulletin,* 1979, *5,* 373–376.

Jourard, S. M. Self-disclosure and other-cathexis. *Journal of Abnormal and Social Psychology,* 1959, *59,* 428–431.

Jourard, S. M., & Landsman, M. J. Cognition, cathexis, and the "dyadic effect" in men's self-disclosing behavior. *Merrill-Palmer Quarterly,* 1960, *6,* 178–186.

Katz, I., Glass, D. C., & Cohen, S. Ambivalence, guilt, and the scapegoating of minority group victims. *Journal of Experimental Social Psychology,* 1973, *9,* 423–436.

Katz, I., Glass, D. C., Lucido, D. J., & Farber, J. Ambivalence, guilt, and the denigration of a physically handicapped victim. *Journal of Personality,* 1977, *45,* 419–429.

Kleck, R. E. Physical stigma and nonverbal cues emitted in face-to-face interaction. *Human Relations,* 1968, *21,* 19–28.

Kleck, R. E. Physical stigma and task oriented interactions. *Human Relations,* 1969, *22,* 53–60.

Kleck, R. E., Ono, H., & Hastorf, A. H. The effects of physical deviance upon face-to-face interaction. *Human Relations,* 1966, *19,* 425–436.

Langer, E. J., Fiske, S., Taylor, S. E., Chanowitz, B. Stigma, staring, and discomfort: A novel-stimulus hypothesis. *Journal of Experimental Social Psychology,* 1976, *12,* 451–463.

Mehrabian, A. *Silent messages.* Belmont, Calif.: Wadsworth, 1971.

Novak, D. W., & Lerner, M. J. Rejection as a consequence of perceived similarity. *Journal of Personality and Social Psychology,* 1968, *9,* 147–152.

Richardson, S. A. Attitudes and behavior toward the physically handicapped. *Birth defects: Original articles series,* 1976, *4,* 15–34.

Worthy, M., Gary, A. L., & Kahn, G. M. Self-disclosure as an exchange process. *Journal of Personality and Social Psychology,* 1969, *13,* 59–63.

Received November 14, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 10, 1798-1809

Characterological Versus Behavioral Self-Blame: Inquiries Into Depression and Rape

Ronnie Janoff-Bulman
University of Massachusetts—Amherst

Two types of self-blame—behavioral and characterological—are distinguished. Behavioral self-blame is control related, involves attributions to a modifiable source (one's behavior), and is associated with a belief in the future avoidability of a negative outcome. Characterological self-blame is esteem related, involves attributions to a relatively nonmodifiable source (one's character), and is associated with a belief in personal deservingness for past negative outcomes. Two studies are reported that bear on this self-blame distinction. In the first study, it was found that depressed female college students engaged in more characterological self-blame than nondepressed female college students, whereas behavioral self-blame did not differ between the two groups; the depressed population was also characterized by greater attributions to chance and decreased beliefs in personal control. Characterological self-blame is proposed as a possible solution to the "paradox in depression." In a second study, rape crisis centers were surveyed. Behavioral self-blame, and not characterological self-blame, emerged as the most common response of rape victims to their victimization, suggesting the victim's desire to maintain a belief in control, particularly the belief in the future avoidability of rape. Implications of this self-blame distinction and potential directions for future research are discussed.

In a study by Bulman and Wortman (1977) on the relationship between blame attributions and coping, self-blame emerged as a predictor of good coping among paralyzed victims of freak accidents. A conclusion that is consistent with these results—that self-blame is a positive psychological mechanism—derives primarily from the implications of this attribution for a belief in personal control over one's outcomes. The advantages of perceived control have been repeatedly demonstrated in social psychological experiments (see, e.g., Bowers, 1968; Glass & Singer, 1972; Langer & Rodin, 1976; Schulz, 1976); Walster's (1966) formulation of observers' reactions to victims and Kelley's (1971) view of attributional processes "as a means of encouraging and maintaining [his] effective exercise of control in the world" (p. 22) are also based upon a recognition of the significance of perceived control. The tenuous link between control and self-blame becomes comprehensible as one realizes that in order to maximize a belief in control when attributing blame to particular factors, one's choice is influenced by the perceived modifiability of the potential factor(s). As Medea and Thompson (1974) write in the case of rape, "If the woman can believe that somehow she got herself into the situation, if she can make herself responsible for it, then she's established some sort of control over rape. It wasn't someone arbitrarily smashing into her life and wreaking havoc" (p. 105).

Unfortunately, this adaptive, control-oriented view of self-blame too easily ignores the more popular conception of the phenomenon, by which self-blame is regarded as maladaptive, a correlate of depression, and a reflection of psychological problems. Thus Beck (1967), writing about depressed pa-

The author thanks Philip Brickman, Irene Frieze, and Camille Wortman for their valuable comments on an earlier draft of this article.

Requests for reprints should be sent to Ronnie Janoff-Bulman, University of Massachusetts, Department of Psychology, Amherst, Massachusetts 01003.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3710-1798$00.75

tients, states, "Another symptom, self-blame, expresses the patient's notion of causality. He is prone to hold himself responsible for any difficulties or problems that he encounters" (p. 21). Self-blame as a maladaptive psychological mechanism is generally related to harsh self-criticism and low evaluations of one's worth.

Two Types of Self-Blame

Recognizing that self-blame may be both adaptive and maladaptive is a first step towards the conclusion that there are two different types of self-blame, one representing an adaptive, control-oriented response, the other a maladaptive, self-deprecating response. The primary distinction between these two self-attributions is the nature of the focus of blame, for it is proposed that the control related self-blame focuses on one's own behavior, whereas the esteem related self-blame focuses on one's character, an overall view of the kind of people individuals perceive themselves to be. In other words, individuals can blame themselves for having engaged in (or having failed to engage in) a particular activity, thereby attributing blame to past behaviors; or individuals can blame themselves for the kind of people they are, thereby faulting their character. To facilitate discussion of these two self-attributional strategies, the esteem related blame will be labeled "characterological" self-blame and the control related type, "behavioral" self-blame. In the case of rape, for example, a woman can blame herself for having walked down a street alone at night or for having let a particular man into her apartment (behavioral blame), or she can blame herself for being "too trusting and unable to say no" or a "careless person who is unable to stay out of trouble." This behavioral–characterological distinction parallels findings in the area of the just world theory. In their recent review, Lerner and Miller (1978) state that innocent victims who cannot be characterologically blamed (i.e., derogated) by virtue of their reputedly good character are instead blamed for some behavior in which they engaged (i.e., behavioral blame).

While this distinction between characterological and behavioral self-blame appears related to the state–trait distinction in clinical psychology (see, e.g., Spielberger, 1972), it more specifically corresponds to the distinctions drawn by Weiner and his colleagues (Weiner et al., 1971) in their scheme of attributions in the area of achievement. In attributing failure to oneself (internal attribution), one can point to his/her own lack of ability or effort, attributions that have very different implications for perceived control. Individuals who make an attribution to poor ability believe that there is little they can do to control the situation and succeed, for ability is stable and relatively unchangeable. Effort attributions, on the other hand, will lead one to believe that as long as he/she tries harder, he/she will be able to control outcomes in a positive manner (see Dweck, 1975). Similarly, characterological self-blame corresponds to an ability attribution, and behavioral self-blame corresponds to an effort attribution, having very different implications for perceived personal control. While the dimension used by Weiner and his colleagues to distinguish between ability and effort is that of stability (stable–unstable), the differences between the attributions may also be captured through the use of a controllability dimension (cf. Elig & Frieze, 1975; Weiner, 1974). The primary distinction to be drawn between behavioral and characterological self-blame is the perceived controllability (i.e., modifiability through one's own efforts) of the factor(s) blamed.

In a recent reformulation of learned helplessness, Abramson, Seligman, and Teasdale (1978) have posited a third dimension of attributions—global–specific—that is important to specify in determining subsequent perceived helplessness. While this global–specific dimension characterizes one of the differences to be noted between characterological and behavioral self-blame, it is proposed that the dimension of significance distinguishing these two types of self-blame is perceived controllability, and the global–specific and stable–unstable dimensions are important because of their contribution to perceived control. Abramson, Seligman, and Teasdale (1978), however, note that the dimension of "controllability is logically orthogonal to the

Internal × Global × Stable dimensions . . ." (p. 62). The position presented here is consistent with a comment by Wortman and Dintzer (1978) in their recent reaction to the learned helplessness reformulation. They state, "We feel that assessments of the controllability of the causal factor may be of the utmost importance in predicting the nature and magnitude of subsequent deficits" (p. 82).

In a discussion of self-blame by Abramson, Seligman, and Teasdale (1978) these authors state that self-blame in helplessness and depression follows from the "attribution of failure to factors that are controllable" (p. 62). The self-blame they are dealing with is that which is consistent with "self-esteem deficits" and "self-criticism" and thus parallels characterological self-blame. The authors do not recognize a second type of self-blame, behavioral self-blame, which according to the present analysis is the type of self-blame following from attributions to controllable factors. Contrary to the assertions of Abramson, Seligman, and Teasdale, it is proposed that characterological self-blame follows from attributions to uncontrollable factors.

A further distinction between behavioral and characterological self-blame lies in the time orientation of the attributor. It is proposed that in blaming one's behavior, an individual is concerned with the future, particularly the future avoidability of the negative outcome. This concern for future avoidability is consistent with the control-motivated basis for behavioral self-blame. The future-oriented concerns of behavioral self-blamers need not focus exclusively on the future avoidability of the negative outcome for which the attributor is blaming him/herself; rather, behavioral self-blame may promote a general belief in one's ability to avoid negative outcomes and to effect positive outcomes in the future. Thus, the paralyzed victims in the Bulman and Wortman (1977) study were apt to be better copers if they blamed themselves, but self-blame was more likely to be in the service of a general belief in future control (e.g., I'll be able to improve my physical condition through physical therapy), rather than a more specific belief in the future avoidability of their own paralysis, which was medically regarded as irreversible in all cases.

In blaming himself or herself characterologically, the individual is not concerned with control in the future, but rather with the past, particularly deservingness for past outcomes. Individuals who engage in behavioral self-blame are apt to have an eye towards the future and what they can do to avoid a recurrence of the negative outcome (or the occurrence of negative outcomes in general). Individuals who engage in characterological self-blame are apt to focus more on the past and what it was about them that rendered them deserving of the negative outcome for which they are blaming themselves.[1] Perceived avoidability and behavioral self-blame are thus assumed to be part of the same blame cluster, whereas characterological self-blame and feelings of deservingness are representative of another blame cluster.

Self-Blame and Depression: Toward the Resolution of a Paradox

Distinguishing between characterological and behavioral self-blame may be a first step toward resolving the "paradox in depression" recently recognized and discussed by Abramson and Sackeim (1977). According to these authors, there are two symptom clusters of depression, one represented by hopelessness, powerlessness, and futility, the other by self-blame, self-deprecation, and guilt. Abramson and Sackeim discuss two prominent theories of depression that are based on cognitions of hopelessness and self-blame. Seligman's (1975) theory of learned helplessness suggests that depression results from a belief in the uncontrollability of outcomes. According to Beck's (1967) theory of depression, the depressed individual blames him/herself for negative outcomes, particularly personal failures. It is the conjunction of these two models that accounts for the paradox. That

[1] These distinctions are consistent with a recent analysis of responsibility by Harvey and Rule (1978), in which causal responsibility and deservingness are regarded as conceptually distinct aspects of responsibility.

is, how can individuals blame themselves for outcomes over which they feel they have had no control? How can an individual feel both helpless and self-blaming? Abramson and Sackeim discuss several possible resolutions to this paradox but remain dissatisfied with the alternatives presented to date. However, a recognition of self-blame, not as a unitary phenomenon, but rather as a label for two very different self-attributional strategies, may inform and resolve the apparent paradox in depression.

One reason why a resolution to the depression paradox has not been forthcoming is suggested by Abramson, Seligman, and Teasdale's (1978) assertion (presented above) that self-blame follows from attributions to controllable factors. In assuming that self-blame naturally involves blaming controllable factors, the possibility that individuals can simultaenously feel they do not have control and blame themselves is foreclosed. Instead, if we recognize the distinction between behavioral and characterological self-blame, then the paradox begins to disappear. Essential to an understanding of this assertion is the proposition that in blaming himself or herself for the kind of person he/she is, the individual is not necessarily placing blame for an event regarded as personally controllable. A person can believe that he/she deserves what happened and is therefore "responsible" for it (see Harvey & Rule, 1978), without believing that he/she is capable of altering the outcome in the past, present, or future.

In the case of personal failures, the characterological blamers will point to deficits in themselves that are believed to account for these failures. The deficits are likely to lie in the realm of characteristics that generally define them, characteristics that are relatively nonmodifiable, stable, and global. Thus, in achievement tasks, an ability attribution would represent a characterological self-blaming strategy. In the case of self-blame for failures that are further removed from the individual, represented by the "delusions of depressives" who blame themselves for the "violence and suffering in the world" (see Beck, 1967), the individuals appear to regard themselves as being punished for who or what they are. In this case, rather than perceive themselves as responding, active organisms, depressed individuals seem to perceive themselves as passive stimuli. They do not believe they actively bring about outcomes that remain under their control. Rather, negative outcomes occur in reaction to them by other people and the world at large. In sum, self-blame by depressives is proposed as characterological in nature. Since characterological self-blame and feelings of helplessness are not logically inconsistent, their conjunction in depressed individuals should not be regarded as paradoxical.

Self-Blame Among Rape Victims

The association between self-blame and depression is probably well recognized and accepted within this culture. While the association between self-blame and rape is probably not as strong, the more or less popular image of the self-blaming rape victim may be more accurate than many would like to believe. The pervasiveness of self-blame has been well documented in literature on rape (see, e.g., Burgess & Holmstrom, 1974a, 1974b, 1976; Griffin, 1971; Hursch, 1977; Weis & Weis, 1975; Bryant & Cirel, Note 1). Although fear (of injury, death, and the rapist) is the primary reaction to rape, self-blame may be second only to fear in frequency of occurrence; perhaps surprisingly, it is far more common than anger.

In considering the few existing facts on victim precipitation in the crime of rape, however, it becomes obvious that the victims' self-attributional strategies (i.e., self-blame) do not reflect an accurate appraisal of the woman's causal role in the assault. The National Commission on the Causes and Prevention of Violence (1969) concluded that only 4.4% of all rapes are precipitated by the victim. Although a higher figure, 19%, has been proposed by Amir (1971), he used a considerably broader definition in establishing his criteria for victim precipitation. Thus, criteria such as "risky situations marred with sexualty" were used, affording the interpreter of data considerable discretion. In light of these percentages, the pervasiveness of self-blame becomes a puzzling phenomenon.

An attempt to account for the pervasiveness of such feelings has involved the proposition that women have been socialized to accept blame for their own victimization. As Brownmiller (1975) suggests, women are conditioned to a female victim mentality. Brownmiller discusses the psychologies of Deutsch and Horney and concludes that masochism is a female trait, one that has been socialized by men. Similarly, Burgess and Holmstrom (1974a) contend that women are socialized to the attitude of "blaming the victim," a perspective shared by Bryant and Cirel (Note 1). While there is no doubt much truth to this socialization hypothesis, it may paint a very incomplete picture of the factor(s) responsible for self-blame in women and the rape victim in particular. It fits nicely with a portrait of women as helpless and masochistic and may unwittingly perpetuate a view of women too consistent with the role of rape victim. In particular, this view entirely overlooks the possibility that self-blame by victims of rape may represent an adaptive response, an attempt to reestablish control following the trauma of rape.

A common reaction to rape is the feeling of a loss of control over one's life (Bard & Ellison, 1974; Bryant & Cirel, Note 1). The woman does not feel sure of herself and questions her self-determination. She needs to feel a sense of control (Hilberman, 1976), for she feels extremely vulnerable and particularly fears the rapist and a recurrence of rape. In blaming herself, perhaps the rape victim is engaging in a type of self-blame that maximizes a belief in control; that is, perhaps rape victims engage in behavioral self-blame rather than characterological self-blame. Whereas the latter type of blame would provide some support for a view of women as helpless and masochistic, the former would foster a different image of the rape victim and her reactions, that of an individual reacting in an adaptive manner to her recent loss of control.

If the rape victim engages in behavioral self-blame and attributes her victimization to a modifiable behavior (e.g., I should not have walked alone, I should have locked the windows), she is likely to maintain a belief in the future avoidability of a similar misfortune, while simultaneously maintaining a belief in personal control over important life outcomes. If, on the other hand, the rape victim blames herself characterologically, attributing the victimization to more or less unchangeable factors (e.g., I'm a weak person and can't say so, I'm the type of person who attracts rapists), she will presumably be considerably less likely to believe that she is capable of alleviating her vulnerability in the future and may begin to perceive herself as a chronic victim.

Two studies were conducted in order to test the usefulness of the distinction between behavioral and characterological self-blame in the areas of depression and rape. Study 1 was designed to determine whether characterological self-blame is a distinguishing characteristic of depressed individuals and whether it co-occurs with decreased beliefs in personal control among female college students. Study 2 involved surveying rape crisis centers across the country in order to determine which type of self-blame—behavioral or characterological—more accurately characterizes the reactions of rape victims served by these centers.

Study 1: Depression

Method

Subjects. Subjects were 129 undergraduate women at a large state university who were volunteers drawn from a number of undergraduate psychology courses. Each received one experimental credit for her participation. Responses from 9 of the subjects lacked much data, and these were eliminated from the analyses, leaving the responses of 120 subjects.

Procedure. Data [2] were collected during group sessions that generally ranged from 10 to 15 students. Subjects were told that we were interested in the relationship between personality variables and artistic taste, and that there would be three parts

[2] These data were collected by Laurie Gunsolley for her senior honors thesis, which was designed and completed under the direction of the author. While Gunsolley was particularly interested in self-esteem, the data have been reanalyzed for this presentation, using the responses to the Zung Self-Rating Depression Scale (1965) as the basis for distinguishing between the two groups of interest.

to the study: completing personality scales, viewing and rating a series of art slides, and reacting to several "real-life" types of situations. The subjects were asked to complete three personality scales. The Zung Self-Rating Depression Scale (Zung, 1965) was used to measure depression.[3] In addition, subjects completed the Janis–Field Feelings of Inadequacy Scale (Eagly, 1967), a self-esteem measure, and the Rotter Internal–External Locus of Control Scale (Rotter, 1966). Having completed these, the subjects were then asked to rate seven art slides on "aesthetic appeal." An overhead projector was used to show the slides, and ratings were made on 5-point scales. These artistic ratings not only provided a "cover" for the study but, more important, served as a distraction between the first (personality scales) and third (self-blame measures) parts of the experiment.

In the third phase of the study, subjects were asked to read four scenarios and to imagine that the various situations described had actually been experienced by them; that is, they were told to react to the scenarios given that they were the target people presented. In each situation the outcome was negative and the role of the target person was intentionally ambiguous. The scenarios involved the following situations: (a) a car driven by the target person is in an accident on a snowy winter day; (b) a social invitation by the target person is rejected (on the basis of false excuses) by an individual she recently met and regarded as a friend: (c) an urgent call for a roommate results in the target person's taking down the wrong number; the roommate is subsequently unable to return the call successfully; (d) an intense love relationship is ended when the target person's boyfriend leaves her and immediately gets involved with another woman.

Subjects were asked to respond to five questions following each scenario; responses were made on 6-point scales with endpoints *not at all* and *completely*. Subjects were asked to indicate how much they blamed themselves, other people, the environment (i.e., impersonal world), and chance, for the situations described. The question that tapped characterological self-blame asked, "Given what happened, how much do you blame yourself for the kind of person you are (e.g., the kind of person who is in an accident [Scenario A], the kind of person who has invitations turned down [Scenario B], the kind of person who causes inconveniences for others [Scenario C], the kind of person who is rejected in relationships [Scenario D])?" The appropriate "kind of person" was included separately for each scenario, so that for scenario D only "the kind of person who is rejected in relationships" was included. Question 3 sought to tap behavioral self-blame and asked, "Given what happened, how much do you blame yourself for what you did (e.g., your driving behavior [Scenario A], how you acted when you first met the person [Scenario B], how you acted when taking down the telephone number [Scenario C], how you acted with your boyfriend [Scenario D])?" Question 4 asked, "How much do you think you deserved what happened?" and Question 5 following each scenario was, "If the same situation arose in the future, to what extent do you believe that you could avoid what happened in this case?" All subjects were thoroughly debriefed upon completion of the session.

Results

Using a median split, subjects were divided into nondepressed (responses ranged from 6 to 21 on the Zung scale) and depressed (22 to 45 on the Zung scale) groups.[4] On the Janis–Field Feelings of Inadequacy Scale, the depressed group scored lower (i.e., had lower self-esteem) than the nondepressed group (64.05 vs. 74.97), $F(1, 118) = 36.72$, $p < .001$, and the depressed group was found to be more external than the nondepressed group on Rotter's Internal–External Locus of Control Scale (13.36 vs. 10.57), $F(1, 118) = 12.32$, $p < .001$.

Parallel attributional and self-blame measures were summed across the four scenarios; for example, a score for characterological blame was derived by adding the individual responses to each of the four questions (one following each scenario) that asked about characterological self-blame. In order to justify adding the four scales, alpha reliability coefficients were calculated for each of the eight summed scores. Despite the fact that each was composed of four scores, only the perceived avoidability measure failed to reach a reliability of .50. The general self and other people attributions were less than .60, and the other five measures had alpha reliabilities between .62 and .74.[5] Each

---

[3] In accordance with work on depression by Bonnie Strickland, a clinical psychologist in the Department of Psychology, University of Massachusetts, Amherst, a response category labeled "none of the time" was added to the Zung Self-Rating Depression Scale (1965). According to Strickland (personal communication), this renders the scale particularly sensitive to depression in a college population.

[4] Male and female college students completed the same depression scale in a study by Haley and Strickland (Note 2); their data had a median of 23.

[5] Nunnally (1967) writes that in early work on "hypothesized measures of a construct", reliabilities of .50 or .60 are sufficient standards (see p. 226).

summed score could range from a total of 0 to 24.

The depressed and nondepressed groups did not differ in the amount of blame they attributed to themselves in general, nor did they differ in the amount of behavioral self-blame reported, $F(1, 118) = 2.47$, *ns*. However, the groups did differ significantly in the amount of characterological self-blame reported, with more characterological self-blame reported by the depressed than the nondepressed group (11.59 vs. 10.03), $F(1, 118) = 4.33$, $p < .05$. Other differences on the total scores were found for attributions to chance; consistent with their greater externality on Rotter's scale, the depressed group blamed chance more than the nondepressed group (11.38 vs. 9.73), $F(1, 118) = 4.54$, $p < .05$. Further, there was a marginally significant difference between the two groups on the question of how much they felt they deserved what happened, with the depressed group reporting greater deservingness than the nondepressed group (8.31 vs. 7.24), $F(1, 118) = 3.68$, $p = .057$.[6]

A stepwise discriminant analysis was conducted in order to determine the variables that differentiated best between the two groups. All blame attribution measures were entered, with the exception of general self-blame, since characterological and behavioral self-blame were assumed to be finer distinctions of the general measure. Attributions to chance emerged as the best discriminator, $F(5, 114) = 4.54$, Wilks $\lambda = .963$, and characterological self-blame emerged as the next strongest differentiator, $F(5, 114) = 3.29$, Wilks $\lambda = .937$. These were followed, respectively, by attributions to other people, environment, and behavioral self-blame.

The correlations between deservingness, perceived avoidability, and the two types of self-blame were all strong. As an exploratory tool, an analysis of variance was conducted in order to further investigate the relationship between the variables. It should be noted that the low reliability of the avoidability measure calls for caution in interpreting the results of this analysis. Median splits were performed on the behavioral self-blame and characterological self-blame totals, and deservingness and avoidability totals were each analyzed by behavioral (high–low) and characterological (high–low) self-blame. A main effect for characterological self-blame emerged from the analysis of deservingness, with less deservingness reported by those who engaged in low characterological self-blame as compared with those who engaged in high characterological self-blame (6.20 vs. 9.26), $F(1, 119) = 29.78$, $p < .001$. On the other hand, a main effect for behavioral self-blame emerged from the analysis of perceived future avoidability, with less perceived future avoidability reported by those who engaged in low behavioral self-blame as compared with those who engaged in high behavioral self-blame (11.89 vs. 13.88), $F(1, 118) = 6.85$, $p < .01$.

Discussion

When self-blame was treated as a single entity (i.e., "self" as one of several possible factors tapped for blame attributions), no differences were found between the depressed and nondepressed students on this variable. However, when self-blame was divided into two types of self-attributions, behavioral and characterological, differences between the groups emerged. While the depressed and nondepressed students did not differ in terms of behavioral self-blame, they did differ significantly in terms of characterological self-blame; that is, the depressed students blamed themselves more characterologically than did the nondepressed students.

The suggestion that characterological self-blame follows from attributions to uncontrollable factors received strong support. Those who were depressed were more likely to attribute negative outcomes to chance, a variable that differentiated best between the depressed and nondepressed groups. Further, the depressed subjects were more external in locus of control orientation. Low self-esteem and somewhat increased feelings of deserving-

[6] When analyses were rerun using the conservative Scheffé procedure to correct for error rate inflation, significant differences between the depressed and nondepressed groups were again found on self-esteem, internal–external control, characterological self-blame, and chance.

ness characterized the depressed population, suggesting that characterological self-blame is esteem related, not control related.

The results would have been considerably more compelling if it were found that those students who were not depressed engaged in more behavioral self-blame than depressed students, yet this was not the case. However, it can perhaps be argued that the behavioral self-blame reported by the depressed and nondepressed populations differed in an important way; for the depressed group the behavioral self-blame co-occurred with characterological self-blame, and blaming one's behavior was thus an extension of blaming one's character. It may be difficult to blame one's character without blaming one's behavior, yet it may be very possible to blame one's behavior without blaming one's character. In the former instance the behavior may be regarded as uncontrollable in that it is a direct and unalterable extension of one's character (i.e., controlled by one's character). In the latter case the behavioral self-blame does not reflect decreased self-esteem, but rather the belief that one's behavior is modifiable. Perhaps behavioral self-blame, when displayed in conjunction with characterological self-blame, is simply a further reflection of characterological self-blame. However, when it occurs alone it is likely to represent an adaptive response, stemming from a desire to maintain a belief in personal control following a negative outcome.

Study 2: Rape[7]

Method

Respondents. Respondents were rape crisis centers located throughout the United States. Center names were derived primarily from a list located in a federal report on rape and its victims (Brodyaga et al., Note 3); this list was supplemented by names of rape crisis centers found in an informal directory at a local women's center. Services that were hotlines only or were task forces without counseling services were excluded from the final list. Questionnaires were mailed to 120 centers representing 37 states and the District of Columbia. Thirty of the questionnaires were returned "addressee unknown." Of the remaining 90 crisis centers, 48 responded (53% return rate; including those returned "addressee unknown," the return rate was 40%).

Questionnaire. In a cover letter I identified myself as a social psychologist interested in the nature of self-blame among victims of rape; letter recipients were asked to base their questionnaire responses on their experiences as counselors of rape victims. The questionnaire items dealt primarily with the issue of self-blame. Crisis centers were asked to indicate approximately how many rape victims they see yearly and of those they see, the percentage who blame themselves, at least in part, for the rape. The behavioral self-blame question asked, "Of the rape victims you see, what percentage blame themselves for the rape because of some behavior (act or omission) they engaged in at the time of or immediately prior to the rape (e.g., 'I should not have walked alone,' 'I should not have hitchhiked,' 'I should have locked my windows')?" The rape crisis centers were then asked to provide specific examples of behavioral self-blame related by the women they have counseled. The characterological self-blame question asked, "Of the rape victims you see, what percentage blame themselves for the rape because of some character trait or personality flaw they believe they have (e.g., 'I am so stupid, I deserved to be raped,' 'I'm the kind of woman who attracts rapists,' 'I am a weak person and can't say no')?" Specific examples of this type of blame were then requested as well. The centers were also asked to indicate on two 7-point scales, with endpoints *almost not at all* and *completely,* how much self-blaming characterized the women who engaged in behavioral and characterological self-blame, respectively; this was included in order to ascertain whether behavioral and characterological self-blamers differ in terms of the amount of self-blame they attribute to themselves for the rape.

Results

Of the 48 rape crisis centers that responded, 38 completed the questionnaire, 6 wrote letters providing general comments, and 4 wrote that they did not provide direct counseling services and were therefore unable to complete the items. Results were therefore based on the completed questionnaires of 38 centers. The rape crisis centers differed markedly in the scope of their operation, with the 3 smallest serving 12, 30, and 40 rape victims yearly, and the 3 largest serving 1,200, 1,250, and 1,500; the mean number of rape victims seen across the centers was 335.

In general, self-blame was reported as quite common; the reported mean percentage of

---

[7] The results of this study were reported by the author at the symposium "New Directions in Control Research" at the convention of the American Psychological Association, Toronto, 1978. The author thanks Chris Eagan for her invaluable help on the project.

women who blamed themselves at least in part for the rape was 74%. Of those who blamed themselves, behavioral self-blame was reported as considerably more common than characterological self-blame, and the differences between the reported incidence of the two blaming strategies was significant, $F(1, 32) = 140.90$, $p < .001$; an average of 69% of the women were reported as blaming themselves behaviorally, whereas an average of 19% were reported as blaming themselves characterologically. Further, examples of the two types of self-blame provided by the rape crisis centers confirmed the fact that they were readily able to distinguish between the two. Frequently mentioned examples of behavioral self-blame included the following: I shouldn't have let someone I didn't know into the house, I shouldn't have been out that late, I should not have walked alone (in that neighborhood), I should not have hitchhiked, I should not have gone to his apartment, I shouldn't have left my window open, I should have locked my car. Examples of characterological self-blame that were frequently reported included: I'm too trusting, I'm a weak person, I'm too naive and gullible, I'm the kind of person who attracts trouble, I'm not a very aware person, I'm not at all assertive—I can't say no, I'm immature and can't take care of myself, I'm not a good judge of character, I'm basically a bad person. It is perhaps worth noting that examples of behavioral blame were, without exception, reported in the past tense (i.e., I should have/should not have), whereas examples of characterological self-blame were presented in the present tense (I am/am not), perhaps implicitly indicating the presumed modifiability/nonmodifiability of factors associated with behavioral and characterological self-blame respectively (cf. Elig & Freize, 1975). However, the examples of the two types of self-blame provided in the questionnaire were consistent with the different tenses reported in the examples of the crisis centers, and this alone could have accounted for the findings. The author did not realize that she had made these distinctions on the questionnaire until the results clearly differentiated between the tenses used for the two types of self-blame. Finally, in responding to how much women blamed themselves for the rape, the centers reported that characterological self-blamers blamed themselves significantly more for the rape than did behavioral self-blamers (3.92 vs. 3.23), $F(1, 25) = 11.29$, $p < .002$.

Discussion

The rape crisis center counselors reported that the majority of rape victims do blame themselves, at least in part, following the rape. However, the focus of this self-attribution is a behavioral act or omission engaged in at the time of (or immediately preceding) the rape. Fewer than one-fifth of the women served by the centers blamed themselves in a characterological way, evidence that the "popular" view of the masochistic rape victim who perceives herself as worthless is largely unfounded. Rather, the self-blame in which most rape victims engage may represent a control maintenance strategy, a functional response to a traumatic event. Given the large discrepancies between those who blame themselves behaviorally and characterologically, it follows that most women clearly blame themselves in a behavioral manner only and do not combine this response with characterological self-blame, as may be the case with depressives (see above). In suggesting that behavioral self-blame reflects a positive impulse in rape victims, there is no intention of implying that the rape was the woman's fault. It is even likely that the woman who engages in behavioral self-blame does not do so to the exclusion of blaming the rapist, society, or other factors. These blame attributions, instead, would stem from different motivations, control maintenance being the motivation behind self-blame.

Two potentially serious objections to this study require a response. First, many of the women who volunteer or work in rape crisis centers may be ardent feminists who would be more likely to indicate that women blame themselves behaviorally rather than characterologically, for the latter suggests that women see themselves as worthless. In response, if the crisis center workers truly wanted to present women in a positive light,

they would have indicated quite simply that women infrequently blame themselves. Further, there was nothing in the questionnaire or cover letter to indicate that one type of self-blame was "healthier" than another, and several counselors commented that they had never before distinguished between types of self-blame but that it appeared interesting to them. Comments by the counselors indicated that these women were concerned about the health of the women they served and that preserving a positive image of womanhood in general was clearly not central to their activities as rape counselors. The second criticism that could be raised is potentially more serious. It is that women who go to rape crisis centers are most likely to be individuals who do not blame themselves characterologically and do not feel they deserved to be raped. Thus, there is a self-selected population of behavioral self-blamers served by rape crisis centers. It is difficult to counter this claim, for there is probably much truth to it. One must realize, however, that the literature written on rape is almost entirely derived from women who seek help after rape and not from women who quietly keep the trauma to themselves, ashamed to talk about it or admit it. The pervasiveness of self-blame documented in the rape literature is drawn primarily from observations of women at rape crisis centers, from women's centers, or from women who agree to be interviewed by researchers, also a population likely to be self-selected. Thus, the negative image of the rape victim engaging in masochistic, maladaptive self-blame derives from a rape victim population likely to be very similar to that served by the rape crisis centers surveyed. It might also be mentioned that those women who have least difficulty coping with the rape and who are apt to be behavioral self-blamers are probably also missing from the rape center population, for they may not require help (outside their own circle of family and friends) following the rape. Perhaps it is sufficient to point out that within the population of women served by rape crisis centers, self-blame has been improperly understood as self-derogating, reflecting the woman's belief in her own worthlessness, rather than as a response that reflects a positive attempt to reestablish personal control.

General Conclusions and Implications

Self-blame appears to be a label for two very different self-attributions, characterological self-blame being esteem related, and behavioral self-blame being control related. Self-blame as a predictor of good coping and self-blame as a concomitant of depression are no longer inconsistent in light of the two types of self-blame. Further, the paradox in depression—that individuals are simultaneously helpless and self-blaming—can be resolved if characterological self-blame characterizes depressives and differentiates them from nondepressed individuals. The division of self-blame into two different phenomena even has political or cultural implications, for self-blame by a victimized group such as rape victims can now be understood in such a manner as to preclude the perpetuation of a negative image of the group in question. It is perhaps unfortunate that one term has been used as a label for these two different self-attributions, for the singular term *self-blame* blurs important distinctions between adaptive and maladaptive responses to failures and victimizations. Since popularly the term has negative connotations, it would perhaps be desirable to provide a more neutral label for behavioral self-blame. Particularly in the case of rape, this would render more politically palatable the proposition that behavioral self-blame is of functional value for victims of rape.

The recognition of two types of self-blame may have therapeutic implications. Seligman's (1975) control-oriented strategies continue to seem appropriate for depressives, whose self-blaming does not imply high perceived control, but rather lack of control. Further, a cognitive therapy that entails reattributing the focus of one's attributions (e.g., from character to behavior) might be of value in treating depressives. In general, leading people to focus on behaviors that are alterable, rather than on their relatively nonmodifiable, more global character, may increase perceived future avoidability of negative events and

perceived control in general, outcomes that would presumably be of positive value. Dweck's (1975) successful reattribution training with helpless students, involving reattributing their ability attributions for failure to effort attributions, suggests the potential of such cognitive strategies using self-blame.

In the case of rape, the control concerns that may be implicit in the rape victim's self-blame often seem to be ignored in counseling, not because they are regarded as unimportant, but because they may go unrecognized. One counseling technique for rape victims includes repeatedly telling a woman that there is nothing she could have done to avoid the rape, that it was entirely the rapist's doing and outside of her control. Although meant to be reassuring, these statements could conceivably be not at all helpful, in light of the proposition that the women are seeking to reestablish a sense of control. Rather, counselors should perhaps recognize the functional value of behavioral self-blame and concentrate on enabling the victim to reestablish a belief in her relative control over life outcomes (e.g., discussing possible ways of minimizing the likelihood of a future rape). Too often, behavioral self-blame is regarded as detrimental to mental health. Rather, it may serve as an indicator of the victim's psychological needs at the time.

Behavioral and characterological self-blame appear to be distinct reactions yet are far from fully understood. Ideas raised in this paper have been tested only with female subjects and thus may not generalize to other populations; this issue of generalizability particularly calls for research with male subjects. Further, the relationship between the two types of self-blame would appear to be a fruitful area for future study. Does behavioral self-blame that occurs with characterological self-blame, for example, lose its adaptive value, or is it similar to behavioral blame that occurs without characterological self-blame? Is characterological self-blame that occurs without behavioral self-blame more or less maladaptive than characterological self-blame that occurs with behavioral self-blame? In addition, longitudinal studies designed to tap the coping implications of these two types of self-blame would be important contributions to our understanding of the relationship between coping and attributional strategies. Another possible direction lies in the area of blaming strategies by help-givers. Brickman and his colleagues (Brickman et al., Note 4) have presented a compelling case for the psychological tensions that exist between conditions that render helping appropriate (i.e., regarding the recipient of help as not responsible) and conditions that render helping effective (i.e., attributing responsibility to the recipient of help). That is, one is apt to help an individual who is not to blame for a misfortune, yet this attribution minimizes the belief that one's help will be effective. Perhaps training both help-givers and recipients of help to hold behavioral blame orientations (as opposed to characterological blame orientations) would help resolve the existing tensions. Last, the therapeutic implications of the two types of self-blame—behavioral and characterological—remain an area ripe for future study.

Reference Notes

1. Bryant, G., & Cirel, P. *A community response to rape: An exemplary project (Polk County Rape/Sexual Assault Care Center)*. Washington, D.C.: National Institute of Law Enforcement and Criminal Justice, 1977.
2. Haley, W. E., & Strickland, B. R. *Locus of control and depression*. Paper presented at the meeting of the Eastern Psychological Association, Boston, 1977.
3. Brodyaga, L., Gates, M., Singer, S., Tucker, M., & White, R. *Rape and its victims: A report for citizens, health facilities, and criminal justice agencies*. Washington, D.C.: National Institute of Law Enforcement and Criminal Justice, 1975.
4. Brickman, P., Rabinowitz, V. C., Coates, D., Cohn, E., Kidder, L., & Karuza, J. *Helping*. Unpublished manuscript, University of Michigan, 1979.

References

Abramson, L. Y., & Sackeim, H. A. A paradox in depression: Uncontrollability and self-blame. *Psychological Bulletin*, 1977, *84*, 838–851.

Abramson, L. Y., Seligman, M. E. P., & Teasdale, J. D. Learned helplessness in humans: Critique and reformulation. *Journal of Abnormal Psychology*, 1978, *87*, 49–74.

Amir, M. *Patterns in forcible rape*. Chicago, Ill.: University of Chicago Press, 1971.

Bard, N., & Ellison, K. Crisis intervention and investigation of forcible rape. *The Police Chief*, 1974, *41*, 68–73.

Beck, A. T. *Depression: Clinical, experimental, and theoretical aspects*. New York: Harper & Row, 1967.

Bowers, K. Pain, anxiety, and perceived control. *Journal of Consulting and Clinical Psychology*, 1968, *32*, 596–602.

Brownmiller, S. *Against our will: Men, women, and rape*. New York: Simon & Schuster, 1975.

Bulman, R. J., & Wortman, C. B. Attributions of blame and coping in the "real world": Severe accident victims react to their lot. *Journal of Personality and Social Psychology*, 1977, *35*, 351–363.

Burgess, A. W., & Holmstrom, L. L. Rape trauma syndrome. *American Journal of Psychiatry*, 1974, *131*, 981–985. (a)

Burgess, A. W., & Holmstrom, L. L. *Rape: Victims of crisis*. Bowie, Md.: Robert J. Brady, 1974. (b)

Burgess, A. W., & Holmstrom, L. L. Coping behavior of the rape victim. *American Journal of Psychiatry*, 1976, *133*, 413–318.

Dweck, C. S. The role of expectations and attributions in the alleviation of learned helplessness. *Journal of Personality and Social Psychology*, 1975, *31*, 674–685.

Eagly, A. H. Involvement as a determinant of response to favorable and unfavorable information. *Journal of Personality and Social Psychology Monograph*, 1967, *7*(3, Pt. 2).

Elig, T., & Frieze, I. A multidimensional scheme for coding and interjecting perceived causality for success and failure events: The coding scheme of perceived causality (CSPC). *JSAS: Catalog of Selected Documents in Psychology*, 1975, *5*, 313. (Ms. No. 1069)

Glass, D., & Singer, J. *Urban stress*. New York: Academic Press, 1972.

Griffin, S. Rape: The all-American crime. *Ramparts Magazine*, 1971, *10*, 26–35.

Harvey, M. D., & Rule, B. G. Moral evaluations and judgments of responsibility. *Personality and Social Psychology Bulletin*, 1978, *4*, 583–588.

Hilberman, E. Rape: The ultimate violation of the self. *American Journal of Psychiatry*, 1976, *133*, 436–437.

Hursch, C. H. *The trouble with rape*. Chicago: Nelson-Hall, 1977.

Kelley, H. H. *Attribution in social interaction*. Morristown, N.J.: General Learning Press, 1971.

Langer, E. J., & Rodin, J. The effects of choice and enhanced responsibility for the aged: A field experiment in an institutional setting. *Journal of Personality and Social Psychology*, 1976, *33*, 951–955.

Lerner, M. J., & Miller, D. T. Just world research and the attribution process: Looking back and ahead. *Psychological Bulletin*, 1978, *85*, 1030–1051.

Medea, A., & Thompson, K. *Against rape*. New York: Farrar, Strauss & Giroux, 1974.

National Commission on the Causes and Prevention of Violence. *Crimes of violence* (Vol. 2). Washington, D.C.: U.S. Government Printing Office, 1969.

Nunnally, J. C. *Psychometric theory*. New York: McGraw-Hill, 1967.

Rotter, J. B. Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs*, 1966, *80*(1, Whole No. 609).

Schulz, R. Effects of control and predictability on the physical and psychological well-being of the institutionalized aged. *Journal of Personality and Social Psychology*, 1976, *33*, 563–573.

Seligman, M. E. P. *Helplessness: On depression, development and death*. San Francisco: Freeman, 1975.

Spielberger, C. D. Anxiety as an emotional state. In C. D. Spielberger (Ed.), *Anxiety: Current trends in theory and research* (Vol. 1). New York: Academic Press, 1972.

Walster, E. Assignment of responsibility for an accident. *Journal of Personality and Social Psychology*, 1966, *3*, 73–79.

Weiner, B. Achievement motivation as conceptualized by an attribution theorist. In B. Weiner (Ed.), *Achievement Motivation and attribution theory*. Morristown, N.J.: General Learning Press, 1974.

Weiner, B., Frieze, I., Kukla, A., Reed, L., Rest, S., & Rosenbaum, R. M. *Perceiving the causes of success and failure*. New York: General Learning Press, 1971.

Weis, K., & Weis, S. Victimology and the justification of rape. In I. Drapkin & E. Viano (Eds.), *Exploiters and exploited: Vol. 5. Victimology: A new focus*. Lexington, Mass.: Lexington Books, 1975.

Wortman, C. B., & Dintzer, L. Is an attributional analysis of the learned helplessness phenomenon viable? A critique of the Abramson–Seligman–Teasdale reformulation. *Journal of Abnormal Psychology*, 1978, *87*, 75–90.

Zung, W. K. A self-rating depression scale. *Archives of General Psychiatry*, 1965, *12*, 63–70.

Received November 21, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 10, 1810-1821

Explaining Attributions for Achievement: A Path Analytic Approach

William M. Bernstein
University of Texas at Austin

Walter G. Stephan
New Mexico State University

Mark H. Davis
University of Texas at Austin

A nonmotivational, expectancy-confirmation model of asymmetrical achievement attributions is compared with a theory of attributional egotism. Data permitting tests of the theories and comparisons between them were collected over an entire semester from students taking examinations in a large undergraduate course. Students provided expectancies, bases for expectancies, and attributions for three examinations. It was found that students' expectancies were unrealistically high at the beginning of the course but became more accurate over time. Path analysis indicated that expected scores were based on prior performance and, increasingly throughout the semester, on expected effort. Analysis of variance further revealed that subjects were generally egotistical in their outcome attributions, stressing internal factors following success and external ones following failure. Attributions were also influenced by the degree to which expectancies were confirmed if such expectancies were clearly based on the attribution factor in question.

Are attributions for achievement biased by a motive to protect and enhance self-esteem? Bradley was more satisfied in 1978 that the existence of the motive had been demonstrated than Miller and Ross were in 1975. Both reviews sensibly concluded, however, that probably motivational and cognitive factors together determine attributions. "The challenge remains," said Miller and Ross (1975), "for future researchers to assess the relative explanatory values of the motivational and nonmotivational interpretations of assymetrical attributions." The present investigation, which focuses on the attributions students make for their performance on examinations, attempts such a comparison of cognitive and motivational theories. Specifically, a cognitive model maintaining that attributions for exam performance are a function of expectancy confirmation is compared with an egotism model holding that good outcomes should be attributed internally and bad outcomes externally, regardless of expectancy confirmation.

The expectancy-confirmation model, first proposed by Miller and Ross (1975), is derived from Heider's (1958) common sense theory of action. According to Heider, an outcome that disconfirms an expectancy calls the basis of that expectancy into question. Miller and Ross assume that in achievement situations people generally expect and intend to succeed. If these high expectations are based on the individuals' beliefs that they possess the appropriate skills and will exert sufficient effort, then success will be attributed to the internal factors of ability and effort. Failure, on the other hand, disconfirms high expectations, weakens the individuals' faith in the internal bases for those expectations, and leads to external attributions such as luck and task difficulty.

The authors thank Frank Bean, John Loehlin, Melvin Snyder, and Robert Helmreich for commenting on earlier versions of this manuscript. This study was supported by a grant from the Hogg Foundation for Mental Health to the second author.

Requests for reprints should be sent to William M. Bernstein, 330 Mezes Hall, University of Texas at Austin, Austin, Texas 78712.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3710-1810$00.75

A comparative analysis of expectancy confirmation and egotism involves two types of inquiry. We must not only decide how well the models predict attributions in situations where we know their preconditions are satisfied (cf. Stephan, Bernstein, Stephan, & Davis, 1979, Experiment I), but we must also consider whether the models' preconditions themselves are likely to obtain in real achievement contexts. Egotism is a simple theory, applicable in most any situation where performance is relevant to self-esteem (Snyder, Stephan, & Rosenfield, 1978). Expectancy confirmation, on the other hand, is built upon a more complex set of interlocking assumptions.

The present study analyzed students' attributions for their own performance on academic examinations. We think it safe to assume that test performance is relevant to students' self-esteem and hence that egotism can make meaningful predictions of attributions in the examination context. Whether students' expectations for exam performance tend to be high and based on internal factors seems less certain. Our analysis will therefore focus first on verifying the assumptions of the expectancy–confirmation theory.

Two previous studies, both conducted in university exam settings, have measured expectations for success and bases of expectations. Simon and Feather (1973) asked students to estimate their chance of passing an upcoming exam and to rate the anticipated contributions of ability, effort (study), test difficulty, and luck on their future performance. Although high expectations of passing the exam did tend to be based on internal factors (ability and effort), the expectations did not appear to be especially high. The mean rating of confidence of passing the exam was "moderate," and 80% of all students did actually pass. Davis and Stephan (in press) examined expectancies and their bases in two separate classrooms. While the expectancies of students in both classes were overoptimistic (combined mean expected score 88%; actual score 80%), expectations in one class were strongly based on ability alone and expectations in the other were weakly related only to effort.

The results of past research do not provide clear answers to our questions about the assumptions of the expectancy-confirmation model. The present study, which examined students' expectations, their bases, and attributions over the course of an entire semester, was designed to exploit more fully the potential of the university field setting. Students from a large introductory psychology course completed questionnaires before and after each of three examinations given over the course of a semester. For each exam students reported their expected scores and the bases for these expectancies; then, after the exam, they made attributions for their actual performance. The acquisition of expectancies, their bases, actual performance, and attributions for three successive exams allows us to chronicle potential changes in these variables stemming from increasing exam experience and feedback. Path analysis (Duncan, 1966; Wright, 1921) will be used to plot the relations between the variables across the three tests.

The decision to use path analysis was based largely on two factors. First, the path approach adds graphic, descriptive continuity to the flow of related events. Second, path analysis makes it possible to examine a number of potentially significant determinants of expectations that have been ignored in previous research. Simon and Feather (1973) and Davis and Stephan (in press) examined only the four attribution factors of Weiner, Frieze, Kukla, Reed, Rest, and Rosenbaum (1971). It seems likely, however, that previous performance, previous expectations, and attributions for previous performance would also affect subsequent expectations.

After the presentation of the path model, the need for refining the expectancy-confirmation predictions of posttest attributions will be discussed. Additional analyses designed to assess the relative merits of the refined cognitive model and egotism will then be presented.

Path Analysis Method

Subjects

The subjects were 469 undergraduates enrolled in an introductory psychology class at the University

of Texas at Austin. Grades in the course were based on scores obtained on three out of four multiple choice tests administered during the semester. Pretest expectations and posttest attributions were collected for the first three tests given.

Pretest Questionnaire

Immediately before each test was administered, a short questionnaire was given to each student. Subjects were asked what score (0% to 100%) they expected to obtain on the test. They were also asked to judge the anticipated effect of ability, effort (study), test difficulty, and luck on their upcoming performance in the following manner. First they indicated whether a factor would help or hinder performance. They then estimated the extent to which the factor would influence performance on a 10-point scale running from *will slightly influence* to *will greatly influence*. If the student marked that a factor would hinder performance, the amount the factor was judged to influence performance was given a negative value. If a factor was expected to help performance, a positive value was assigned.

Posttest Questionnaire

During the first class period following the exam, each student was provided with a second questionnaire that included his or her actual percentage score on the exam as well as the predicted score. Students were asked to make attributions for their actual performance, using the same four causal factors that appeared on the pretest questionnaire. Again, helping factors were scored positively and hindering factors negatively.

Model Construction

The intercorrelations of bases of expectation, expected score, actual score, and attributions were studied by means of a path model. In path analysis it is assumed that there is a weak causal ordering among the variables. In the present study, the temporal sequence in which the questionnaires were administered (which followed the temporal sequence of real events) was used as the basis for causal ordering of the variables. Only one ordering decision could not be made by appealing to the questionnaire order. This involved the precedence of expected score and basis of expected score on the pretest questionnaires. The assumption was made that bases of expectancy precede expectations. The ordering of variables in the model then is: bases of expected score for test one, expected score for test one, score for test one, attributions for test one outcome, bases of expected score for test two, expected score for test two, score for test two, and so on. Given the large number of variables considered, the original model from which Figure 1 was derived did not include estimates of every possible path. The criteria used to construct the original model were as follows:

1. No paths were calculated *between* pretest bases of expectations or between posttest attributions (e.g., no path was estimated between effort expectations and ability expectations, or between effort attributions and ability attributions).

2. The only paths estimated from a given basis of expectancy were to (a) the immediately following expected score, (b) the immediately following actual score, (c) the immediately following attribution, and (d) the next basis of expectancy (e.g., paths were considered from XEF2 to XSCOR2, SCOR2, ATEF2, and XEF3).[1]

3. The only paths estimated from a given expected score were to (a) the actual score for which the expectancy was made, and (b) the immediately following expected score (e.g., paths estimated from XSCOR1 were to SCOR1 and XSCOR2 only).

4. The only paths estimated from an actual score were to (a) the following scores, (b) the immediately following expected score, (c) the immediately following attributions, (d) the bases of the next expected score (e.g., paths were estimated from SCOR1 to SCOR2, SCOR3, XSCOR2, ATEF1, ATAB1, ATTD1, ATLK1, XEF2, XAB2, XTD2, XLK2).

5. The only paths estimated from a given attribution were to (a) the immediately following score, (b) the immediately following expected score, (c) the immediately following basis of expectancy, and (d) the next attribution (e.g., paths were estimated from ATEF1 to SCOR2, XSCOR2, XEF2, ATEF2).

Paths were trimmed from the model constructed by the above criteria if the probability level of their associated F ratios was greater than .05 (Kerlinger & Pedhazur, 1973, p. 318). If a path was deleted from the original model, we assumed for the sake of parsimony that the contribution of the nonsignificant predictors was zero and performed new regressions with only the significant predictors included to obtain the path coefficients. A complete correlation matrix for all the variables in the model appears in Table 1. Means and standard deviations for the variables are in Table 2. The path analysis was based on data collected from the 179 students from the entire class who responded to all six questionnaires. Generalization from the restricted population seems justified, since the distribution of all scores and all expected scores for both the restricted group and the entire population is substantially the same.

---

[1] The symbol system used to refer to the variables in Figure 1 is straightforward. The prefix X stands for an expectancy. It is used in conjunction with both expected scores (XSCOR) and expected effects of a factor like ability (XAB). The numerical suffixes 1, 2, and 3 indicate the test for which the variable is relevant. Hence, XSCOR1 would be the expected score for Test 1; XEF2 would be the expected effect of effort or study on Test 2 performance. Attributions for performance are indicated by the prefix AT. For example, attributions to the task difficulty factor for Test 3 are symbolized ATTD3. Attributions to ability, luck, and effort for Test 2 performance would be indicated by ATAB2, ATLK2, and ATEF2, respectively. Actual scores are indicated as follows: Test 1 actual score is SCOR1; Test 2 is SCOR2; Test 3 is SCOR3.

Table 1
Correlation Matrix for Variables in the Path Model

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29
1. XEF1																													
2. XAB1	.25																												
3. XTD1	−.10	.23																											
4. XLK1	.41	.16	.13																										
5. XSCOR1	.31	.23	.14	.38																									
6. SCOR1	.11	.13	.07	.06	.35																								
7. ATEF1	.39	.09	.03	.11	.12	.11																							
8. ATAB1	−.01	.21	.02	.02	−.03	.45	.25																						
9. ATTD1	−.09	.04	.05	−.04	−.01	.34	.16	.42																					
10. ATLK1	.05	−.05	−.01	−.05	.06	.33	.19	.24	.31																				
11. XEF2	.20	.15	.00	.06	.16	−.04	.12	−.01	.13	.01																			
12. XAB2	.10	.35	.02	.11	.19	.30	.16	.30	.22	−.08	.21																		
13. XTD2	−.07	.13	.29	.00	.09	.24	.10	.14	.22	.08	.10	.21																	
14. XLK2	−.01	.11	.00	.16	.12	.08	−.01	.01	.13	−.02	.19	.09	.13																
15. XSCOR2	.11	.14	.10	.20	.57	.37	.14	.12	.29	.02	.47	.35	.22	.25															
16. SCOR2	.02	.17	.02	.03	.38	.49	.24	.05	.13	.05	.19	.20	.00	.13	.40														
17. ATEF2	.15	.21	.02	.05	.13	−.01	.16	−.08	−.06	.04	.49	.08	.10	.12	.27	.43													
18. ATAB2	.01	.32	.18	.06	.27	.23	.17	.30	.15	.05	.12	.46	.13	.07	.20	.49	.26												
19. ATTD2	.01	.11	.04	.06	.10	.08	.04	.01	.23	.10	−.03	−.06	.03	.06	.03	.46	.23	.29											
20. ATLK2	−.07	.21	−.01	−.03	.08	.06	.06	−.02	−.05	.04	−.10	.04	−.01	.03	−.13	.42	.22	.42	.41										
21. XEF3	.28	.15	.14	.27	.16	.06	.37	−.02	−.02	−.09	.11	.12	−.03	−.04	.18	.03	.11	.02	−.06	−.04									
22. XAB3	.11	.43	.14	.12	.17	.22	.15	.34	.16	.04	.11	.33	.21	.21	.19	.20	.17	.31	.08	.03	.25								
23. XTD3	−.07	.13	.18	.14	.09	.12	.01	.06	.33	.15	−.04	.07	.43	.19	.04	.03	.00	.22	.20	.15	.04	.31							
24. XLK3	.12	.10	−.01	.18	.14	.08	.11	.10	.03	−.04	.09	.08	.01	.18	.10	.01	.19	−.03	−.07	.02	.25	.34	.25						
25. XSCOR3	.11	.15	.09	.11	.36	.25	.20	.04	.14	−.01	.07	.19	.03	.09	.11	.27	.04	.11	.09	.06	.53	.34	.14	.15					
26. SCOR3	.12	.04	.01	.12	.22	.41	.33	.13	.12	.09	.04	.06	.00	.00	.22	.35	.07	.08	.14	.01	.38	.10	−.01	.09	.42				
27. ATEF3	.23	.08	−.02	.19	.16	.04	.31	.04	−.04	−.05	.16	.04	−.10	−.02	.12	.13	.18	.08	.04	−.05	.53	.13	−.09	.17	.34	.58			
28. ATAB3	.15	.23	.02	.16	.22	.22	.14	.26	.09	.10	.08	.31	.09	.04	.18	.22	.13	.28	.13	.03	.04	.28	.10	.09	.12	.44	.36		
29. ATTD3	.04	.06	.07	.02	.09	.16	.12	.19	.26	.09	−.01	.01	.17	.08	.04	.14	−.01	.11	.32	.11	−.07	.03	.22	.04	−.06	.42	.26	.41	
30. ATLK3	.09	.06	−.08	.02	.01	.12	.12	.14	.03	.03	.04	.06	−.02	−.02	.00	.08	.19	−.01	.05	−.02	.02	.04	−.03	.14	−.10	.45	.35	.44	.52

Note. $n = 179$. (XSCOR1, XSCOR2 and XSCOR3 are student's expected scores for Tests 1, 2, and 3, respectively. SCOR1, SCOR2, and SCOR3 are the actual scores achieved. XEF1, XEF2, and XEF3 are the expected effects of effort [study] on Test 1, 2, and 3 performance, respectively. XAB1, XAB2, and XAB3 are the expected effects of ability on Test 1, 2, and 3 performance, respectively. XTD1, XTD2, and XTD3 are the expected effects of test difficulty on Test 1, 2, and 3 performance. XLK1, XLK2, and XLK3 are the expected effects of luck on Tests 1, 2, and 3. ATEF1, ATEF2, and ATFE3 are effort attributions for the three tests. ATAB1, ATAB2, and ATAB3 are ability attributions. ATTD1, ATTD2, and ATTD3 are test difficulty attributions. ATLK1, ATLK2, and ATLK3 are luck attributions.)

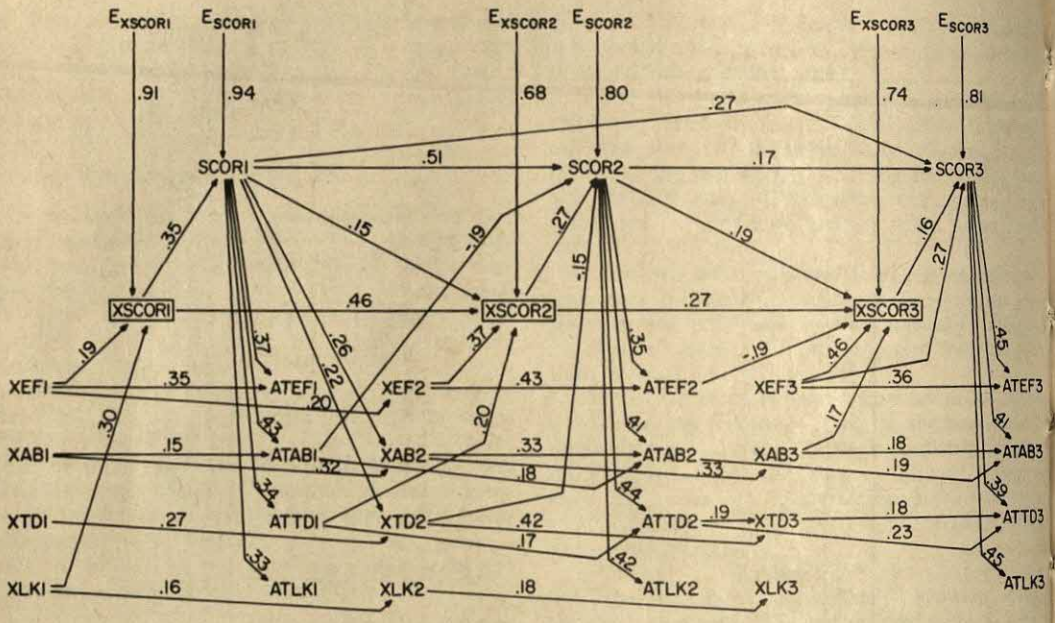


Figure 1. Path Model. ($n = 179$. XSCOR1, XSCOR2 and XSCOR3 are students' expected scores for Tests 1, 2, and 3, respectively. SCOR1, SCOR2, and SCOR3 are the actual scores achieved. XEF1, XEF2, and XEF3 are the expected effects of effort [study] on Test 1, 2, and 3 performance, respectively. XAB1, XAB2, and XAB3 are the expected effects of ability on Test 1, 2, and 3 performance, respectively. XTD1, XTD2, and XTD3 are the expected effects of test difficulty on Test 1, 2, and 3 performance. XLK1, XLK2, and XLK3 are the expected effects of luck on Test 1, 2, and 3. ATEF1, ATEF2, and ATEF3 are effort attributions for the three tests. ATAB1, ATAB2, and ATAB3 are ability attributions. ATTD1, ATTD2, and ATTD3 are test difficulty attributions. ATLK1, ATLK2, and ATLK3 are luck attributions.)

Examination of Expectations

Before presenting the path analysis results, we can examine the assumption made by Miller and Ross that expectancies in achievement situations tend to be relatively high. An inspection of the discrepancies between the students' expected and actual scores indicates that whereas they initially overestimated their actual scores by a considerable margin, they became increasingly accurate across the three tests (Table 2). Test 1 predictions showed an overestimation of 9 points. By Test 2, students were still overestimating their scores but this time by only 5 points, significantly more accurate than on Test 1, $t(178) = 4.09$, $p < .001$. By Test 3, expected–actual discrepancy was less than 1 point, and the error consisted of an underestimation of performance. Test 3 expectations were significantly more accurate than those for Test 2, $t(178) = 5.01$, $p < .001$.

Path Analysis Results [2]

Bases of Expectations

Across the three tests there is an increased reliance on perceived study (and to a lesser degree on perceived ability) as a basis of expectation (Figure 1). Whereas the overoptimistic Test 1 expectations were primarily based on luck and to a lesser extent on study, the more realistic expectations for Tests 2 and 3

[2] Interpretation of changes in path coefficients is more meaningful when we can safely assume that the structure of the causal system regulating the variables in a model remains constant over time. Structural stability between explicitly considered variables is, however, typically accompanied by stability among latent, unanalyzed sources of variability. Since stable latent variables create correlations among the residuals of explicit variables, the existence of a constant factor structure over time practically assures some violation of path model assumptions concerning disturbance terms (see Heise, 1970).

Table 2
Means and Standard Deviations for the Variables in the Path Model

Variable	*M*	*SD*	Variable	*M*	*SD*	Variable	*M*	*SD*
XEF1	5.91	5.42	XEF2	3.53	7.00	XEF3	3.02	6.76
XAB1	6.21	3.59	XAB2	4.23	4.82	XAB3	3.97	5.37
XTD1	2.28	7.30	XTD2	−1.28	7.78	XTD3	−0.26	7.44
XLK1	2.70	4.33	XLK2	1.93	5.30	XLK3	2.45	4.86
XSCOR1	81.89	7.30	XSCOR2	75.78	10.15	XSCOR3	74.02	12.55
SCOR1	72.72	11.11	SCOR2	70.82	11.01	SCOR3	74.78	11.73
SCOR1–XSCOR1	−9.17	10.92	SCOR2–XSCOR2	−4.96	11.56	SCOR3–XSCOR3	0.77	13.07
ATEF1	1.88	7.18	ATEF2	1.38	7.07	ATEF3	1.97	6.88
ATAB1	1.44	6.48	ATAB2	2.42	5.46	ATAB3	3.47	5.33
ATTD1	−4.37	6.24	ATTD2	−3.94	6.11	ATTD3	−1.70	6.56
ATLK1	−1.73	4.16	ATLK2	−0.51	4.72	ATLK3	0.91	4.78

Note. $n = 179$. (XSCOR1, XSCOR2 and XSCOR3 are students' expected scores for Tests 1, 2, and 3, respectively. SCOR1, SCOR2, and SCOR3 are the actual scores achieved. XEF1, XEF2, and XEF3 are the expected effects of effort [study] on Test 1, 2, and 3 performance, respectively. XAB1, XAB2, and XAB3 are the expected effects of ability on Test 1, 2, and 3 performance, respectively. XTD1, XTD2, and XTD3 are the expected effects of test difficulty on Test 1, 2, and 3 performance. XLK1, XLK2, and XLK3 are the expected effects of luck on Tests 1, 2, and 3. ATEF1, ATEF2, and ATEF3 are effort attributions for the three tests. ATAB1, ATAB2, and ATAB3 are ability attributions. ATTD1, ATTD2, and ATTD3 are test difficulty attributions. ATLK1, ATLK2, and ATLK3 are luck attributions.)

were directly and strongly related to perceived study. Previous expectations also played a role in determining subsequent expectations, although their influence diminished over the semester (from .46 to .27) as expectations became more accurate. No paths were found between anticipated ease or difficulty of any test (XTD) and expected score, but task difficulty attributions for Test 1 performance (ATTD1) were positively related to Test 2 expectancies. That is, students who said the ease of Test 1 aided Test 1 performance tended to predict higher scores for Test 2.

The basis-of-expectation results differ from those of Simon and Feather (1973) and Davis and Stephan (in press) in that effort emerged as a much stronger determinant than ability. There is at least one important difference in procedure that may account for the discrepancy between those studies and the present study. In the earlier studies, previous expectation and previous performance were not entered into the regression equations. When the path coefficients for XSCOR2 are calculated omitting XSCOR1 and SCOR1 as predictors, the path from effort to XSCOR2 remains unchanged, whereas the ability path shows a sizable increase (from .09 to .21). These findings suggest that had previous studies investigated the role of past expectancy and performance on subsequent expectations, the role assigned to perceived ability as a basis of expectations would have been smaller.

Despite the significantly larger paths from effort to expectancy than from ability to expectancy, the mean perception (Table 2) of ability's anticipated effect on performance was somewhat larger than that of effort for every test. That is, although most students said their ability would help them perform well on

Two types of analyses performed on the present data indicate that while factor constancy does exist, the accompanying violation of disturbance assumptions is minimal. Factor analyses performed on the 10 variables represented at each of the three time periods all yielded highly similar three-factor solutions. Accordingly, analysis of residuals revealed significant (though generally small) intercorrelations among disturbance terms within each time period. On the other hand, correlations between the residuals of the same dependent variable measured at separate time points (e.g., ATEF1, ATEF2, and ATEF3) were negligible. Note the lack of correlation between the error terms, $E = (1 - R^2)^{\frac{1}{2}}$, associated with expected and actual scores. These results indicate that an acceptable compromise between the assumptions of structural constancy and uncorrelated errors exists in these data. Thus, while the usual caveats relevant to inferring causality from correlational data of course apply here, violations of the particular assumptions of time series analysis seem minimal.

the tests, only those who thought that they had studied hard actually predicted high scores.

Determinants of Performance

The two factors that were consistently related to actual outcome for Tests 2 and 3 were the expected scores for the tests and performance on the previous test(s). For Test 1 expected score was the only significant antecedent of performance, and the size of the expectancy–performance relation is larger than that for subsequent tests. This makes sense if it is assumed that students' expectations for the initial test include information about performance on similar tasks in the past. XSCOR1 as a predictor of SCOR1 may then be thought of as confounding the two elements that are discretely represented by SCOR1 and XSCOR2 as predictors of SCOR2, and SCOR1/SCOR2 and XSCOR3 as predictors of SCOR3.

Other predictors of actual performance were ability attributions (ATAB1) and anticipated effect of test difficulty (XTD2) for SCOR2, and the anticipated effect of study (XEF3) for SCOR3. Test 1 ability attributions and the expected role of test difficulty were both negatively related to performance on Test 2. Taken together, these two findings may be indicative of complacency among students who attributed high Test 1 scores to ability and those who expected Test 2 to be easy. Both types of inference could conceivably have led to poor performance on Test 2. The positive relation between the anticipated beneficial influence of study (XEF3) on Test 3 performance and actual performance seems straightforward.

Determinants of Attributions

The results clearly show that all four attributions were related to actual performance in a logical manner, that is, they were perceived to have facilitated high scores and hindered low scores. In addition, for attributions to internal factors it was found that the expected contribution of these factors was related to subsequent attributions to them. Thus, if the subjects expected effort (and to a lesser extent, ability) to help them, they subsequently attributed their outcome to that factor. This was not the case for luck on any of the three tests. The expectancy–attribution relation was obtained for task difficulty, the other external factor, only for Test 3 (the path between XTD3 and ATTD3).

The existence of direct paths from internal bases of expectancy to attributions (unmediated by outcome) suggests that to some degree students do not alter their pretest opinions of the expected contributions of their ability and effort to test performance, even after performance feedback. Since the internal factors are knowable to a greater extent before the test than the external factors are, one might expect more pretest–posttest consistency for internal than external factors.

Attribution, Future Expectations, and Future Performance

Attributions for performance had few effects on future expectations, the bases of these expectations, or future performance. We expected attributions at least to affect bases of expectations (e.g., ATEF1 was expected to influence XEF2, ATAB1 was expected to influence XAB2, etc.). Only one such relationship was found, the path between ATTD2 and XTD3, and it was relatively small. Where attributions did affect expected scores and actual performance, their influence was not mediated by the bases of expectancy variables (e.g., ATTD1 did not affect XTD2, which in turn affected XSCOR2; rather, ATTD1 affected XSCOR2 directly). Three such direct relationships were found, and they comprised the entire influence of attributions on future expectancy and actual performance. The paths were between ATAB1 and SCOR2 (−.19), ATEF2 and XSCOR3 (−.19), and ATTD1 and XSCOR2 (.20).

The small effect of attributions on expectancies is surprising given the important role assigned to them in both Weiner's (1974) model and the recent learned helplessness formulation (Abramson, Seligman, & Teasdale, 1978). More detailed analysis of the role of attributions in expectancy formation seems warranted by these results.

Reformulation and Test of the Expectancy Model

The path model and the associated table of means suggests that a revision of the dual assumptions of the expectancy-confirmation approach to attribution is needed. Initially, people tend to believe that they will be lucky and are overoptimistic regarding their chances for success. Later, the more experience students gain with a task and the more accurate their expectations become, the more expectancies are based on internal factors, especially effort.

Given what we have learned about students' expectations and their bases, we can now make more specific predictions concerning the role of confirmation of expectations on attributions.

1. Confirmation will exert an effect on effort attributions for Tests 2 and 3 performance, since expectations were largely based on perceived study for these two tests. Confirmed expectancies should be attributed to effort more than disconfirmed expectancies.

2. Ability and task attributions should be minimally affected by confirmation since expectations were not based on either of these factors.

3. Previous research has consistently shown that unexpected outcomes are attributed more to luck than expected outcomes (McMahon, 1973; Simon & Feather, 1973; Stephan et al., 1979). We expect to find a similar result in this study for all three tests, regardless of whether or not luck was a significant predictor of expected score (as for Test 1).

The prediction of differential effects of the confirmation factor on effort (disconfirmation leading to less attribution) and luck (disconfirmation leading to more attribution) is logical. If one knows he/she has worked hard, that knowledge is not easily changed by receipt of a low grade. Rather, the perceived effect of study merely becomes discounted as a cause of poor performance relative to its role in explaining good performance. Judgments of luck's influence on performance, however, are directly related to the element of surprise and should not be affected by whether or not luck formed the basis of expectation. Disconfirmation of a positive or negative nature seems to automatically imply the operation of a greater amount of capriciousness and luck in the production of outcomes than does confirmation.

The predictions of the egotism model are, of course, not altered by the path model information, since they do not depend upon expectations or bases of expectations. The motivational model simply predicts that the internal factors of ability and effort will be seen as causing good performance and the external factors of task difficulty and luck will be seen as the causes of poor performance on all three tests.

Analysis of Variance Method

In order to test the above predictions and to contrast the effects of expectancy confirmation and egotism on attributions, we separately analyzed the attributions following each test according to a simple 2(Level of Expectancy) × 2(Confirmation of Expecancy) design. Students were informed at the beginning of the semester and on our pretest questionnaire that scores of 85%–100% constituted an A, 75–84 a B, 60–74 a C, 50–59 a D, and 49 and below an F. Students whose expected score for a given test was 75% or above made up the high expectancy group for that test. Those whose expected score was 74% or below made up the low expectancy group. Students with high expectancies whose scores were 75 or above and those with low expectancies whose scores were 74 or below made up the confirmation of expectancy groups. Students with low expectancies who scored 75 or above and those with high expectancies who scored 74 or less made up the disconfirmation groups.

A different method for scoring the attribution questions was used for this analysis than was employed for the path model (see Path Method). In order to study the interaction between expectancy and performance (i.e., the confirmation factor), the scoring of the attributions was modified to take into account the relationship between the perceived effect of a factor (did it contribute to the outcome or not) and the valence of the outcome (success or failure). Accordingly, when the rating indicated that the subject felt his outcome, high performance or low performance, could be attributed to a given factor, its rating was scored positively. If a factor was perceived to work in the direction opposite from the subject's ultimate outcome, it was scored negatively (Stephan et al., 1979; Stephan, Rosenfield, & Stephan, 1976). Thus, when the subject succeeded, all factors the subject perceived as having helped him were scored positively, and all hindering factors were scored negatively. When he failed, hindering factors were scored positively, and helping factors negatively. The reason for reversing the directionality

Table 3
Expected Scores, Actual Scores, and n *of Subjects*

Expectancy	Confirmed			Disconfirmed		
	Expected score	Actual score	*n*	Expected score	Actual score	*n*
			Test 1			
High	85.7	82.7	158	82.6	63.5	157
Low	66.4	61.2	41	66.6	78.9	10
			Test 2			
High	83.1	83.1	108	80.8	63.6	119
Low	64.2	60.1	109	66.4	79.5	28
			Test 3			
High	83.1	84.3	123	80.5	66.7	68
Low	61.9	62.5	85	66.0	81.0	54

of the scoring for subjects who failed is derived from Kelley's (1971) analysis of facilitatory and inhibitory causes. Facilitative causes are those that make the outcome more likely, whereas inhibitory causes tend to prevent the occurrence of certain outcomes. For example, high ability facilitates success but inhibits failure. Similarly, low ability facilitates failure and inhibits success. The scoring system, then, scores facilitative causes positively and inhibitory causes negatively.

Analysis of Variance Results [3]

Table 3 shows the results of the group assignment procedure for the three examinations. Three separate Expectancy × Confirmation analyses of variance (ANOVAS) show that the mean expected score of the high expectancy groups is significantly greater than that of the low expectancy groups, as planned. Test 1: $F(1, 362) = 321.43$, $p < .001$; Test 2: $F(1, 360) = 424.53$, $p < .001$; Test 3: $F(1, 326) = 590.41$, $p < .001$. In addition, high expectancy – confirmation and low expectancy – disconfirmation students on Tests 2 and 3 have higher expected scores; Test 2: $F(1, 360) = 7.57$, $p < .01$; Test 3: $F(1, 326) = 21.25$, $p < .001$.

The confirmation group assignment was equally effective. Subjects in the confirmation cells had significantly smaller absolute discrepancies between their expected and actual scores for all three tests than did students in the disconfirmation conditions. Test 1: $F(1, 362) = 33.65$, $p < .001$; Test 2: $F(1, 360) = 91.51$, $p < .001$; Test 3: $F(1, 326) = 111.22$, $p < .001$. There were three other significant effects observed for discrepancy scores. On Tests 1 and 2, high expectancy – confirmation and low expectancy – disconfirmation students tended to be more accurate with their predictions than were students in the other conditions; Test 1: $F(1, 362) = 16.08$, $p < .001$; Test 2: $F(1, 360) = 14.11$, $p < .001$. On Test 3, low expectancy subjects tended to be more accurate than were high expectancy subjects, $F(1, 326) = 4.69$, $p < .05$.

Attribution measures. Egotism would predict a greater emphasis on internal factors given success (high expectancy – confirmation and low expectancy – disconfirmation) than would occur given failure (low expectancy – confirmation and high expectancy – disconfirmation). The opposite pattern of results would be predicted for external factors. Attributions in accordance with these predictions would result in a Level of Expectancy × Confirmation interaction. As predicted, attri-

[3] All analyses used the Statistical Package for the Social Sciences (SPSS) ANOVA regression model, given the unequal cell *n*s. Also note that each of these analyses included any subjects who made a prediction and took the test relevant to the specific analysis. Hence, the *n*s vary for each analysis. This also means that the population assessed in these ANOVAS includes extra subjects whose data were not used to calculate the path model. Similar ANOVAS carried out on the restricted population used to construct the path model yielded substantially the same pattern of results as that obtained from the total population, although significance levels varied somewhat.

butions for Test 1 performance (Table 4) exhibit egotism effects for all four factors. The Expectancy × Confirmation interaction for all four factors and for an internal minus external composite (Stephan et al., 1976) were all highly significant: Effort $F(1, 349) = 5.17$, $p < .05$; ability $F(1, 356) = 13.16$, $p < .001$; task $F(1, 351) = 30.15$, $p < .001$; luck $F(1, 356) = 4.43$, $p < .05$; internal–external $F(1, 334) = 58.93$, $p < .001$. Ability and effort, the internal factors, were emphasized as the causes of good outcomes, whereas task difficulty and luck, the external factors, were perceived to be the cause of poor performance.

The only other significant effect for Test 1 was the predicted emphasis on luck by students whose expectations were disconfirmed, $F(1, 354) = 4.98$, $p < .05$.

Attributions for Test 2 also generally confirmed predictions. The egotism interaction again emerged for three of the four factors and the composite: Effort $F(1, 353) = 19.04$, $p < .001$; ability $F(1, 351) = 68.58$, $p < .001$; task $F(1, 353) = 69.18$, $p < .001$; internal–external $F(1, 335) = 116.68$, $p < .001$. Luck attributions, however, did not show a motivational effect. Effort attributions were influenced more by the confirmation factor, as predicted. Confirmed expectancies were attributed more to effort, the basis of expectations, than were disconfirmed expectancies, $F(1, 349) = 4.47$, $p < .05$. Luck attributions were again influenced by confirmation, with disconfirmation subjects stressing luck more than confirmation subjects, $F(1, 354) = 13.65$, $p < .001$. The composite measure was also influenced by the confirmation variable. Disconfirmed expectancies led to a greater reliance on external than on internal factors, $F(1, 335) = 12.19$, $p < .001$.

Attributions for Test 3 showed egotism effects on all five dependent variables: Effort $F(1, 321) = 36.69$, $p < .001$; ability $F(1, 320) = 106.69$, $p < .001$; task $F(1, 314) = 17.52$, $p < .001$; luck $F(1, 317) = 5.01$, $p < .05$; internal–external $F(1, 303) = 136.75$, $p < .001$. Confirmation again resulted in a greater reliance on effort than did disconfirmation, as predicted, $F(1, 312) = 5.43$, $p < .05$. Again, luck attributions—and, surprisingly, task attributions—were emphasized by subjects whose expectations were disconfirmed following Test 3: Luck $F(1, 317) = 15.72$, $p < .001$; task $F(1, 314) = 4.51$, $p < .05$. As on Test 2, summed external factors were stressed more than internal factors by disconfirmation subjects, $F(1, 303) = 20.92$, $p < .001$.

Table 4
Attributions to Effort, Ability, Task Difficulty, Luck, and Internal–External Factors

Attribution	High expectancy (A or B)		Low expectancy (C or less)	
	Confirmed	Disconfirmed	Confirmed	Disconfirmed
Effort				
Test 1	5.1	.9	1.8	3.2
Test 2	5.1	−.2	1.3	3.2
Test 3	5.9	.2	1.2	3.8
Ability				
Test 1	4.5	1.4	−.9	3.7
Test 2	5.1	−.1	−.2	5.4
Test 3	5.2	−1.4	−1.7	4.0
Task				
Test 1	−2.3	6.0	4.1	−.8
Test 2	−1.1	6.0	5.5	.9
Test 3	1.0	5.4	2.8	1.3
Luck				
Test 1	−.2	2.9	2.4	2.5
Test 2	1.0	2.7	1.5	3.8
Test 3	2.1	5.4	2.8	3.7
Internal–external				
Test 1	11.9	−6.9	−7.0	5.9
Test 2	10.0	−8.8	−5.9	3.7
Test 3	8.2	−12.4	−6.3	2.7

Note. Expectancy is rated according to test grade anticipated (A, B, C, or less).

Discussion

These results help to define the boundaries of two major theories concerned with the determination of attributions for performance in achievement settings. Students in the present study were egotistical in their attributions for three examination scores. They attributed A and B scores to ability and effort, and C, D,

and F performance to test difficulty and bad luck. However, attributions were also influenced by less motivated concerns. The path analysis indicated that the most important factor in determining expectations for Tests 2 and 3 was perceived amount of study. Given this clear basis of expectancy, the students' subsequent attributions to effort were influenced by their expectations in a manner consistent with the expectancy confirmation approach outlined by Miller and Ross (1975). Students whose expectations were confirmed on Tests 2 and 3 did make more attribution to effort, the basis of their expectations, than did students who received unexpected scores.

The results of this study highlight the important role of effort in the attribution process. While a number of previous studies have shown that attributions are affected by outcomes and expectancy confirmation (e.g., Frieze & Weiner, 1971; McMahon, 1973), the present study shows that over a series of trials, effort emerges as the single most important attribution factor. On Tests 2 and 3 effort attributions were influenced by both motivational and nonmotivational attribution processes. Effort is distinguished from the other attribution factors by the fact that it alone is under the actor's control. It may be easier for the actor to perceive the covariance between effort and outcome than to perceive the relationship between outcome and the other attribution factors (Kelley, 1967). For instance, an actor may attempt to assess the difficulty of the test while he/she is taking it, but an accurate assessment of its difficulty is probably dependent on social comparison information provided by subsequent informal discussion and feedback on class performance.

Ability and task difficulty attributions, in contrast to effort ascriptions, were relatively uninfluenced by the confirmation factor. Concern with self-protection and enhancement were the most important determinants of ability and task attributions in the present study. The path analysis suggested that unlike effort, neither ability nor task difficulty judgments were important in the formation of expectancies. According to our interpretation of the expectancy-confirmation model, lack of a direct relation between the expected effect of a factor and the expected level of performance makes it impossible for that model to make meaningful predictions of posttest attributions. This is because achieving or not achieving the score one expects has no clear implications for attributions to a factor unrelated to the production of the expectancy.

An exception to the last statement seems warranted only in the case of the luck factor. Attributions to luck, like those to effort, were generally influenced by both egotism and expectancy confirmation. But, unlike effort, luck was stressed as a causal factor by students whose expectancies were disconfirmed rather than those whose expectancies were confirmed. In further contrast to effort, the effect of confirmation on luck attributions did not depend upon whether luck was (as for Test 1) or was not (as for Tests 2 and 3) a basis for expectancy. This is because luck bears a more direct relation to the confirmation factor than do the other three factors. Estimates of luck's role are in a sense equivalent to judgments of surprise and hence should be greatest given disconfirmation of expectancy.

The results for the internal minus external composite showed that on all three tests successful students stressed internal factors to explain their performance and less successful students used external causes. For Tests 2 and 3, in addition to the egotism effect, it was shown that disconfirmed expectancies were associated with external attributions. In the present context, the internal–external composite might also be seen as a stable–unstable index. In Weiner's scheme, ability and task difficulty are regarded as stable factors and effort and luck as unstable factors. It is clear, however, that in a university examination setting test difficulty may vary throughout a semester. Additionally, we have demonstrated that effort formed the strongest basis of expectancy for Tests 2 and 3 and that effort attributions for these tests were consistent with the idea that judgments of one's amount of prior study remain stable across time. It therefore seems reasonable to interpret the emphasis on external factors by disconfirmation subjects on Tests 2 and 3 as an emphasis on unstable factors (see Frieze & Weiner, 1971).

A third and perhaps more parsimonious way to conceptualize these results is to view the composite as a comparison between attributions to factors that were important in forming expectancies versus those that were not. Given disconfirmation of Tests 2 and 3 expectancies, task difficulty and luck, factors unrelated to Tests 2 and 3 expectancies are emphasized relative to effort and ability.

Deciding upon a way to conceptualize these results depends upon the particular expectancy attribution theory one adopts. Some theorists maintain that the stability of attribution factors is the dimension most relevant for explaining the effects of confirmation and disconfirmation on causal attributions (Frieze & Weiner, 1971; McMahon, 1973; Valle & Frieze, 1976). Others stress the importance of the internal–external dimension (Feather, 1969; Miller & Ross, 1975). Still others have deemphasized the role of expectancies and have stressed the motive to make self-serving attributions regardless of the degree of expectancy confirmation (Bradley, 1978; Stephan et al., 1979). The present study suggests that closer examination of the antecedents of expectancies may be a useful way to disentangle, compare, and reintegrate these competing hypotheses.

References

Abramson, L. Y., Seligman, M. E. P., & Teasdale, S. D. Learned helplessness in humans: Critique and reformulation. *Journal of Abnormal Psychology,* 1978, *87,* 49–74.

Bradley, G. W. Self-serving biases in the attribution process: A reexamination of the fact or fiction question. *Journal of Personality and Social Psychology,* 1978, *36,* 56–71.

Davis, M. H., & Stephan, W. G. Attributions for exam performance. *Journal of Applied Social Psychology,* in press.

Duncan, O. P. Path analysis: Sociological examples. *American Journal of Sociology,* 1966, *72,* 1–16.

Feather, N. T. Attribution of responsibility and valence of success and failure in relation to initial confidence and task performance. *Journal of Personality and Social Psychology,* 1969, *13,* 129–144.

Frieze, I., & Weiner, B. Cue inhibition and attributional judgments for success and failure. *Journal of Personality and Social Psychology,* 1971, *39,* 591–605.

Heider, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

Heise, D. R. Causal inference from panel data. In E. F. Borgatta and G. W. Bohrnstedt (Eds.), *Sociological Methodology 1970.* San Francisco: Jossey-Bass, 1970.

Kelley, H. H. Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 15). Lincoln: University of Nebraska Press, 1967.

Kelley, H. H. *Attribution in social interaction.* Morristown, N.J.: General Learning Press, 1971.

Kerlinger, F. N., & Pedhazur, E. J. *Multiple regression in behavioral research.* New York: Holt, Rinehart & Winston, 1973.

McMahon, I. D. Relationship between causal attributions and expectancy of success. *Journal of Personality and Social Psychology,* 1973, *28,* 108–114.

Miller, D. T., & Ross, M. Self-serving biases in the attribution of causality: Fact or fiction? *Psychological Bulletin,* 1975, *82,* 213–225.

Simon, J. G., & Feather, N. T. Causal attributions for success and failure of university examinations. *Journal of Educational Psychology,* 1973, *64,* 46–56.

Snyder, M. L., Stephan, W. G., & Rosenfield, D. Attributional egotism. In J. Harvey, W. Ickes, & R. Kidd (Eds.), *New Directions in Attribution Research* (Vol. 2). Hillsdale, N.J.: Erlbaum, 1978.

Stephan, W. G., Bernstein, W. M., Stephan, C., & Davis, M. H. Attributions for achievement: Egotism vs. expectancy confirmation. *Social Psychology Quarterly,* 1979, *42,* 5–17.

Stephan, W. G., Rosenfield, D., & Stephan, C. Egotism in males and females. *Journal of Personality and Social Psychology,* 1976, *34,* 1161–1167.

Valle, V., & Frieze, I. Stability of causal attributions as a mediator in changing expectations for success. *Journal of Personality and Social Psychology,* 1976, *33,* 579–587.

Weiner, B. Achievement motivation as conceptualized by an attribution theorist. In B. Weiner (Ed.), *Achievement motivation and attribution theory.* Morristown, N.J.: General Learning Press, 1974.

Weiner, B., Frieze, I., Kukla, A., Reed, L., Rest, S., & Rosenbaum, R. *Perceiving the causes of success and failure.* Morristown N.J.: General Learning Press, 1971.

Wright, S. Correlation and causation. *Journal of Agricultural Research,* 1921, *20,* 557–585.

Received November 28, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 10, 1822-1834

Emergencies: What Are They and Do They Influence Bystanders to Intervene?

R. Lance Shotland and Ted L. Huston
Pennsylvania State University

Social psychological research on helping has, in part, been concerned with the intervention of bystanders into emergencies. Pertinent empirical literature does not seem to be available on what factors bystanders use to define an emergency nor the effect of such a decision on the rate of helping. A series of four studies was conducted to answer these questions. We found that (a) Emergencies are a subclass of problem situation that usually result from accidents; (b) there is a high degree of agreement concerning what problem situations are definitely an emergency; (c) emergency situations are differentiated from other problem situations by threat of harm or actual harm worsening with time, unavailability of an easy solution to the problem, and necessity of obtaining outside help to solve the problem; (d) disagreement on whether a problem situation is an emergency or not results from differing perceptions of the degree to which threat of harm or actual harm worsens with time; (e) bystanders are more likely to help in emergency than in nonemergency problem situations. The results were interpreted as indicating that the need of the victim is a salient feature used by bystanders in determining whether or not to help.

Social psychological research on helping has, in part, been concerned with the intervention of bystanders into emergencies. Surprisingly, the question of whether onlookers are more apt to help in emergencies than in nonemergencies has not been investigated, nor does there seem to be any firm theoretical basis for predicting the effect of whether or not a given situation is seen as an emergency on the rate of helping. Moreover, no evidence is available concerning the considerations bystanders use in labeling a situation as an emergency. These are the problems we wish to investigate.

This research was supported by a grant from the Central Fund for Research, College of Liberal Arts, Pennsylvania State University. We thank Steven Marks, Carolyn Edwards, and Mark Peterson for their work on various phases of the project. We also thank Jerald Barnett, Craig Bowman, Kenneth Falkenbach, Gary Foley, Christine Pogue, Alice Pope, Adrienne Ratella, and Lisa Tate for serving as confederates in Study 4.

Requests for reprints should be sent to R. Lance Shotland, Department of Psychology, Pennsylvania State University, University Park, Pennsylvania 16802.

Latané and Darley (1970), in their monograph on the unresponsive bystander, come to an equivocal position on whether a person is more or less likely to respond to an emergency situation. In laying out the reasons why bystanders may fail to help in emergencies, these writers make the following observation about emergencies: "The picture is a grim one. Faced with a situation in which he (the bystander) can gain no benefit, unable to rely on past experience, on the experience of others, or on forethought and planning, denied the opportunity to carefully consider his course of action, the bystander to an emergency is in an unenviable position. It is perhaps surprising that anyone should intervene at all" (p. 31).

Latané and Darley, in proposing their decision-making model, however, suggest that perceiving an event as an emergency is a critical determinant of whether a bystander will provide help. "If an individual is to intervene in an emergency," according to Latané and Darley (1970), "he must make not just one, but a *series* of decisions. Only one particular set of choices will lead him to

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3710-1822$00.75

take action in the situation" (p. 31). In order to intervene the onlooker must (a) notice that something is happening, (b) *interpret the event as an emergency,* (c) assume responsibility for taking action, (d) decide on the appropriate form of assistance, and (e) decide how to implement the assistance.

Piliavin, Piliavin, and Rodin (1975) take a somewhat different stance concerning the impact of emergencies on helping. According to their model, emergencies produce arousal that becomes more unpleasant as it increases. In order to reduce the unpleasant arousal, bystanders take action, the form of which depends on the line of action they perceive to be most likely to reduce their arousal quickly and effectively. Emergencies ought to intensify the actions of bystanders, but whether the onlookers run toward or away from the scene is predictable from elements in the situation other than its emergency character.

Since we are interested in determining whether emergencies stimulate helping, our first task was to identify such events and to differentiate them from other situations where help is not a matter of urgency or necessity. The emergency intervention literature has confronted onlookers with a variety of incidents, including a collapsed subways rider (Piliavin, Rodin, & Piliavin, 1969), a physical assault (Shotland & Straw, 1976), and a victim of stomach cramps (Staub, 1974). No systematic attempt has been mounted to identify the range of events perceived to be of an emergency nature.

This article reports data from four studies. The first three studies were concerned with identifying emergencies and defining their properties. The fourth study examined whether the emergency character of an event influences helping.

Study 1

The first study had two phases. In the first phase (Part A) participants were asked to identify the kinds of events that come to mind when they think of an emergency. The events identified by these respondents, as well as a sample of the kinds of events simulated by researchers studying bystander intervention, were rated as to their degree of emergency by a new set of respondents during the second phase (Part B) of the study.

Part A

Method

Subjects. The subjects were 39 men and 16 women enrolled in a social psychology course, winter term 1976, at The Pennsylvania State University.

Procedure. Subjects were asked to list events they considered an emergency that they had witnessed, imagined, or knew about. The term *emergency* was left undefined, given our interest in determining the subject's own definition.

Results

In order of frequency, the following events were identified: auto accident, fire, heart attack, drowning, flood, poisoning, airplane crash, hurricane/tornado, earthquake, rape, and war. The events can be classified into the following categories, listed in order of the number of events identified: (a) accidents; (b) illness, disease, and death due to natural causes; (c) natural disasters; (d) acts of violence and crime; (e) psychological and emotional problems (e.g., attempted suicide, nervous breakdown); (f) political problems and crises (e.g., overpopulation, oil shortage); (g) everyday problems and crises (e.g., late for an appointment, studying all night for an exam, and (h) miscellaneous (e.g., electrical power failure, financial emergency).

Part B

Method

The emergency situations generated by the subjects in Part A were augmented by a list of emergencies simulated in previous research (Huston & Korte, 1976), as well as by situations generated by the investigators.

Subjects. The subjects were 69 females and 21 males enrolled in a social psychology course, spring term 1976, at the Pennsylvania State University.

Procedure. The final pool of items consisted of 96 descriptions of events. Half of the items were put into Form A, the other half into Form B. This was done to prevent respondents from becoming fatigued. Participants rated each event in terms of whether it is an emergency on the following 5-point scale: (1) definitely an emergency, (2) probably an

Table 1
Classification of Events

Event	*M*	*SD*	Category
Definite emergency			
Cut artery, profuse bleeding	1.00	.00	Accident
House ablaze, people screaming for help	1.00	.00	Accident
Child poisoned	1.00	.00	Accident
Child swallowed razor blade	1.00	.00	Accident
Person in shock, lying on ground	1.07	.25	Accident
Heart attack	1.02	.15	Illness, disease
Airplane crash, some passengers still alive	1.02	.15	Accident
Rape in progress	1.09	.36	Act of violence
Car accident, driver motionless on ground	1.09	.29	Accident
Drug overdose	1.09	.46	Accident
Not sure[a]			
Terminal cancer, 3 mos. to live	2.60	1.35	Illness, disease
Lost person in woods shouting for help	2.64	1.03	Miscellaneous
Car broken down on side of road	2.72	.98	Everyday problem
Emergency landing by airplane, passengers shaken but unhurt	2.78	1.08	Accident
Burning leaves, no one attending, windy day	2.81	.95	Accident
Rock stars in danger of being trampled by fans	2.84	1.18	Miscellaneous
Mildly intoxicated friend wants to drive home	2.84	1.18	Everyday problem
Blackout in neighborhood	2.98	1.03	Miscellaneous
Man on first date with woman he likes, no money	3.07	1.08	Everyday problem
Friend states he is miserable and depressed	3.18	.92	Psychological and emotional problems
Probably not an emergency			
Car door chips other car's paint	3.75	1.06	Everyday problem
Telethon wants to get $20 million for muscular dystrophy	3.75	1.17	Miscellaneous
Car parked by fire hydrant	3.98	.98	Everyday problem
Scraped knee	4.05	.84	Everyday problem
Tramp panhandling	4.05	.99	Everyday problem
Tray of hors d'oeuvres drops at party	4.09	.91	Everyday problem
Child holds breath in protest	4.18	.97	Everyday problem
Person loses dime in pay phone	4.23	.94	Everyday problem
Cat is stuck in a tree	4.23	.73	Everyday problem
Someone has cigar, no match	4.87	.54	Everyday problem

[a] *M*s closest to 3.

emergency, (3) not sure, (4) probably not an emergency, (5) definitely not an emergency.

Results

The most striking finding is that our participants agree considerably concerning what events are emergencies. In fact, four items were seen as definitely an emergency by *all* respondents; several other events were seen as definite emergencies by *all but a few* of the respondents. As can be seen in Table 1, the events seen as emergencies were primarily accidents. At the other end, problems seen as probably not an emergency were categorized, for the most part, as everyday problems. Situations that elicited uncertainty concerning their emergency nature fit a number of categories, including everyday problems, psychological and emotional problems, and other miscellaneous occurrences. From Table 1 it is apparent that emergencies appear to be a special class of problem situations.

As noted earlier, a sampling of events simulated in the laboratory in previous research was rated by our respondents in terms of their degree of emergency. These events tended to be rated as probably an emergency. Representative events are as follows: a collapsed subway rider ($M = 1.50$; $SD = .73$), a person lying on the floor having fallen from a ladder ($M = 1.61$; $SD = .64$), an individual suffering from an asthmatic attack ($M = 2.00$; $SD = 1.10$), a person groaning and lying in a doorway ($M = 2.02$; $SD = .77$), and an individual complaining of severe stomach cramps ($M = 2.38$; $SD = .86$).

Study 2

The next step is to identify the defining qualities of emergencies that distinguish them from other problem situations. Latané and Darley (1970) have drawn a portrait of emergencies. We shall draw heavily from their observations, elaborating them in terms of insights gleaned from Table 1. First, as can be seen in Table 1, events seen as definite emergencies *all happen suddenly and unexpectedly*. Victims and bystanders alike find themselves unprepared. Eight of the 10 events most consistently seen as definite emergencies were accidents; the remaining 2—heart attack and rape—also catch their victims by surprise. The fact that something happens without warning, however, does not necessarily indicate an emergency. Dropping a tray of hors d'oeuvres at a party may be embarrassing, but according to our respondents, it is probably not an emergency.

A second important feature of emergencies is that they involve *threat of harm* or actual harm to the victim. Events seen by our respondents as definite emergencies all threaten the life of the victim. A cut artery, poisoning, and drug overdose are unambiguous with regard to their life-threatening potential. In each case, in addition, *immediate, urgent action is required* if the victim is to be saved. Too much loss of blood or the failure to find an antidote quickly to counteract the poison or the drug would be fatal. In contrast, events that were seen as nonemergency problems or that our respondents were unsure how to classify pose little or no threat of harm to the victim. A child holding his breadth in protest may put a scare into his parents, but our respondents did not think of it as much of an emergency. Similarly, a person who scrapes a knee or loses a dime in a pay phone may feel distressed, but others are unlikely to see this situation as an emergency.

Another important feature of an emergency is that *outside help is necessary*. Without the intervention of an onlooker, the victim is unable to keep the situation from worsening. A helpless victim is characteristic of all situations classified as definite emergencies in Table 1. The fact that a victim is helpless, however, does not necessarily indicate that an emergency is at hand. Victims of terminal cancer are helpless, but our respondents were uncertain whether to classify their circumstances as an emergency. This suggests, perhaps, that in order for an event to be seen as an emergency, *it must be possible for onlookers to do something* to help the victim.

The above analysis suggests that emergencies involve several essential features:

1. They occur suddenly and unexpectedly.
2. They involve threat of harm or actual harm to the *victim*.
3. The threat increases with time.
4. The victim is helpless.
5. Potentially effective intervention must be possible.

According to the data in Study 1, it appears that events that involve one of the above features but not others are unlikely to be seen as definite emergencies. We expect to see "threat of harm" (2) and "threat of harm increasing with time" (3) to be moderately to highly correlated. A moderate correlation is predicted between "something can be done to correct the situation" (4) and "outside help is necessary" (5).

Method

Brief scales comprised of both positively and negatively worded Likert-type items were constructed to measure each of the factors that were hypothesized to define an event as an emergency. The scale designed to measure the degree to which an event occurred suddenly and unexpectedly consists of 10 items such as "the incident occurred without warning" and "the situation could have been anticipated." Guttman's lamda-3 index of reliability was measured

as .88 and .80 on two data sets using different clusters of situations. The scale measuring threat of harm or actual harm consisted of 13 items yielding a Guttman lambda-3 index of reliability of .83 and .88. A representative item from this scale is "individuals or property involved will only benefit from this situation." The scale indexing threat of harm or actual harm increasing with time had reliabilities of .79 and .79 with a total of 8 items. An example from this scale is "the probability of threat or harm to the person or property increases with time." The scale designed to tap whether something can be done to correct the situation consisted of 9 items yielding a Guttman lambda-3 of .83 and .79. An example of an item from this scale is: "help will be effective in solving the problem." The scale designed to ascertain whether outside help is necessary contained 10 items and had Guttman reliabilities of .75 and .79. "Others are needed to help change the situation" is an item taken from this scale.

The items from the five scales were randomized along with an additional 5-point Likert-type item designed to measure subjects' perceptions of a problem situation as an emergency. The anchor points of the latter measure were (1) definitely an emergency and (5) definitely not an emergency, with (3) being not sure.

To increase the generalizability of the results, we decided to investigate the defining features of three problem situations. Two situations that were rated as "probably an emergency" by subjects in Study 1 were randomly chosen. We used situations rated as "probably an emergency" for two reasons. First, situations rated as "definitely an emergency" had either no or very little variance. Second, as we pointed out previously, this is the category that most of the situations used by previous researchers investigating emergency intervention phenomena had used in their research. The first emergency situation was "you see a driver weaving from one side of the road to the other. He is apparently drunk" ($M = 2.07$; $SD = 1.04$). The second situation was: "you see the classmate next to you have an asthmatic attack" ($M = 2.00$, $SD = 1.10$). We also wished to investigate crimes as emergencies, as one might expect that crimes might be emergencies with different defining characteristics. However, the crimes on our list were all rated as "definitely an emergency" and had little or no variance. So a crime was chosen from Rossi, Waite, Bose, and Berk's (1974) list of crimes ordered in terms of seriousness. We chose a crime that was not seen as extremely serious. This problem situation was: "you see a young man grab the purse from an elderly woman and run away."

Subjects. The subjects were 279 undergraduates from an introductory psychology course at The Pennsylvania State University during the spring term of 1977. Subjects were randomly assigned a given emergency situation: 57 males and 43 females were assigned to the drunk driver situation; 50 males and 38 females were assigned to rate the asthmatic attack; and 45 males and 45 females, plus one respondent whose sex was not ascertained, rated the purse-snatching incident.

Procedure. The subjects were given the questionnaire described above. The questionnaire contained a front page that listed one of the three problem situations and provided instructions to the subject to rate the situation on each of the items listed on the following pages. A sample of the Likert scale format was included on the front page of the questionnaire as well as on each succeeding page.

Results

As Table 2 illustrates, if we view across all three emergency situations and look at the half of the matrix corrected for attenuation, the scale measuring the degree to which an event is seen as sudden and unexpected bears little relationship to the other scales. The remaining four scales are moderately related with one another and share between 15.2% of the variance (threats of harm with outside help necessary) and 38.4% of the variance

Table 2
Attenuated and Unattenuated Pearson Correlation Matrix Indicating the Interrelationships Between the Variables Hypothesized to be Components of an Emergency Across the Three Emergency Situations

Variable	1	2	3	4	5	6
1. Sudden and unexpected	—	−.19	−.07	.08	.14	−.06
2. Threat of harm	−.23	—	.35	.30	.31	.24
3. Harm increases with time	−.09	.43	—	.41	.48	.51
4. Something can be done	.10	.40	.56	—	.36	.25
5. Outside help necessary	.17	.39	.62	.50	—	.30
6. Emergency	−.06	.26	.57	.30	.35	—

Note. Below the diagonal are correlations corrected for reliability, above it are the attenuated correlations. To test for significance use the attenuated matrix. Correlations equal to or above .14 are significant at the .05 level.

Table 3
Regression Analyses of the Variables That Lead to a Decision That a Problem Situation Is an Emergency (Corrected for Attenuation)

Variable in order of insertion into regression equation	Multiple R	R^2	β	F
Problem Situation 1: Asthmatic attack				
Harm increases with time	.613	.376	1.060	41.17*
Something can be done	.677	.458	−.568	13.22*
Sudden and unexpected	.695	.484	−.158	3.54
Threat of harm	.704	.496	.140	2.13
Outside help necessary	.705	.496	−.040	.11
Problem Situation 2: Drunk driver				
Harm increases with time	.717	.514	.984	38.75*
Outside help necessary	.743	.552	−.283	4.94
Sudden and unexpected	.756	.572	.141	3.44
Something can be done	.758	.574	.093	.85
Threat of harm	.759	.576	−.080	.34
Problem Situation 3: Purse snatching				
Harm increases with time	.360	.130	.344	7.24*
Something can be done	.369	.136	.148	1.25
Threat of harm	.382	.146	−.150	1.20
Sudden and unexpected	.392	.154	−.103	1.00
Outside help necessary	.395	.156	.070	.25

* $p < .01$.

(harm increases with time with outside help necessary).

Regression was employed to determine the impact of various perceptual elements in our subjects' definitions of events as emergencies. In Table 3, R and R^2 are ordered as they were produced by a forward stepwise regression. The beta weights and F ratios are the product of the summary analysis in which the final regression equation, which had all the variables entered, was used. Thus, the F ratios and beta weights reflect the contributions of each of the variables by holding constant all of the other variables. Since differences in the reliabilities of the scales can cause disparities due to measurement error rather than to the variables themselves, we decided to conduct the analyses on matrices corrected for attenuation due to imperfect reliability. However, one must be cautious in interpreting tests of significance based upon data corrected for attenuation, in that the test statistic, as well as other coefficients, is inflated as a result of increased sampling error (Cohen & Cohen, 1975). Therefore, we have set alpha at the .01 level, and we will report differences in conclusions regarding each study that are reached by doing the regression analysis on attenuated data using the standard confidence interval. Since no sex differences were found for either the emergency question or with regard to any of the scales, the samples were combined for the regression analysis. The results are shown in Table 3.

The asthmatic attack was generally seen as probably an emergency ($M = 2.10$, $SD = .87$). There were two statistically significant predictors of the degree to which the asthmatic attack was viewed as an emergency. As shown in Table 3, the most important predictor was harm increasing with time, which was entered first and accounted for approximately 38% of the variance. Whether the incident was seen as one in which something could be done to help the victim was the second statistically significant variable and accounted for another 6.4% of the variance.

Table 4
Regression Analysis Across 96 Problem Situations of the Variables That Lead to a Decision That a Problem Situation Is an Emergency (Corrected for Attenuation)

Variable in order of insertion into regression equation	Multiple R	R^2	β	F
Harm increases with time	.856	.733	.654	11.78*
Something can be done	.908	.824	−.255	13.54*
Outside help is necessary	.916	.838	.222	7.34*
Threat of harm	.917	.840	.130	.84
Sudden and unexpected	.917	.842	.039	.80

Note. Reliabilities were calculated across emergencies rather than subjects and these reliabilities were used to correct for attenuation.
* $p < .01$.

The drunk driver situation was again seen by most of our respondents as "probably an emergency" ($M = 2.05$, $SD = .97$). Again, harm increasing with time was entered into the equation first. It accounted for approximately 51.4% of the variance. No other scale was a significant predictor of perceiving the event as an emergency.

The purse-snatching situation was rated as probably an emergency by most of our subjects ($M = 1.96$, $SD = .91$). As Table 3 shows, harm increasing with time was the only scale that was significantly tied to the perception of the event as an emergency accounting for 13% of the variance in the emergency ratings.

Discussion

Only one of the variables, threat of harm increasing with time, was a significant predictor of perceiving the event as an emergency across all three situations. No other variable was significant across any two of the situations. The same conclusion is reached if the attenuated matrix is used. We find these results surprising. Perhaps the discrepancy between our expectations and the results reflects differences in our conceptualization and our test of the model. In formulating the model we were attempting to identify the critical dimensions that differentiate emergency situations from other problem situations. The test provided in this study asked the question: What are the considerations that cause different people to see the same situation differently? This is an important question, particularly if individual differences in perceiving an event as an emergency relate to the likelihood that onlookers respond. We would now like to focus our attention on answering the question: What features do people use to differentiate emergency situations from other nonemergency problems?

Study 3

Method

The same questionnaire used in Study 2 was used in Study 3. Instead of selecting 3 problem situations, all of which were about equal in the degrees to which they were likely to be seen as an emergency, all 96 problem situations used in Study 1 were again employed. Each situation was rated by a minimum of four subjects ($M = 5.44$), who were randomly assigned to the emergency situations. The responses of subjects who rated a given emergency were averaged, thereby producing average scale scores and an average rating of the situation as an emergency. The correlational analysis, then, is across 96 problem situations rather than subjects. The variance generated with this procedure is basically due to differences in situations rather than to individual differences in the perceptions of the subjects.

Subjects. The subjects were 522 male and female undergraduates at The Pennsylvania State University. The students participated during the fall term of 1977 and winter term of 1978 and earned supplemental credit for their introductory psychology course.

Results

The regression procedures used in Study 2 were also employed in Study 3. As Table 4 indicates, and as found in Study 2, "harm

increases with time" was an important predictor accounting for 73.3% of the variance. The second statistically significant variable was "something can be done," which accounted for another 9.1% of the variance and yielded a negative beta weight. The last statistically significant variable was "outside help is necessary," accounting for another 1.4% of the variance.

The zero-order (Pearson) correlation matrix illustrating the interrelationships among the variables shown in Table 5 provides additional information. This correlation matrix, although similar to the one presented in Table 2 in some respects, differs in others.

Again, it appears that the sudden and unexpected variable bears a small relationship to the other variables. In general, however, the relationships among the other variables are much higher. Of particular note is the corrected correlation between "harm increases with time" and "threat of harm." The correlation between those two scales was .89, indicating that "threat of harm" and "harm increases with time" are basically the same variable. The correlation between "harm increases with time" and "outside help is necessary" ($r = .80$) indicates that these two variables also show a great deal of common variance.

The negative beta weight of "something can be done" along with its zero-order correlation of .05 with the emergency question and its zero-order correlation of .38 with "harm increases with time" and .22 with "outside help is necessary" indicates that "something can be done" is working as a suppressor variable in the multiple regression equation (see Nunally, 1967). This is consistent with the negative beta weight associated with the "something can be done" variable found in the asthmatic attack situation in Study 2. The regression analysis based upon the attenuated correlation matrix leads to the same general conclusion. In this analysis, threat of harm proved to be a significant predictor of perceiving events as emergencies.

Table 5
Attenuated and Unattenuated Correlation Matrix Indicating the Interrelationships Between the Variables Hypothesized to be Components of an Emergency Across the 96 Emergency Situations

Variable	1	2	3	4	5	6
1. Sudden and unexpected	—	.16	.13	.04	.20	.21
2. Threat of harm	.19	—	.74	.03	.53	.78
3. Harm increases with time	.15	.89	—	.30	.65	.78
4. Something can be done	−.05	.04	.38	—	.17	.04
5. Outside help necessary	.24	.65	.80	.22	—	.70
6. Emergency	.23	.85	.86	.05	.78	—

Note. To test for significance use the attenuated matrix (i.e., the correlations above the diagonal). Correlations equal to or above .17 are significant at the .10 level, correlations equal to or above .20 are significant at the .05 level.

Discussion

It appears that a major determinant of whether or not a problem situation is seen as an emergency is the presence of harm or threat of harm. If there is a sufficient amount of harm, then almost by definition it becomes worse with time. In addition, if the amount of harm is sufficiently grave, then outside help is necessary. The negative beta weight associated with "something can be done" indicates that if there is an easy solution to the problem, it is less likely to be viewed as an emergency.

For example, Table 1 illustrated that a scraped knee was generally not seen as an emergency ($M = 4.05$). Harm is minimal, providing that infection doesn't set in. It does not get worse with time, and no outside help is required in that the victim generally can take care of the problem of cleaning the abrasion and putting a Band-Aid on it. On the other hand, Table 1 indicates that all of our subjects saw a cut artery with profuse bleeding as a definite emergency ($M = 1.00$). In this incident, there is a greater degree of harm than with a scraped knee—at the least the cut is deeper. The situation is worsening with time, in that if nothing is done the person may well bleed to death. With this type of injury an outside source of help with medical expertise is usually required to solve the

problem directly. One might expect that a person that knows that "something can be done" and precisely what to do, for example, a doctor with a cut artery and proper equipment, may regard the incident as much less of an emergency.

Of the three considerations used by respondents in Study 3 to discriminate between emergency situations and nonemergency problems, only one differentiated respondents in Study 2 in terms of the degree to which they saw events as emergencies. This factor was the degree to which they saw harm to the victim increasing with time.

Study 4

Our original intention was to unravel the concept of an emergency and, once we understood its important features, to design a factorial experiment assessing the relative importance of the elements in accounting for helping. The findings of the third study, however, clearly indicate that the defining ingredients of an emergency are tightly intertwined (see Table 5), thereby making it impossible to separate them factorially. Study 3 demonstrated that the perceptions of threat of harm or actual harm, harm increasing with time, and outside help as necessary were highly intercorrelated. Furthermore, the perception that "something can be done" acted as a suppressor variable, meaning that taken alone it did not correlate with people's ratings of events as emergencies. Although we now know what people mean by an emergency, we could not experimentally isolate its component parts. The key question remains, however, as to how people behave in emergency versus nonemergency situations, and that is what the next study was designed to answer.

Method

Given the high intercorrelations between two of the three factors used to predict an emergency and the indirect effect of the third factor, it did not seem reasonable to use a factorial design to try to decipher the relative importance of the three factors in providing help. We decided simply to examine emergencies versus nonemergencies. The major difficulty is to construct problem situations of an emergency and nonemergency nature that require the same kinds of help, so that the two types of situations are comparable. We decided that the most reasonable procedure was for the persons who needed help always to ask for a ride to a given location following their explanations of why the ride was necessary.

The setting was the student parking lot at The Pennsylvania State University. Four male and four female undergraduate and graduate students served as the help solicitors. The solicitors varied in age (from 19 to 30 years) and in physical characteristics, although none was extreme as to size, weight, or attractiveness. Males varied on length of hair and amount of facial hair. The solicitors were stationed at various points in a parking lot capable of holding 1,340 automobiles. When a subject approached an automobile in the solicitor's area, or when a car was being parked in the solicitor's area, the solicitor would approach the potential subject and explain his or her predicament following a script and keeping to a standardized, moderate level of affect. The solicitor would then ask for a ride to his or her home in a small town approximately 5 miles away. Seven emergency and nonemergency situations were used. Some examples of emergency and nonemergency situations are:

1. "Excuse me. I need some help. I'm a diabetic and I forgot to take my insulin this morning. I'm long overdue. Can you give me a ride home to Boalsburg?" (emergency).

2. "Excuse me. I need some help. I forgot to take my allergy medicine this morning and it's starting to bother me. Can you give me a ride home to Boalsburg?" (nonemergency).

3. "Excuse me. I need some help. My roommate called and she took an overdose of sleeping pills. Can you give me a ride home to Boalsburg?" (emergency).

4. "Excuse me. I need some help. My roommate called and sounded very depressed. Can you give me a ride home to Boalsburg?" (nonemergency).

Once the subject's response was ascertained he or she was immediately debriefed. A response was counted as helping if the subject offered to take the solicitor to the appointed destination.

Subjects. The subjects were 286 male and female students (208 males and 78 females) who were alone and were either entering or exiting from their automobile in a student parking lot on The Pennsylvania State University campus during spring term of 1978.

Hypotheses

The key hypothesis concerns whether people are more inclined to help in emergencies or in nonemergencies. As we indicated in the introduction to the paper, there is no theoretical basis for making a prediction. Nonetheless, we expect people to help more frequently in emergencies compared to nonemergencies because bystanders should perceive a greater

need for their assistance. In order to generalize, we employed a number of different emergency and nonemergency situations and used both male and female solicitors. Given that the subjects were of both sexes and were either arriving at or leaving the lot, we recorded this information as well. Since we collected this information, the following additional hypotheses were formulated.

1. Based upon previous research (e.g., Borofsky, Stollak, & Messé, 1971; Piliavin & Piliavin, 1972), we hypothesized that males will help more than females.

2. Latané and Dabbs' (1975) research suggests that females should receive more help than males.

3. We expect subjects who are arriving at the parking lot to be less helpful than subjects who are leaving the lot. Commuters who are arriving should be, on average, going to class, whereas those who are leaving should have greater time flexibility. Darley and Batson (1973) found that subjects in a hurry help less frequently than subjects with more time.

Results

Manipulation check. The first step is to determine whether our emergency events are seen as emergencies in comparison to our nonemergency events. To accomplish this, 168 subjects rated the situations. Each of our 14 situations was described to separate groups of 6 males and 6 females earning extra credit in an introductory psychology class. They were asked to imagine themselves walking near their car in a parking lot and having a person approach them who then made one of the 14 statements used in the study. The subjects were then asked to fill out a short questionnaire concerning the situation. The emergency question used in each of the earlier studies was again used here. The results showed that our emergency situations were indeed seen as emergencies ($M = 1.94$) in comparison to our nonemergency situations ($M = 3.14$), $F(1, 164) = 74.35$, $p < .001$. The degree to which an event was seen as an emergency was strongly correlated (.75) with judgments concerning the severity of the situation. There was no difference between males ($M = 2.64$) and females ($M = 2.44$), $F(1, 164) = 2.11$, *ns*, nor were differences found due to the specific incidents or to interactions between and among the nature of the event, the sex of the respondent, and the specific events.

Tests of hypotheses. The data from the factorial study were analyzed by means of Goodman's techniques for the assessment of interaction effects in multivariate contingency tables (Bishop, Fienberg, & Holland, 1975). Using analysis of variance terminology, the model of the four-way interaction was significant, $\chi^2(1) = 8.04$, $p < .005$. In nonemergency situations when female subjects were arriving and a male solicitor asked for a ride, *none* of the 12 subjects in that cell helped in comparison to the average helping rate of 55% across all cells. This is precisely the cell one would predict the lowest rate of help from. The models of the three-way and two-way interactions were not significant. In fact, if one lumps all four-, three-, and two-way interactions together and drops them from the model, leaving main effects, one loses a nonsignificant chi-square, $\chi^2(11) = 15.47$. Since the four variables only interact if all the variables are present and do not interact in any form with a subset of these variables, it is not useful to consider this model except in the unlikely possibility that another researcher is going to combine those four variables.

Three of the main effects were also statistically significant. More help was produced in emergency conditions (64%) than in nonemergency conditions (45%), $\chi^2(1) = 11.17$, $p < .001$. Female solicitors received more help (64%) than male solicitors (47%), $\chi^2(1) = 9.04$, $p < .01$. Males helped (61%) more than females (40%), $\chi^2(1) = 11.48$, $p < .01$. In addition, subjects who were arriving on campus (50%) were marginally less likely to help than subjects who were leaving (63%), $\chi^2(1) = 3.84$, $p < .10$. If we drop the data from the cell that caused the four-way interaction where no one helped, and use a chi-square procedure to recalculate the main effects to determine if the main effect is independent of the interaction, we find that the main effect due to the presence or absence of an emer-

gency is independent of the four-way interaction, $\chi^2(1) = 5.76$, $p < .02$. The sex of the solicitor is marginally significant, $\chi^2(1) = 3.74$, $p < .10$, as is the sex of the subject, $\chi^2(1) = 3.76$, $p < .10$. The main effect due to whether the subject is coming or going, however, is dependent on the interaction, $\chi^2(1) = 2.37$, *ns*.

Discussion

The data from Study 4, when considered together with the findings of Study 3, indicate that a prime concern motivating potential helpers is the degree to which the victim is in need of help. From Study 3 it was determined that our subjects defined an emergency primarily in terms of the presence of a sufficient amount of harm or threat of harm such that the situation was both worsening with time and that outside help was necessary.

The results of Study 4 indicate that people are more inclined to help in an emergency as compared to nonemergency situation. This research is not the only line of research that indicates that the bystander's perceptions of the need of the victim are an important determinant of whether or not help is offered. Weiss, Boyer, Lombardo, and Stitch (1973) found that the more pain a person is in, the more rapidly people will provide help. In another study, Staub and Baer (1974) investigated the severity of an incident by having a confederate feign either a heart attack or a twisted knee. Although the results were not entirely consistent, they concluded that the greater the severity, the more the helping. West and Brown (1975) used blood as a sign of severity and found more help when blood was used; in contrast, Piliavin and Piliavin (1972) found less help with the use of blood. A recent study by Clark (1975) found that the less a person is able to help himself, the more inclined people are to provide help. His data showed that a man on crutches who had dropped a book was much more likely to be helped than a man walking normally. Finally, Shotland and Johnson (1978) varied the severity of a fall and found that onlookers were more apt to help the victim after a bad fall than after a mild one. As a group, these studies document the importance of severity of need as a determinant of bystander intervention.

General Discussion

The data reported in the four studies bears on both Latané and Darley's (1970) decision-making model and the arousal reduction model of bystander intervention set forth by the Piliavins and their colleagues (Piliavin et al., 1969; Piliavin et al., 1975; Walster & Piliavin, 1972; Piliavin & Piliavin, Note 1). It will be recalled that Latané and Darley (1970) suggested that the decision to intervene is based on a series of decisions: (a) notice that something is happening, (b) interpret the event as an emergency, (c) assume responsibility for taking action, (d) decide on the appropriate form of assistance, and (e) decide how to implement the assistance. Our data suggest that the decisions are not conceptually distinct. Latané and Darley's (1970) second decision in the series—that an event is an emergency—is affected by the conclusions onlookers come to regarding the third and fourth decisions—that is, the degree to which onlookers believe they are in a position to help and take responsibility for providing help in one form or another. If the victim is able to extricate himself without the aid of a bystander, or if no form of assistance is likely to help correct the situation, then there is no action to take and therefore no reason to take responsibility for the situation. The interpretation that a problem situation is an emergency is not only based on harm or threat of harm to the victim increasing with time but also on the degree to which outside help is both necessary and feasible. Thus, the decision that an event is an emergency is colored by the judgments onlookers make concerning the importance and feasibility of taking action to correct the situation.

The work of the Piliavins has emphasized the idea that onlookers confronted with an emergency choose a course of action based on their assessment of the rewards and costs that they believe flow from various lines of behavior. Their research (Piliavin et al., 1969; Piliavin et al., 1975) has focused particularly

on the effects on bystander response of variations in the costs to the bystander of direct intervention. The data reported herein, when combined with the material summarized earlier, suggest that another important factor influencing the response of onlookers is their perception of the degree to which the victim is in need of help. Indeed, such a perception lies at the cornerstone of Schwartz's (1977) recent model of the decision processes that precede taking action in a distress situation. According to Schwartz (1977), the following sequential steps are involved in arriving at a decision regarding whether or not to take one or another form of action: (a) awareness of a person in a state of need, (b) perception that there are actions that could relieve the need, (c) recognition of own ability to provide relief, (d) apprehension of some responsibility to become involved, (e) activation of preexisting or situationally constructed personal norms of moral obligation, and (f) assessment of costs and evaluation of probable outcomes of taking action. Depending on the assessment of costs, onlookers are expected either to take action or not, or to reassess and redefine the situation by doing such things as denying the victim's need or by denying their own responsibility to respond. The fourth study in our series varied the potential helper's perception of the degree to which the victim was in need and assessed the degree to which the perception of need influenced the frequency of helping behavior. Models of bystander intervention have emphasized the hedonistic view of onlooker behavior and underplayed the altruistic features of the act. Although a plausible case can be made for assuming that reducing another's distress is rewarding, a strong argument can be set forth that bystander behavior can be understood more readily by focusing on objective costs to the bystander weighted against the benefits to the victim. If this is, in fact, what bystanders do, then it would appear that a model positing "limited altruism" might be more accurate than the hedonistic model.

Reference Note

1. Piliavin, J. A., & Piliavin, I. M. *The good samaritan: Why does he help?* Unpublished manuscript, University of Wisconsin, 1976.

References

Bishop, Y., Fienberg, S., & Holland, P. *Discrete multivariate analysis: Theory and Practice.* Cambridge: MIT Press, 1975.

Borofsky, G., Stollak, G., & Messé, L. Bystander reactions to physical assault: Sex differences in socially responsible behavior. *Journal of Experimental Social Psychology,* 1971, *7,* 313–318.

Clark, R. D. The effects of reinforcement, punishment, and dependency on helping behavior. *Personality and Social Psychology Bulletin,* 1975, *1,* 596–599.

Cohen, J., & Cohen, P. *Applied multiple regression/correlation analysis for the behavioral sciences.* Hillsdale, N.J.: Erlbaum, 1975.

Darley, J. M., & Batson, C. D. "From Jerusalem to Jericho": A study of situational and dispositional variables in helping behavior. *Journal of Personality and Social Psychology,* 1973, *27,* 100–108.

Huston, T. L., & Korte, C. The responsive bystander: Why he helps. In T. Lickona (Ed.), *Moral development and behavior.* New York: Holt, Rinehart & Winston, 1976.

Latané, B., & Dabbs, J. Sex, group size, and helping in three cities. *Sociometry,* 1975, *38,* 180–194.

Latané, B., & Darley, J. M. *The unresponsive bystander: Why doesn't he help?* New York: Appleton-Century-Crofts, 1970.

Nunally, J. C. *Psychometric Theory.* New York: McGraw-Hill, 1967.

Piliavin, I. M., Piliavin, J. A., & Rodin, J. A. Costs, diffusion, and the stigmatized victim. *Journal of Personality and Social Psychology,* 1975, *32,* 429–438.

Piliavin, I. M., Rodin, J., & Piliavin, J. A. Good samaritanism: An underground phenomenon? *Journal of Personality and Social Psychology,* 1969, *13,* 289–299.

Piliavin, J., & Piliavin, I. Effects of blood on reactions to a victim. *Journal of Personality and Social Psychology,* 1972, *23,* 353–361.

Schwartz, S. H. Normative influences on altruism. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10). New York: Academic Press, 1977.

Shotland, R. L., & Johnson, M. P. Bystander behavior and kinesics. *Environmental Psychology and Nonverbal Behavior,* 1978, *2,* 181–190.

Shotland, R. L., & Straw, M. K. Bystander response to an assault: When a man attacks a woman. *Journal of Personality and Social Psychology,* 1976, *34,* 990–999.

Staub, E. Helping a distressed person: Social, personality, and stimulus determinants. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 7). New York: Academic Press, 1974.

Staub, E., & Baer, R. Stimulus characteristics of a sufferer and difficulty of escape as determinants of helping. *Journal of Personality and Social Psychology,* 1974, *30,* 279–284.

Rossi, P., Waite, E., Bose, C., & Berk, R. The seriousness of crimes: Normative structure and individual differences. *American Sociological Review*, 1974, *30*, 279–284.

Walster, E., & Piliavin, J. Equity and the innocent bystander. *Journal of Social Issues*, 1972, *28*, 165–189.

Weiss, R. F., Boyer, J. L., Lombardo, J. P., & Stitch, M. H. Altruistic drive and altruistic reinforcement. *Journal of Personality and Social Psychology*, 1973, *25*, 390–400.

West, S., & Brown, T. Physical attractiveness, the severity of the emergency and helping. *Journal of Experimental Social Psychology*, 1975, *11*, 531–538.

Received November 29, 1978 ▪

Journal of Personality and Social Psychology
1979, Vol. 37, No. 10, 1835-1846

Self-Awareness and Transgression in Children: Two Field Studies

Arthur L. Beaman and Bonnel Klentz
University of Montana

Edward Diener
University of Illinois at Urbana-Champaign

Soren Svanum
Indiana University—Purdue University at Indianapolis

Two field studies explored the relationship between self-awareness and transgressive behavior. In the first study, 363 Halloween trick-or-treaters were instructed to only take one candy. Self-awareness induced by the presence of a mirror placed behind the candy bowl decreased transgression rates for children who had been individuated by asking them their name and address, but did not affect the behavior of children left anonymous. Self-awareness influenced older but not younger children. Naturally occurring standards instituted by the behavior of the first child to approach the candy bowl in each group were shown to interact with the experimenter's verbally stated standard. The behavior of 349 subjects in the second study replicated the findings in the first study. Additionally, when no standard was stated by the experimenter, children took more candy when not self-aware than when self-aware.

The theory of objective self-awareness (Duval & Wicklund, 1972; Wicklund, 1975) proposes that a person's attention may be focused inward toward the self (self-aware) or outward toward the environment. According to the theory, when self-aware, a person is likely to focus on the standards of behavior appropriate to the setting and upon the consistency of his/her behavior with these standards. Self-aware individuals are more likely to self-regulate (Kanfer, 1970) their behaviors to be consistent with normative standards when such standards are salient. Several laboratory studies have demonstrated such effects. Diener and Wallbom (1976) found college students cheated less on an "intelligence test" if they were made self-aware by observing their reflection in a mirror and listening to their own tape-recorded voice. It has also been shown that self-awareness may either increase or decrease aggression in a laboratory situation, depending on the direction of the experimenter-produced norm (Carver, 1975; Scheier, Fenigstein, & Buss, 1974).

Transgressing a Standard

The present research extended the theory to a new population and further examined the relationship between self-awareness and self-regulation. A standard of behavior was made salient to Halloween trick-or-treaters by instructing them to take only one candy. They were then presented with an opportunity to "transgress" by taking extra candy. (It should be remembered that self-awareness theory itself is not a moralistic theory but one of consistency between behaviors and any standard of interest.) By varying the presence or absence of a mirror, it was possible to examine the effect of self-awareness on the rate of transgression (taking extra candy). The field study would reduce demands that might occur in laboratory studies and offer generalization to a noncollege population. It was hypothesized that the presence of the mirror would decrease the incidence of transgression.

We thank the many competent assistants who aided in the implementation of these two studies.

Requests for reprints should be sent to Arthur L. Beaman, Department of Psychology, University of Montana, Missoula, Montana 59812.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3710-1835$00.75

We noted, however, that the children in this research would be in costume. If the children focused on their costumes, which are aspects of the external environment, mirror presence might not elicit self-awareness. Thus half of the children were asked their names and addresses to individuate them and make their identities salient just prior to entering the room where the candy and mirror were. It was expected that the mirror might have its predicted effect only in the individuated condition.

Age

By collecting age data on the children, the present study also provided some evidence concerning the developmental course of self-awareness as it affects self-regulation. It is not known whether children, either young or old, experience self-aware states or whether these states influence their behavior. The ability to be self-aware and to focus on behavior/standards discrepancies in order to regulate behavior probably develops over a period of time. Duval and Wicklund (1972) suggested that children must learn the objectlike nature of the self before they can become self-aware. Although not speculating about the age at which this occurs, Duval and Wicklund argue that self-awareness arises as the child repeatedly encounters "situations that cause him to examine dimensions of himself . . . to build up a unified conception of himself, a conception that will constitute a comprehensive causal agent self. . ." (p. 52). It is also possible that children have not developed internalized standards at young ages. Therefore we hypothesized that the self-awareness effects would vary as a function of age, with the older children being relatively more influenced by the manipulation.

Naturally Occurring Standards

The present design also allowed an examination of the influence of self-awareness on naturally occurring modeling. The behavior of the first child to approach the candy bowl in each group might provide the subsequent children with a standard of behavior. This naturally occurring standard might support or might conflict with the admonition of the experimenter to take only one candy, depending on whether the first child did or did not transgress. The distinction between verbally stated standards and the standards expressed implicitly in the behavior of peers could prove theoretically valuable. It is possible, for example, that self-awareness may focus persons' attention differentially on one or the other standard. The present study allowed a post hoc examination of the influence of self-awareness on unstated norms as well as on those that were verbally specified.

Deindividuation Theory

A paradigm similar to the present one has been used previously (Diener, Fraser, Beaman, & Kelem, 1976; Diener, Westford, Diener, & Beaman, 1973) to investigate deindividuation theory (Zimbardo, 1970). Since only half of the trick-or-treaters in the current study were individuated (asked their name and address), the behavior of the remaining children could offer additional information on the effects of anonymity, which is a deindividuation theory input variable.

A deindividuation theory perspective would argue that transgressions increase as anonymity increases, although data on this point have been somewhat inconsistent (see Diener, 1977). This prediction is quite similar in some respects to the self-awareness theory perspective. Both imply that people are least likely to match their behavior to norms when they are not aware of themselves as individuals. Indeed, several authors (e.g., Diener, 1977) have suggested the possibility of merging self-awareness and deindividuation theories.

Because the present study included two variables traditionally considered to be antecedents of deindividuation (anonymity and group presence), predictions of deindividuation theory could also be tested. This might provide the basis for further integration of the two theories.

Experiment 1

Method

Participants and Setting

The participants were 363 children who arrived to trick-or-treat between 5:00 and 8:00 p.m. on Hal-

loween, 1976, at any of 18 selected homes. Children were used only if they were not with a parent. Thus all were old enough to be able to state their name and some description of where they lived. Furthermore, only children who came disguised by a costume, mask, or painted face were used. When too many children appeared at one time (more than 7), or if a second group of children arrived before the experimenter could leave the room, these children were not included as subjects.

The 18 homes were arranged in the same basic pattern. Inside a room near the front door was a low table approximately 5 feet (1.5 m) long. On the table was a large bowl full of wrapped, bite-sized candy bars that was replenished frequently during the evening. A large mirror was used in the self-awareness condition. The mirror rose above the table to a height of at least 5 feet (1.5 m) from the floor. The mirror was placed directly behind the table at a 90° angle such that the children would always see themselves while reaching for the candy. Within full view of the bowl was a decorative backdrop with a peephole that camouflaged an unobtrusive observer.

Procedure

A female experimenter greeted all children who came trick-or-treating, and an assistant served as the unobtrusive observer who recorded the data. Whenever a child or children arrived, the experimenter would greet the child(ren) amicably and would comment on their costumes. The experimenter then told each child, "You (or each of you) may take one of the candies. I have to go back to my work in another room." The purpose of this, as in past studies, was to make the standard of taking only one candy clearly salient. If any child had any questions about what was supposed to be done, the experimenter repeated the instructions to take one candy. She then exited and made sure she was out of sight of the children.

The observer recorded the number of children who entered the house, the number of candies taken by each child, and the estimated age and sex of each of the children. In past research and pilot studies, estimates of number of candies taken have been found to be highly reliable. With regard to age estimations, any inaccuracies would tend to obscure real differences between the age groups. If observers tend consistently to overestimate or underestimate ages, then the age data would just be shifted consistently up or down and would be less specific but would not obscure the trends present.

Since the experimenters in each house knew whether or not a mirror was present and at least the greeter knew whether or not children were individuated (see below), experimental bias was possible. However, a number of controls minimized this concern. Both mirror and no mirror conditions were run at each home, and a large number of homes and experimenters were used. Although all experimenters were carefully trained and rehearsed in the procedures, most were not on the research team involved in designing the study and hence were blind to the predictions made. Also, due to the relatively large number of conditions, systematic inaccurate recording of the data was less likely. Additionally, since the mean number of subjects per house was 26, trends in the various conditions should not have been apparent to any group of experimenters at a particular home.

Self-awareness manipulation. In this condition a large mirror was placed behind the table. Half of the homes were randomly assigned to the self-awareness condition from 5:00 to 6:30 p.m. At 6:30 p.m. mirrors were removed in these homes, whereas in the remaining homes, mirrors were placed behind the tables and the self-awareness conditions were run from 6:30 to 8:00 p.m.

Individuation manipulation. There was a concern prior to the experiment that the mirror manipulation of self-awareness might be ineffective for costumed (anonymous) children. The mirror might focus their attention on their costumes,[1] not upon themselves as individuals. Thus, it might be more difficult to make anonymous children self-aware. It was therefore considered important to shift the focus of attention from examining and appreciating one's costume to reflection upon oneself. To do this, after greeting the children and commenting on their costumes, the experimenter explicitly asked each child in the individuated condition what his or her name was and where he or she lived. The experimenter carefully repeated each child's name and address to make it salient that she knew this information about each of them. She then continued with the rest of the basic procedure by telling each child to take one candy and excusing herself to work in another room while the child(ren) entered the room where the candy was. These procedures have been used several times before and require relatively little time. Because the children are asked their names at many other homes on Halloween, they do not seem to find the request strange. No attempt was made to identify any of the costumed children in the anonymous condition. The children were assigned to the anonymous and individuated conditions alternately by groups in each home.

Results

Of the 363 children, 70 (19.3%) transgressed by taking more than one candy.[2] Chil-

[1] The importance of these thoughts received anecdotal support by the statement of one parent to the children: "Oh, look, they have a mirror so you can all see your costumes. How nice."

[2] It should be noted that there are a number of possible ways to analyze the data. The overall rate of transgression can be computed for each relevant condition and contrasted with other conditions. Two other possibilities readily come to mind. One might just examine the behavior of the first child in each group, reasoning that those individuals will be *less*

Table 1
Percentage of Children Transgressing

Group	Condition	
	No mirror	Mirror
Individuated		
%	37.7	8.9
No. of children	69	90
Anonymous		
%	19.1	19.6
No. of children	68	97

dren arriving alone ($n = 39$) transgressed somewhat less frequently than groups of children did, 10.3% versus 20.4%, $z = 1.51$, $p < .07$. Since the distribution among all the various conditions of the few children who arrived alone resulted in sample sizes much too small for statistical analysis, and since the children who arrived alone seemed to transgress less often, the subsequent analyses of self-awareness effects are based upon the 324 children who arrived in groups.[3]

Self-Awareness and Deindividuation

Because of the apparently strong interaction between self-awareness and anonymity, an examination of the interaction by a z test on the arc sine transformation of the proportion differences (Langer & Abelson, 1972) was completed. The highly significant $z = 3.41$ ($p < .001$) revealed that there was indeed a strong interaction between the two variables, and so comparisons of individual conditions were computed. Children in the no mirror condition transgressed significantly more than those in the mirror condition, 28.5% versus 14.4%, $z = 3.10$, $p < .01$.[4] As is shown in Table 1, the self-awareness effect seemed only to occur in the individuation condition, where each child's identity was made salient. In the individuated, no mirror condition 37.7% of the children transgressed, compared to 8.9% in the individuated mirror condition, $z = 4.39$, $p < .0001$.[5] However, in the anonymous no mirror condition, 19.1% transgressed, compared to 19.6% in the anonymous mirror condition (*ns*).

Deindividuation theory would not only predict lower stealing rates for children who arrive alone (as was found) but would also predict that groups of children who are anonymous would steal more than groups consisting of individuated children. However, the data indicated that overall the anonymous groups actually produced a transgression rate, 19.4%, about equal to that for the individuated conditions, 21.4%. In the no mirror condition, the anonymous children stole significantly *less* often, 19.1% versus 37.7%, $z = 2.42$, $p < .02$, than individuated children did. However, in the mirror condition the individuated groups' norm violations were reduced below those of the anonymous groups

influenced by the group than the following children would be, and hence a more conservative analysis will emerge. One could instead assign a value to each group (such as the proportion of the group transgressing) and make groups, not individuals, the unit of observation. Since the group members certainly could influence each other, a group-as-unit analysis could be preferred. However, three important considerations argue for the use of the data from all individuals. First, past research employing the Halloween paradigm has found group and individual data to provide identical statistical outcomes. Second, since groups are comprised of mixed ages and sexes, hypotheses made concerning these variables could not be examined using groups as units. Third, the potential modeling effects (which work against independence of group members) require investigating each member's behavior, and this was one intent of the present design. Nevertheless, the analyses were performed when possible using the first child and also the proportion-per-group data. In almost every case the conclusions remained the same. These results will be noted in footnotes.

[3] Even though the manipulations were administered randomly with respect to sex, age, and group size, it is possible that the children did not arrive in a random pattern. Thus it would be possible for one of these variables to be disproportionately represented in some cells. Fortunately, such was not the case. The correlation of group size by condition was .096 (*ns*). The correlation for age by condition was .15 (*ns*). And the chi-square for the Sex × Condition interaction was .3 (*ns*).

[4] In the analysis using only first children, this comparison reached only marginal significance, 22.9% versus 12.9%, $z = 1.28$, $p < .10$. Using proportion data with groups as units for each group maintained significant results, $t(103) = 1.69$, $p < .05$.

[5] Data from first children yielded significant results, 35.3% versus 9.7%, $z = 2.16$, $p < .02$. Similar results occurred with the proportion data with group-as-units analysis, $t(50) = 2.95$, $p < .005$.

Table 2
Percentage of Individuated Children Transgressing as a Function of Age and Self-Awareness

Age (years)	Overall		No mirror		Mirror		
	%	*n*	%	*n*	%	*n*	*p*
1–4	0	11	0	5	0	6	*ns*
5–8	12.2	49	13.3	15	11.8	34	*ns*
9–12	27.0	63	46.4	28	11.4	35	.01
13 and above	40.0	20	72.7	11	0	9	.001

8.9% versus 19.6%, $z = 2.09$, $p < .05$. As will be discussed later, the unexpectedly low transgression rates among anonymous subjects have also occurred in several laboratory studies.

The results for the amount of candy taken per child replicated the earlier research of Diener, Fraser, Beaman, and Kelem (1976) in obtaining a mean of roughly three candies (two extra) for each transgressing child. They tend to take the amount their hands can hold easily. However, children in the condition with the reduced stealing rate, the *individuated* mirror condition, did take fewer candies compared to those in the individuated no mirror condition, 2.68 versus 2.0, $t(32) = 1.96$, $p < .05$, one-tailed. It appears that self-awareness reduced both the rates and the amount of candy taken when children were individuated.

Age Data

The median estimated age of the children was 8. Based upon the median, the age data were divided into four categories, 1–4, 5–8, 9–12, 13 and older, prior to examining the transgression data. These age categories were examined to investigate developmental trends. Analysis of the four categories indicated a significant age effect, $\chi^2(3) = 18.9$, $p < .001$, with the percentage of children stealing increasing over the four age categories, 6.5%, 9.7%, 23.6%, 41.9%, respectively. The finding of an increase of transgression with age should be interpreted cautiously, because it may be specific to the type of transgression involved, and there may be greater self-selection of those children who trick-or-treat in the older age groups.

More interesting, however, is the interaction between self-awareness and age. Self-awareness produced its largest effect on the transgressions of older children. Whereas self-awareness decreased the transgression rate by 55% among teenagers, $z = 3.06$, $p < .001$, it decreased transgression by only 15% for preadolescents, $z = 2.05$, $p < .02$, and not at all for younger children. Based on the overall analyses, these effects should be due to the behaviors of the individuated children, since apparently only these children were self-aware in the presence of the mirror. Table 2 presents the data for the individuated children. Again, the developmental trends were apparent. Self-awareness exerted its strongest influence on children aged 9 and above, and not at all on the younger children. Similar analyses done on the anonymous children (who we suspected were not made self-aware by mirror presence) yielded no significant difference at any age grouping.

Naturally Occurring Standards

The present design allowed an opportunity to examine the effect of self-awareness both on experimenter-imposed verbal standards and behaviorally set standards of peers (i.e., the behavior of the first child to approach the bowl could serve as a modeled standard for the rest of the children). All children received the experimenters' standard instructing them to take only one candy, and the results above reveal that self-awareness enhanced the impact of this norm. Peer-induced standards via the model effects were also apparent. Whatever the first child did in each group (transgress or not), a large percentage of subsequent children did also. Whereas children ex-

Table 3
Percentage of Children Not Transgressing as a Function of Modeled Behavior and Condition

Condition	Modeled behavior: Not transgress %	n	Transgress %	n
Individuated				
No mirror	76.9	26	50.0	12
Mirror	96.4	55	25.0	4
Anonymous				
No mirror	88.0	33	33.3	6
Mirror	89.0	54	36.4	11

posed to a first child who took more than one candy transgressed frequently (60.6%), those exposed to a nontransgressing model rarely did so, 10.7%, $\chi^2(1) = 41.6$, $p < .001$. In the anonymous condition *only* this modeling effect is apparent (see Table 3), once again revealing that the self-awareness manipulation had no effect for anonymous children. However, in the individuated condition where the mirror did affect behavior, the interaction with modeling is more interesting. The interaction was again analyzed by a z test on the differences in arc sine transformed proportions, and a significant value was found, 1.83, $p < .05$. Because of the use of a one-tailed test and the small sample in two cells, the Modeling × Self-Awareness interaction should be cautiously interpreted. It should be noted that the data in Table 3 for the mirror condition indicate that *both* self-awareness and modeling exerted effects on the self-aware children. That is, independent of the overall reduction in transgression caused by mirror presence, mirror presence also enhanced the tendency of subjects to conform to the behavior exhibited by the model.

Sex

The estimation of each child's sex was obtained for 326 of the children (including children who arrived alone, where age data were available). Due to their costumes, the sex data for the remaining 37 children could not be reliably determined. There were 190 boys and 136 girls. It was found that boys transgressed significantly more than girls, 24.2% versus 10.3%, $z = 3.16$, $p < .001$. The effect of the mirror were also examined for each sex. Boys' stealing rate decreased from 35.8% (no mirror) to 15.6% (mirror), $z = 3.2$, $p < .00$. The stealing rate for girls only dropped from 13.2% (no mirror) to 8.4% (mirror), $z = .87$, $p < .20$.

One last consideration pertains to a group size analysis. There were no differences in the transgression rates of the groups of various sizes (two to seven children), except for the high rate of 44% observed in groups of size five. Removing these groups from the analyses did not alter the significance levels of the analyses. It appeared that the high rate occurred because these few groups of five happened to occur only in no mirror conditions hence self-awareness was not present to reduce the transgressions. It should be emphasized again that since there was no other relationship between group size and transgression rates, the analyses are not influenced by the distribution of groups to various conditions. In a similar vein, the distribution of children of various ages to conditions was investigated and found to have no effect on the data analyses.

Discussion

As the theory of objective self-awareness predicts (Duval & Wicklund, 1972; Wicklund, 1975), standard-consistent behavior was increased in a naturalistic setting by focusing children's attention upon themselves. This occurred only when children had first been individuated by being asked their name and address. This reduction in transgression seems to reflect an increase in compliance to a salient behavioral standard. The standard instituted in the present study was the instruction given by the experimenter to "take one candy only." The children in the present study altered their behavior to be more consistent with the standard of correctness when they were self-aware. Thus the present findings add external validity to the earlier research showing greater adherence to standard-appropriate behaviors (Carver, 1974, 1975;

Diener & Wallbom, 1976; Scheier et al., 1974).

Anonymity

In cases where the children remained anonymous, the presence of the mirror had no impact on behavior. We interpret this as an indication that the child's focus of attention was directed to aspects of his/her costume rather than to the self. Hence, no state of self-awareness seemed to develop. Although these findings are consistent with the contention that anonymity may inhibit self-awareness, this may occur only in the presence of a disguise, not as a general function of anonymity. Future research in which self-awareness is induced and anonymity is experimentally manipulated in several different ways should help clarify the way in which anonymity affects self-awareness. It may be that when one is disguised and a group member, anonymity inhibits feelings of self-awareness. If one were alone in a costume on a downtown street at noon, the disguise might enhance self-awareness. Both the manner in which one is made anonymous and the situational context in which the anonymity occurs probably influence the effect this variable has on self-awareness.

Data from the anonymous conditions were not what one would predict from deindividuation theory (Zimbardo, 1970). It was hoped that the present study might, in addition to assessing various self-awareness effects, shed light on the interaction of the two theories. Earlier research (e.g., Diener et al., 1976) has shown that group presence, and also anonymity, increases transgressions. Although some weak support for increasing transgressions among groups compared to children arriving alone ($p < .07$) was found in the present study, no overall effects were found for anonymity. In fact, anonymous children actually transgressed significantly less in the no mirror condition than individuated children did (a similar result was also reported by Zimbardo, 1970, p. 277). One possible explanation for this is that in the present study the individuated no mirror condition may have had high rates of transgression due to some artifact introduced by the procedures used.[6]

A close inspection of the individuation literature reveals that previous results for anonymity effects have been quite mixed, with anonymity manipulations sometimes increasing and sometimes decreasing antinormative behavior (Diener, 1977). Thus our results are actually not inconsistent with the previous research findings. However, our findings do not illuminate why anonymity produced unexpected effects. The puzzle is heightened by the fact that our manipulation was identical to that used in an earlier study in which anonymity produced heightened transgression. Future research on anonymity should explore the interaction of other variables with anonymity in order to shed light on the past contradictory findings. One possible explanation of the problematical data in the anonymity versus individuated no mirror conditions is that asking children their name actually made salient to them the fact that they could not be easily identified by the experimenter, especially since she was rarely known by the children. Another possibility is that asking children their names heightened the children's awareness of the identity of others in their group, thereby enhancing cohesiveness, and thus possibly strengthened a transgressive norm if it was held by members of the group.

Development of Self-Awareness

One of the major benefits of this naturalistic study was the opportunity to gather data

[6] However, there are a number of considerations that rule against this possibility. First, the same individuation manipulation has been used by some of the authors in earlier research (e.g., Diener et al., 1976) which produced expected results. Thus we doubt that in this study the same manipulation would produce a radically different psychological state. Also, as mentioned earlier, asking children their names is not an unusual practice on Halloween. One might argue that sampling error led to the self-awareness effects. In order to provide an even more conservative test of the hypothesis, the anonymous no mirror transgression rate was used as a baseline to assess the decrease noted in the individuated mirror condition. Again, the mirror condition was shown to be significantly lower, 19.1% versus 8.9%, $z = 1.88$, $p < .04$. It thus appears unlikely that self-awareness results occurred due to some unknown artifact present in one of the individuated groups.

from children of various ages, thus allowing an examination of self-awareness from a developmental viewpoint. As was seen in Table 2, individuated children 8 years of age or younger were apparently unaffected by their self-reflections in the mirror. The mirror effects appeared to begin in our study somewhere around the age of 9, and then the effects were substantial. However, since stealing rates were lower among the younger children, it is possible that a floor effect was responsible for the absence of larger reductions in the lower age groups. Moreover, one must view the specific ages in our age data cautiously, as they are based on estimates. Further, our data do not indicate whether children under a given age do not enter the self-aware state or whether they have not yet gotten to the point where the presence of their reflection in a mirror cues other self-relevant thoughts. It seems relatively certain, though, that by ages 9–12, self-awareness has an effect in reducing transgressions, and by 13 and above self-awareness substantially reduces antinormative behavior.

Behavioral Standards

In the present study self-awareness could focus one's attention on a standard set by the experimenter and/or one set by the behavior of other children. The modeling effects were examined by analyzing the transgression rates for the children who approached the candy bowl after the first child (model) did in each group. If the first child did not transgress (honest model), 89.3% of the subsequent children were also honest. If the first child in a given group did transgress (dishonest model), the honesty rate was reduced to 39.4% for the followers. These effects replicate an earlier field study (Diener et al., 1976). However, these data are potentially confounded by the possibility that "honest" and "dishonest" children may tend to trick-or-treat together.

Self-awareness in the present study magnified the modeling effects, among individuated subjects. This Modeling × Mirror interaction should not be due to cohort effects, since the mirror was randomly assigned and cohorts should not have been any more similar in the mirror than the no mirror condition. Moreover, this finding suggests that the modeled behavior provided an additional norm at the time the followers approached the candy bowl. Carver (1974) has shown that aggression, which might normally be considered antinormative, can be facilitated by self-awareness in the laboratory if the salient standard of behavior favors aggressing. Our results in the field support Carver's in suggesting that transgressing can be enhanced by self-awareness (if the model makes salient a transgressing standard). The present results suggest that self-awareness can heighten the salience of norms implicit in the behavior of others, even when these norms are not verbally stated. However, because of the small sample size in the transgressive model groups, this interpretation should be accepted cautiously until replicated. Further research might vary orthogonally the types of standards in order to explore modeling further.

Summary

Overall, the findings form a consistent pattern of support relating antinormative behavior to the theory of objective self-awareness. An alternative explanation for the reduction of stealing in the mirror condition might be that the children were more worried about getting in trouble for taking extra candy when the mirror was present than when it was not. If this were the case, however, one would expect the fear of getting in trouble to be more salient in the individuated condition than in the anonymous conditions, since it was clear to the children that their names and addresses were known by the experimenter in the individuated condition. This explanation would predict a lower transgression rate in individuated no mirror conditions, compared to the anonymous conditions. Because such a result was not found, our self-awareness analysis offers a more complete explanation for the findings than does an analysis based on fear of reprisal.

Experiment 2

To add confidence to our interpretations, we designed a second study to replicate the

self-awareness findings. Thus, conditions were included where the standard of taking only one candy was made salient and self-focus was varied. In this second study, children were sent into the candy room one at a time, and all children were individuated by asking them their names and *ages*. The request-for-age information was used instead of asking their addresses in order to provide actual age data instead of estimates (as in Experiment 1). Finally, conditions were included in which no standard was made salient, to explore what standards, if any, may be present to guide childrens' behavior when trick-or-treating, without the experimenter's suggesting a specific behavior as appropriate.

It was hypothesized, as before, that mirror-induced self-awareness would decrease the percentage of children transgressing. The age trends revealed in Experiment 1 were expected to occur again with the *actual* ages given by the children. In the no standard conditions, the experimenter did not make any statement about the appropriate number of candies to take. We expected that children would be inclined to think that more than one candy was appropriate (since the candies were smaller than in Experiment 1, see below). Therefore, overall a larger percentage of children should take more than one candy in these conditions than in the standard conditions.

Whatever standard (if any) is common among trick-or-treaters should be most apparent in the no standard – mirror condition. These children should be self-aware and therefore more likely to behave consistently with the standards that they brought with them into the situation. One possibility is that children trick-or-treating on Halloween might think it is a night when they can "get all the free candy they want" or that "it is okay to steal." If such reasoning is correct and this sort of standard is salient, then *more* children in the no standard condition should take more than one candy in the mirror condition than would do so in the no mirror condition, since such behavior would be consistent with their standards. However, we believed it reasonable to suggest that children on Halloween have the cognitive set that they are to be "given free candy." Further, children realize that what they will receive at each house is governed by the people there. In addition, most children have in general an internal standard that it is inappropriate to take more than their own share or to steal. Thus we expected a *decrease* for the mirror condition in the number of candies taken. Which of these two possibilities is correct should be revealed by the directional difference between the no mirror and mirror conditions when no experimental standard was made salient.

Method

Participants and Setting

The participants were 349 children who arrived to trick-or-treat between 5:00 and 9:00 p.m. on Halloween, 1978, at any of 13 selected homes. Again, children were not used as subjects if they arrived with a parent or in groups too large to process or if a second group arrived before the manipulations could be completed on the group already present. The candy rooms were arranged as in Experiment 1.

Procedure

The greeting of children was performed as in Experiment 1. The children were sent into the candy room *one at a time*. The candy room was arranged such that the bowl of candy and child approaching it would not be visible to either the greeter or the other children waiting to enter the candy room. The hidden rater again recorded the number of candies taken. Instead of using age estimates, each child was individually asked his or her name and age and the greeter repeated this to the child prior to sending the child into the candy room. Thus all children were individuated. Based upon the results of Experiment 1, this appeared to be necessary as a precondition for self-awareness when the child was later in front of the mirror. Each child was alone at the candy bowl (thus each child's behavior was statistically independent of other children's).

Independent variables. Self-awareness was again manipulated by the presence or absence of a mirror. Mirrors were physically placed as in Experiment 1, and the same randomization procedures were used. To make a standard salient, children were instructed to take one candy only, as before. However, additional no standard conditions (for mirror and no mirror) were added. In these cases each child was told "the candies are in the living room, go help yourself." For these children the standard for behavior would be whatever norm(s) children generally have when trick-or-treating. The

Table 4
Percentage of Children Taking More Than One Candy and Mean Candies Taken

Verbal manipulation	Self-awareness		
	No mirror	Mirror	p
Standard			
%	34.2	11.7	.001
M	2.41	1.18	.001
No. of children	76	94	
No standard			
%	75.0	58.9	.03
M	4.56	2.17	.001
No. of children	84	95	

standard and no standard conditions were alternated at each house, so all four conditions were in operation at each house at various times throughout the evening.

Dependent variables. Again the percentage of children taking more than one candy was computed. Also the exact number of candies taken by each child was recorded. In Experiment 1, small bite-sized candy bars were used. However, the transgression rate (more than one candy) was relatively low (about 10% for individuated children who arrived alone). Thus it was decided to use smaller candies for this study, to produce a higher transgression base rate, so that decreases caused by self-awareness could more easily be discerned. Several types of candies were chosen that varied in size between that of an individually-wrapped caramel and the candy size used before. This choice should also allow for a more sensitive measure, since the size of a child's hand would not limit as much the number of candies taken.

Results

Mirror Effects

Of the 349 children, 156 (44.7%) took more than one candy. The effectiveness of the verbally stated standard to take only one candy is shown by the highly significant reduction in the standard conditions when compared to the no standard conditions, 21.8% versus 66.5%, $\chi^2(1) = 68.7$, $p < .001$. Similarly, over all conditions the mirror produced a strong reduction inhibiting children from taking more than one candy, 55.6% versus 35.4%, $\chi^2(1) = 37.5$, $p < .001$. However, the comparisons of most interest were those within each of the types of standards.

The percentages of children taking more than one candy and the mean number taken are shown in Table 4. Transgressions were greatly reduced by the mirror, $\chi^2(1) = 11.2$, $p < .001$, when the standard was made salient by the experimenter. The reduction in numbers of candies taken was also significant, $t(168) = 3.29$, $p < .001$. Even though children were not given specific instructions in the no standard condition, fewer children took more than one candy when a mirror was present than when it was not, $\chi^2(1) = 4.5$, $p < .03$, and children took fewer candies on the average, $t(177) = 3.89$, $p < .001$.

Age Data

As is shown in Table 5, self-awareness affected the older children when a standard had been made salient, but not the younger ones. These results compare favorably with Experiment 1, where age estimates were used instead of actual age, as in the present study.[7] Similarly, the presence of the mirror reduced transgressions for the two oldest age categories when no standard was made salient. These decreases did not reach statistical significance.

Sex Data

Overall, there was only a marginal tendency for males to take two or more candies more frequently than females, 49.0% versus 39.7%, $\chi^2(1) = 2.59$, $p < .11$. However, if we examine only the standard conditions, where such behavior may more clearly be labeled a transgression, we find that males did transgress more, 28% versus 13%, $\chi^2(1) = 4.5$, $p < .03$, as was the case in Experiment 1. Also as before, self-awareness in this condition reduced the proportion of male transgressors from 46.8% to 11.3%, $\chi^2(1) = 13.9$, $p < .001$, whereas the reduction was nonsignificant among females (14.3% to 12.2%).

[7] The accuracy of these estimates is not of major concern with respect to the present data. Nevertheless, an estimate of their reliability may be of interest for future research. A correlation coefficient of 0.89 was obtained (based on several homes that included an extra experimenter who independently estimated ages to provide reliability data), indicating a high relationship between estimates and actual age.

Table 5
Percentage of Children in the Standard Conditions Taking More Than One Candy as a Function of Age and Self-Awareness

	Condition				
	No mirror		Mirror		
Age (years)	%	*n*	%	*n*	*p*
1–4	0	7	7.7	13	*ns*
5–8	40	20	25	32	*ns*
9–12	24	25	6.1	33	.04
13 and above	50	24	0	16	.001

Discussion

Since all findings in the second experiment replicated the earlier findings, only the extension and additions will be discussed here, in order to avoid redundancies. Overall, self-awareness again exerted a strong influence on children aged 9 and older. The self-awareness results in Experiment 2 are especially reassuring in that they were clearly unconfounded by modeling effects, since children entered the room individually. The use of actual age data supported the former estimated age findings. The mirror-induced self-awareness effect simply was not present among young children. Why this was so should be an important subject for future research.

The addition of no standard conditions, where children were not given an instruction as to how many candies they might take, yielded two important conclusions. First, the saliency of the simple verbal statement to take just one candy was demonstrated by the large reduction in the standard condition compared to the no standard condition. Second, the first evidence was provided concerning the cognitive standards by which children may generally guide their behavior in this particular setting. Apparently Halloween is a night when a child receives free candy, but that fact does not remove his/her knowledge that it is inappropriate to overindulge or to steal.

It should be reiterated at this point that although the term "transgression" has a moralistic connotation, the theory of self-awareness is a neutral one. It may be in our study that self-awareness inhibits transgressions or increases conformity to standards. Children who transgress may not be so much actively breaking rules as just failing to self-regulate. Self-awareness may be one of many variables that aid in producing self-monitoring of one's behavior. The theory has been primarily developed in terms of producing a state where one evaluates his or her behavior in terms of how consistent it is with salient social or personal standards. In such states one should act in ways that are more consistent with whatever standards are salient and are seen as applying in that situation.

In sum, the present two large-scale field studies offer strong support for the theory of self-awareness as applied to transgressive behavior among children. The first study also provided some leads concerning modeling that should be followed up in future laboratory studies. Finally, the age effects should also encourage researchers to study the developmental course of self-awareness.

References

Carver, C. S. Facilitation of physical aggression through objective self-awareness. *Journal of Experimental Social Psychology*, 1974, *10*, 365–370.

Carver, C. S. Physical aggression as a function of objective self-awareness and attitudes toward punishment. *Journal of Experimental Social Psychology*, 1975, *11*, 510–519.

Diener, E. Deindividuation: Causes and consequences. *Social Behavior and Personality*, 1977, *5*, 143–155.

Diener, E., Fraser, S. C., Beaman, A. L., & Kelem, R. T. Effects of deindividuation variables on stealing among Halloween trick-or-treaters. *Journal of Personality and Social Psychology*, 1976, *33*, 178–183.

Diener, E., & Wallbom, M. Effects of self-awareness on antinormative behavior. *Journal of Research in Personality*, 1976, *10*, 107–111.

Diener, E., Westford, K. L., Diener, C., & Beaman, A. L. Deindividuating effects of group presence and arousal on stealing by Halloween trick-or-treaters. *Proceedings of the 81st Annual Convention of the American Psychological Association*, 1973, *8*, 219–220. (Summary)

Duval, S., & Wicklund, R. A. *A theory of objective self awareness*. New York: Academic Press, 1972.

Kanfer, F. H. Self-regulation: Research, issues, speculations. In C. Neuringer & J. L. Michael (Eds.), *Behavior modification in clinical psychology*. New York: Appleton, 1970.

Langer, E. J., & Abelson, R. P. The semantics of asking a favor: How to succeed in getting help without really dying. *Journal of Personality and Social Psychology*, 1972, *24*, 26–32.

Scheier, M. F., Fenigstein, A., & Buss, A. H. Self-awareness and physical aggression. *Journal of Experimental Social Psychology*, 1974, *10*, 264–273.

Wicklund, R. A. Objective self-awareness. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. New York: Academic Press, 1975.

Zimbardo, P. G. The human choice: Individuation, reason, and order versus deindividuation, impulse, and chaos. In W. J. Arnold & D. Levine (Eds.), *Nebraska Symposium on Motivation* (Vol. 18). Lincoln: University of Nebraska Press, 1970.

Received December 4, 1978

Journal of Personality and Social Psychology
1979, Vol. 37, No. 10, 1847-1858

Variation in Behavior Perception and Ability Attribution

Darren Newtson and Richard J. Rindner
University of Virginia

Three studies were conducted. In the first, 16 subjects segmented a videotape of an actor solving 30 problems in an ascending success pattern under instructions to analyze the first half of the sequence into fine or large units. Estimates of future performance were lower when the initial part of the series was analyzed more finely. In a second study, 72 subjects viewed either an ascending or a descending success pattern; the first half of the series was presented at fast motion, normal speed, or slow motion. Slow-motion presentation induced finer units of analysis than normal speed and enhanced primacy effects; fast motion induced larger units of analysis and attenuated primacy effects. A third study compared fine-unit and large-unit analysis induced by instruction and large-unit analysis induced by fast-motion presentation of the entire ascending and descending series. Results were comparable to those of the first two studies. It was concluded that primacy effects result from a cessation of processing when a point of subjectively sufficient information is reached and that variations in level of analysis set limits on information gain in observation.

Results of several recent studies of behavior perception have supported the view that actions are discriminated by the selection of successive points of definition in the behavior stream. These points of definition, termed breakpoints, are points judged by perceivers as the points at which actions have occurred (Newtson, Engquist, & Bois, 1977; Newtson, Rindner, Miller, & LaCross, 1978). Breakpoints appear to be selected in such a way that they provide a summary of the information gained from observation of the sequence of events (Newtson & Engquist, 1976).

Newtson (1973) hypothesized that perceivers actively control information gain from behavior by selecting a greater or lesser number of breakpoints from the behavior stream. That is, observers selecting more breakpoints from a given sequence (i.e., analyzing the behavior into finer units of action) should gain more information than observers analyzing the behavior into larger units. Newtson (1973) in fact did find that subjects instructed to analyze a 5-minute action sequence into fine units had more confident and differentiated impressions of the stimulus person than those subjects instructed to analyze that behavior into large units. In a second study, it was found that the occurrence of an unexpected event prompted observers to analyze subsequent behavior more finely than observers who had not viewed the unexpected action. This result is also consistent with the view that observers actively control information gain from ongoing behavior. In the control condition of that study, where the behavior remained predictable throughout, a marked decline in number of breakpoints was observed from the beginning to the end of the sequence. Wilder (1978a) reported a similar decline in rate of segmentation over time.

Newtson (1973) noted that one consequence of this decline over time could be that the perceiver's sample of information is biased, in that it would be largely drawn from

This research was supported by U.S. Army Research Institute for the Behavioral and Social Sciences Grant DAHC 19-77-G-0010. This article is based in part on a thesis submitted by the second author to the Department of Psychology, University of Virginia, in partial fulfillment of the requirements for the master's degree. The authors thank Cheryl Miles for her assistance in analyzing the data.

Requests for reprints should be sent to Darren Newtson, Department of Psychology, University of Virginia, Charlottesville, Virginia 22903.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3710-1847$00.75

the first part of a sequence. Thus one might predict primacy effects in attributions based on the observation of behavior. Newtson (1973) raised this possibility as one alternative explanation for the primacy effects in ability attribution reported by Jones, Rock, Shaver, Goethals, and Ward (1968).

The present article reports a series of studies exploring such biased sampling phenomena. Biased sampling was induced in two different ways: by instruction as to level of perceptual analysis, as in Newtson (1973), and by variations in film speed. Filmmakers have long defined a quantity termed *event density* in terms of the number of events happening within a certain time unit. Slow- and accelerated-motion techniques are explicitly defined as manipulations more of event density than of motion and may be realized equivalently with both cutting techniques and actual manipulation of film speed (cf. Zettl, 1973, pp. 280–283). Variations in level of analysis may be thought of as subjective variations in event density that are under the perceiver's control. Thus film speed manipulations may provide a useful converging operation for the manipulation of level of perceptual analysis, although their precise effects on interpretation of events remain to be evaluated.

Experiment 1

In the first experiment, subjects viewed an actor attempting to solve 30 problems, succeeding on 5 of the first 15 and on 10 of the last 15. In one condition, subjects were instructed to analyze the series into fine units for the first half of the series and then into whatever units seemed natural and meaningful to them for the second half of the series. In a second condition, subjects analyzed the series into large units for the first half and then into natural units for the second half of the series. Subjects were next asked to make estimates of the actor's performance on a subsequent series of problems. Thus, it was reasoned, the subjects' information sample from the initial part of the series should be increased or decreased. The result should be that the impact of the initial part of the series on estimates of the actor's future performance should be greater when it is analyzed finely and less when it is analyzed into large units.

Method

Subjects. Subjects were 16 undergraduates (12 males, 4 females) recruited from an introductory psychology course at the University of Virginia.

Stimulus. A videotape was prepared showing a male actor ostensibly taking a computer-scored intelligence test. The actor read each question from a 3 × 5 inch (7.62 × 12.70 cm) card and typed a response into a teletype. Feedback was presented by means of two lights above prominent signs labeled "right" or "wrong." Thirty items were answered in this manner. The actor received "right" feedback on 5 of the first 15 items and on 10 of the second 15 items. The specific feedback pattern was as follows: 1 "right," 4 "wrong," 1 "right," 3 "wrong," 2 "right," 2 "wrong," 1 "right," 2 "wrong," 3 "right," 3 "wrong," 3 "right," 1 "wrong," and 4 "right."

Procedure. Subjects were assembled in groups of two or three persons facing a 23-inch (58 cm) video monitor. Each subject was given a hand-held button connected to an event recorder in another room. Subjects were informed that they would view a videotape of another undergraduate taking a computer-scored intelligence test. All subjects were then instructed as follows: "What I am interested in here are the ways in which people organize or break up another person's behavior into its component actions. By that I mean that people may break up another person's behavior in different ways. For example, he might turn, walk over, push a door closed, turn and walk back, and you might see each of those actions as a separate, meaningful action. Or you might see them as just one action, such as closing the door. In one condition, subjects were then instructed, "What I would like you to do is to mark off for me the *largest* actions in the sequence you will shortly see, as you see them. That is, I want you to press this button firmly whenever, in your judgment, such a meaningful action occurs. There are no right or wrong ways to do this; I simply want to know how you do it."

In the second condition, subjects were given the same instructions, except that the word *smallest* was substituted for *largest*.

At the midpoint of the sequence the tape was halted, and all subjects were instructed, "From now on, I want you to mark off whatever actions seem natural and meaningful to you."

Dependent measures. Data consisted of the number of breakpoints marked for each half of the videotape and responses to the postexperimental questionnaire items. The first item asked subjects to estimate the percent correct that the stimulus person would obtain on a second series of similar items. A second

question asked subjects to report the number of items the stimulus person had answered in the videotape. A third item asked subjects to recall the percentage of those items the stimulus person got right.

An additional comparison of the unit data was conducted to see if the predicted differences in level of analysis varied over a hierarchical structure. Frequencies of marking were tabulated for each 1-sec interval of the sequence. If fine-unit subjects are simply breaking down into smaller components the same actions that subjects using larger units are discriminating, then it follows that where the large-unit condition subjects indicate a breakpoint in the sequence, fine-unit subjects should agree, also indicating a breakpoint (note that the converse need not necessarily be true). Accordingly, consensus breakpoints in the large-unit condition were identified, and the frequency of identification of these same points as breakpoints in the fine-unit condition was compared to chance expectation by chi-squares (see Newtson, 1973). Statistical confirmation of the hypothesis would indicate a significant tendency toward a hierarchical relationship between the two segmentations.

Results

Subjects viewing the first part of the sequence under fine-unit instructions marked an average of 50.13 units, which was significantly more than the quantity subjects marked under large-unit instructions, $M = 18.88$, $t(14) = 5.83$, $p < .001$. Number of units did not differ significantly for the second half of the tape; means were 43.38 and 43.50 for the fine-unit/natural-unit and large-unit/natural-unit conditions respectively, $t(14) = .023$.

Subjects in the fine-unit/natural-unit condition estimated future performance at 35.75%, significantly worse than subjects in the large-unit/natural-unit condition, $M = 61.88$, $t(14) = 3.34$, $p < .01$. Results on recall of past performance were in the same direction, although this difference did not attain significance; means were 34.50 and 43.75, $t(14) = 1.69$, $p < .20$, two-tailed, for the fine-unit/natural-unit and large-unit/natural-unit conditions, respectively.

Comparison of segmentation patterns confirmed the hypothesis that fine-unit and large-unit segmentations would be more hierarchically related than chance, $\chi^2(1) = 9.46$, $p < .01$. That is, fine-unit subjects tended to break down the actions discriminated in the large-unit conditions into their component parts.

Discussion

These results confirmed that variation in level of analysis systematically altered the impact of the behavior on observer judgments. When the initial part of the series was analyzed finely, subjects predicted significantly poorer subsequent performance than when the first part of the series was analyzed in larger units. These results raise the possibility that biased information sampling could underlie primacy effects in ability attribution in the Jones et al. (1968) paradigm.

Evaluation of this hypothesis requires a comparison of the segmentations between the first and second halves of both an ascending and a descending success series. If the biased sampling interpretation is correct, then more breakpoints should be discriminated in the first halves of the sequences than in the second halves, and a typical primacy effect should result. That is, subjects viewing a series in which rate of success is low at the outset and then improves should estimate future performance as poorer than subjects viewing a series in which rate of success begins at a high level and then declines.

In addition, film speed was varied for the first halves of the sequences. The first half of each sequence was presented in slow motion or in fast motion. Our ability to comprehend variable speed films is an interesting phenomenon. Rate of movement in fast-motion and slow-motion films is typically altered by a factor of 7 to 10 times (cf. Miller & Strenge, 1969). This ability has simply not been explained. A careful search of the literature has failed to yield a single study systematically investigating the effects of film speed on the perception or interpretation of action. If, as suggested earlier, manipulations of film speed are similar in important ways to normal variations in level of perceptual analysis, variable speed film techniques may provide a useful convergent operation for the manipulation of level of perceptual analysis that is free of at least some of the possible demand characteristics associated with direct instructions. It would seem reasonable to expect that slow-motion film presentation would induce fine-unit analysis of a behavior sequence, whereas

fast-motion presentation would induce large-unit analysis.

Experiment 2

In this experiment subjects viewed one of two sequences showing an actor attempting to solve 30 problems. One sequence was the ascending success series employed in Experiment 1. A second sequence showing a descending success series was included in this experiment. The first half of each sequence was shown to subjects at slow motion, normal speed, or fast motion. In all conditions, the second halves of the sequences were shown at normal speed. It was anticipated that slow-motion presentation should induce finer unit analysis than normal speed presentation and that fast-motion presentation should induce larger unit analysis.

Results from the previous study suggest that the number of units marked for the first half of the series determines the impact of that part of the series in subsequent judgments. It would follow, therefore, that slow-motion presentation should enhance primacy effects as compared to normal-speed presentation, whereas fast-motion presentation should attenuate primacy effects. Alternately, the differences observed in Experiment 1 could have been unidirectional, such that only one of the instructional conditions affected judgments as compared to normal observation.

Method

Subjects. Seventy-two male undergraduates served as subjects in the experiment.

Stimuli. The study employed the ascending success sequence from the previous experiment and a second, descending success sequence. In this sequence, the actor received "right" feedback on 10 of the first 15 problems, and on 5 of the last 15. The precise pattern of feedback for this sequence was simply the reverse of the pattern for the ascending success series.

Three versions of each series comprised the stimuli for the experiment. One version simply consisted of the entire sequence of 30 problems at normal speed (normal-speed conditions). In a second version, the actor's performance on the first 15 items was copied at seven times normal speed, and the last 15 responses were recorded at normal speed (fast-motion conditions). In a third version, first-half performance was copied at one-seventh speed, whereas the second half was recorded at normal speed (slow-motion conditions).

This resulted in sequences of the following lengths: normal speed, 10.1 min.; slow motion, 42.4 min.; fast motion, 5.9 min.

Selection of these film speeds was dictated by available videotape machinery. Video recorders with continuously variable speed capacity also have "frame roll": each video "frame" is presented successively with a black bar separating them. The machine used for this experiment (a Javelin) is capable of presenting videotapes at 7:1 and 1:7 speeds without frame roll but is limited to these fixed presentation speeds.

Procedure. Subjects were assembled in groups of randomly varying size (1 to 15 persons) facing a 23-inch (58 cm) video monitor. Each subject was given a hand-held button, connected to an event recorder in another room. Subjects were informed that they would view a film of another undergraduate taking a computer-scored intelligence test. All subjects were then given the same segmentation instructions as in Experiment 1, except that subjects were instructed to discriminate the "naturally occurring" actions in the sequence.

Finally subjects were informed that sequences would be presented at varying speeds and were told the speed of the tape that they were about to see. This was done in order to prevent "surprise" effects on marking and to prepare subjects (especially in the fast-motion condition) for marking.

Following the viewing of the tape, subjects completed a questionnaire and were then debriefed.

Dependent measures. The same postexperimental questionnaire was used as in the previous study.

Design and analysis. Number of breakpoints marked for the sequences were analyzed in a $2 \times 2 \times 2$ mixed-effects analysis of variance in which there were two between-subjects factors, success pattern (ascending vs. descending) and speed of presentation (slow, normal, and fast motion) for the first half of the sequences, and a repeated measure film half (first vs. second).

Responses to postexperimental questionnaire items were analyzed in a 2×3 analysis of variance design in which factors were success pattern (ascending vs. descending) and speed of presentation (slow, normal, and fast motion).

Segmentation patterns for the first halves of the films were analyzed as in Experiment 1. This analysis is rather more important in the present study, as it bears on the question of whether or not the film speed manipulations simply disrupted perception of the sequences. Frequencies of marking were tabulated for each 1-sec interval, as before, and consensus breakpoints were identified for each sequence in each condition. Since presentation speed in the fast-motion condition was seven times that of normal-speed presentation, each 1-sec interval of the fast-motion record corresponded to seven 1-sec intervals in the normal-speed condition. Accordingly, co-occurrences of breakpoints were identified according to the following criterion: if an interval was identified as

Table 1
Mean Number of Breakpoints by Success Pattern, Speed of Presentation, and Film Half

Speed	Success pattern					
	Ascending		Descending		Combined average	
	First half	Second half	First half	Second half	First half	Second half
Slow	73.82	65.36	79.42	61.00	76.74_a	63.09_b
Normal	38.46	36.00	55.08	48.42	47.13_c	42.48_c
Fast	21.17	42.42	26.83	43.58	24.00_d	43.00_c

Note. Speed of presentation varied for the first film half only. Means with different subscripts differ by t test, $p < .001$.

breakpoint in the fast-motion condition, then the occurrence of a breakpoint in the normal-speed condition in any of the corresponding seven 1-sec intervals was counted as one, and only one, co-occurrence. Comparison of segmentations between the normal-speed and slow-motion conditions was performed in a completely analogous manner. Statistical confirmation of these hypotheses would indicate a significant tendency toward a hierarchical structure in the segmentation of the behavior.

Results

Unit data. Analysis of variance of unit data indicated a significant main effect for speed of presentation, $F(2, 64) = 23.46$, $p < .001$, and a significant Speed of Presentation × Film Half interaction, $F(2, 64) = 27.15$, $p < .001$. The main effect for success pattern, $F(1, 64) = 1.93$, $p < .20$, and the interaction between this factor and film half, $F(1, 64) = 2.81$, $p < .10$, approached significance. No other effects or interactions approached significance. Means are reported in Table 1. Interaction means are reported so as to facilitate comparison with Experiment 1, which employed the ascending success series only. The biased sampling hypothesis predicts that at normal speed more units should be marked for the first halves than the second halves. This hypothesis was not supported. When the sequence was presented at normal speed, unitization of the two film halves did not differ significantly, $t(64) = 1.45$, $p < .20$.

Under slow motion, more units were marked for the first than the second halves of the sequences, $t(64) = 4.24$, $p < .001$, which were at normal speed. When the first halves of the sequences were presented at fast motion, significantly fewer units were marked for the first halves of the sequences, $t(64) = 5.91$, $p < .001$. It was predicted that speed of presentation would alter level of unitization for the first halves of the sequences. Comparison of these means using t tests confirmed that more units were marked under slow motion than normal speed, $t(64) = 9.20$, $p < .001$, and more units were marked at normal speed than at fast motion, $t(64) = 7.91$, $p < .001$.

An unexpected result was obtained for the second half of the films in the slow-motion condition. Apparently because they had been induced to analyze the behavior finely in the first half by the slow-motion presentation, subjects persisted in finer unit analysis than that performed in the other condition (see Table 1).

Comparisons of segmentation patterns for the first halves of each sequence across film speeds indicated that the film speeds simply altered segmentations over a hierarchical structure. For the ascending series, 90% of the breakpoints marked under fast motion were also marked at normal speed, $\chi^2(1) = 153.31$, $p < .001$, whereas 45% of the breakpoints marked at normal speed were also marked at slow motion, $\chi^2(1) = 35.78$, $p < .001$. In the descending series, 60% of the breakpoints identified at fast motion were also breakpoints at normal speed, $\chi^2(1) = 63.68$, $p < .001$, whereas 64% of the breakpoints marked at normal speed were also breakpoints at slow motion, $\chi^2(1) = 56.43$, $p < .001$. This level of consensus is comparable to that observed for different levels of segmentation in previous studies (cf. Newtson, Enquist, & Bois, 1976). These results confirm that the film speed manipulations were not so

Table 2
Mean Estimates of Future Performance, Number of Problems Attempted, and Percent Correct by Speed of Presentation and Success Series

Success pattern and speed of presentation	Measure		
	Future performance (% correct)	% correct	No. attempted
Ascending			
Slow	35.97	32.27	23.82
Natural	46.36	39.09	27.91
Fast	67.08	60.00	27.33
Descending			
Slow	71.83	69.33	24.67
Natural	64.58	60.00	24.42
Fast	45.42	43.75	27.08

Note. Speed of presentation varied for the first film half only.

extreme as to disrupt perception of the sequences but that they did, as hypothesized, alter level of perceptual organization in a lawful manner.

Estimates of future performance. Analysis of variance of estimates of future performance indicated a significant main effect for success pattern, $F(1, 64) = 13.78$, $p < .001$, and a significant interaction between this factor and speed of presentation, $F(2, 64) = 34.11$, $p < .001$. The main effect of speed of presentation was not significant ($F < 1$). Means for this measure and two others are presented in Table 2. Comparison of the means in the natural-speed conditions across the ascending and descending success patterns indicated a strong primacy effect. Subjects viewing the ascending success series predicted significantly poorer subsequent performance than subjects viewing the descending success series, $t(64) = 3.61$, $p < .01$. This result replicates the findings of Jones et al. (1968).

It was predicted that slow-motion presentation of the first halves of the series would accentuate primacy effects. In the ascending series condition, slow-motion presentation resulted in significantly lower estimates of future performance than natural-speed presentation, $t(64) = 2.07$, $p < .05$, as expected. In the descending series condition, however, while slow-motion presentation tended to produce higher estimates of future performance, this difference was only of marginal significance, $t(64) = 1.44$, $p < .10$, one-tailed.

The second prediction, that fast motion should reduce primacy effects, was unequivocally supported. In the ascending series condition, fast-motion presentation resulted in significantly higher estimates of performance than natural-speed presentation, $t(64) = 4.10$, $p < .001$. In the descending series condition, fast-motion presentation resulted in significantly lower estimates of future performance, $t(64) = 3.79$, $p < .001$.

Recall of percent correct. A second questionnaire item asked subjects to recall the percentage of problems that the stimulus person had gotten right. Results on this measure generally paralleled those of the first but were slightly weaker statistically. A significant primacy effect was again obtained in the natural-speed conditions on this measure (see Table 2), with subjects recalling better performance in the descending series than in the ascending series conditions, $t(64) = 3.92$, $p < .001$.

As on the previous measure, differences between slow-motion presentation and natural speed were consistently in the predicted direction. In the ascending condition, the difference in recall of number correct did not differ significantly between the natural and slow-motion conditions, $t(64) = 1.28$, $p < .15$, one-tailed, although this difference did attain significance in the descending condition, $t(64) = 1.75$, $p < .05$, one-tailed.

Comparison of the fast-motion and natural-speed presentation conditions confirmed the predicted effect on this measure. Subjects in the natural-speed conditions recalled significantly lower performance than fast-motion condition subjects from the ascending series, $t(64) = 3.92$, $p < .001$, and recalled significantly higher performance at natural speed from the descending series than at fast motion, $t(64) = 3.04$, $p < .01$.

A third item requested subjects to recall the number of items attempted. Analysis of variance of these responses indicated no significant main effects or interactions. Means are reported in Table 2. Apparently, the ve

iations in film speed did not alter perception of problems attempted.

Discussion

Results did not support a simple biased sampling interpretation of primacy effects in ability attribution. Differences in number of units marked between sequence halves were small, despite a strong replication of the Jones et al. (1968) primacy effect on the cognitive measures. This failure to confirm the hypothesis is especially intriguing in view of the significant impact of the film speed manipulations on primacy effects in the second experiment, and the effects of level of analysis in Experiment 1.

Comparisons between the two experiments are also interesting in view of the tremendous differences in processing time between the two experiments. Subjects viewing the first half of the ascending success series under fine-unit instructions estimated future performance ($M = 35.75$) as comparable to those viewing that part of the sequence under slow motion ($M = 35.91$). Similarly, estimates of future performance resulting from large-unit instructions ($M = 61.88$) closely paralleled those obtained when the initial part of the sequence was presented in fast motion ($M = 67.08$). As the sequences were about 10 minutes long, the first half took 35 minutes to view at slow motion and about 43 seconds at fast motion.

Despite the large differences in presentation time, analysis of segmentation patterns confirmed that the fast-motion and slow-motion presentations did not destroy the intelligibility of the event. No differences were observed on estimates of the number of problems attempted, although it might be argued that subjects in all conditions inferred that the sequences were divided in half for presentation and simply doubled their estimates from the second halves of the films. More simply, however, inspection of the films themselves revealed no difficulty in following the events. These film speeds might be considered extreme. As indicated by their choice as a fixed speed by the manufacturer of the video recorder, however, these speeds are fairly typical for fast-motion and slow-motion film presentations.

The comparability of the instructional variations and film speed variations on the perceptual and cognitive measures does not indicate that there was some fortuitous "match" between the film speeds used and these instructional extremes. Rather, a hierarchy of structures exists in behavior, in this instance at a step-by-step level, at a trial-by-trial level and, as suggested by the natural-speed conditions, at at least one intermediate level. The variations in film speed apparently caused observers to select the level of analysis that was least difficult to employ under the circumstances.

With respect to the mechanisms of information integration underlying the cognitive data, the present findings are suggestive, if not conclusive. One possibility could be a simple kind of attention decrement or enhancement effect. That is, if fine units provide more information to the observer, subjects in the fine-unit/natural-unit and slow-motion/normal-speed conditions would have, in comparison to subjects in the normal-speed conditions, a sample of information from the behavior that is biased toward the first part of the sequence. For this to occur, however, it would also be necessary for perceivers to weight outcomes according to the number of actions associated with them. That is, one would have to assume that the experience of following someone through a step-by-step problem-solving attempt and then seeing him or her fail causes the outcome (i.e., the failure) to have more impact on the perceiver's judgment than if the perceiver viewed the event in large units of action (e.g., the actor tried and failed). The smaller magnitude of differences between the slow-motion/normal-speed conditions and the normal-speed conditions could thus be seen to follow from the fact that fine-unit analysis tended to persist in the slow-motion/normal-speed conditions. This interpretation could not, however, account for the occurrence of primacy effects in the normal-speed conditions, since segmentation did not differ across film halves in these conditions. Thus such an interpretation would have to assume that these effects occur in addition to primacy effects due to other mechanisms.

A second, perhaps more likely interpreta-

tion is that differences in level of analysis affect the point in the series at which a subjectively sufficient quantity of information is reached. That is, observers in these studies are simply told to view the series of events. If we assume, following Jones and Davis (1965), that perceivers take in information about others only until they feel they have "reason enough" to predict or explain the other's behavior, then it could be that fine-unit perceivers obtain sufficient information (subjectively) about the actor sooner in the series than large-unit perceivers. This could include other information about the actor than just ability information as reflected in outcomes of trials. Since the distribution of outcome information in the series is biased, the sooner the perceiver stops incorporating new information, the poorer (in an ascending success series) the impression of the actor's ability would be. This interpretation would have the advantage of being a more parsimonious one and might also account for the parallel results on both estimates of future performance and on recall of number of problems solved, in the absence of differences in recall of number of problems attempted. That is, although subjects continued to monitor the problem-solving behavior throughout the series, it would appear that they ceased to integrate outcome information into their impression of the actor's ability beyond a certain point.

It might be argued, however, that the results on the cognitive measures in both experiments are artifactual, in that there was a confound between the shift in content between film halves (the changing success rate) and the shift in instructions (Experiment 1) or in film speed (Experiment 2). That is, perhaps the changes in instruction or film speed caused subjects to conclude that the second half of the sequences was more or less important. A third experiment was thus conducted to rule out this interpretation.

Experiment 3

In this experiment subjects viewed either the ascending or descending success series. Subjects were instructed to analyze normal-speed films into fine units or large units or to segment fast-motion films naturally. The same segmentation instructions and film speed were maintained throughout the series. If the results on the cognitive measures in the first two experiments were due to an artifact of changes in film speeds or instructions, differences on these measures as a function of these manipulations should not be observed. Failure to find accentuation or attenuation of primacy effects as a result of these manipulations, however, could also be consistent with the interpretation of the previous results as due to enhancement of or decrements in attention. That is, if level of analysis simply affects the weighting of outcomes throughout the series, then differences in levels of analysis, if maintained throughout, should leave the overall judgment unchanged.

If, however, differences in level of analysis affect the point in the series where information integration ceases, fine-unit analysis should produce strong primacy effects, whereas large-unit analysis should attenuate them.

The present study did not include slow-motion conditions. Slow-motion presentation of these sequences would take more than 70 minutes. Our experience with the previous study suggested that slow-motion presentation of only the first halves of the sequences tested the limits of human endurance. Since the fast-motion conditions are adequate to test the alternative interpretations noted above, slow-motion conditions were excluded.

Method

Subjects. Subjects were 53 undergraduates (20 males, 33 females).

Stimuli. The same sequences were used that were employed in the previous study. Two versions of each sequence were used: (a) the normal-speed version of the ascending and descending series (length was 10.1 min.), and (b) a fast-motion version, in which the entire sequence was presented at seven times normal speed (length was 1.45 min.).

Procedure. Subjects in the fine-unit and large-unit conditions were instructed as in Experiment 1; subjects in the fast-motion conditions were instructed as in Experiment 2. One alteration in the procedure was made from the previous two studies. Subjects were asked to record units by making marks on a piece of paper, rather than by pressing a button as in the previous two studies. This modification of the recording procedure has been used by Wilder (1978a,

Table 3
Mean Number of Breakpoints, Estimates of Future Performance, Recall of Percent Correct, and Recall of Number of Problems Attempted by Condition and Success Series

Success pattern and condition	Measure			
	Number of breakpoints	Future performnace	% correct	No. attempted
Ascending				
Fine unit	104.70	35.50	33.50	27.30
Large unit	62.70	56.50	51.30	28.30
Fast motion	80.40	60.75	53.88	24.13
Descending				
Fine unit	147.22	82.56	68.89	25.00
Large unit	64.00	49.29	55.71	28.86
Fast motion	60.44	39.22	35.89	29.11

1978b) with good results and is more efficient if one is primarily concerned with the number, rather than the patterning, of breakpoints.

Dependent measures. The same postexperimental questionnaire was used as in the previous study.

Design and analysis. A one-way analysis of variance was conducted between the six groups (fine-unit, large-unit, and fast-motion conditions for the ascending and descending series) to obtain an estimate of variance, and t tests were conducted between groups on each of the variables.

Results

Results of the one-way analyses of variance indicated significant differences between groups in number of breakpoints marked, $F(5, 47) = 13.11$, $p < .001$; estimates of future performance, $F(5, 47) = 21.94$, $p < .001$; and recall of percent correct, $F(5, 47) = 9.05$, $p < .001$. As in the previous studies, no significant differences were obtained between groups on recall of number of problems attempted, $F(5, 47) = .40$.

Means for the six groups on all measures are reported in Table 3. Tests of mean differences indicated that for the ascending success series, significantly more breakpoints were marked in the fine-unit condition than in the large-unit condition, $t(47) = 3.33$, $p < .005$, or in the fast-motion condition, $t(47) = 1.85$, $p < .05$, one-tailed. Number of breakpoints marked did not differ between the large-unit and fast-motion conditions for the ascending series, $t(47) = 1.29$, $p < .20$.

Similarly, in the descending success series conditions, significantly more breakpoints were marked in the fine-unit condition than in the large-unit condition, $t(47) = 5.86$, $p < .001$, or in the fast-motion condition, $t(47) = 6.53$, $p < .001$. The latter two conditions did not differ in number of breakpoints marked, $t(47) = .25$, $p < .80$.

Results on estimates of future performance paralleled those obtained in the previous two studies, ruling out interpretation of those results as due to artifacts or attentional variations. In the ascending series, subjects in the fine-unit condition predicted significantly lower performance than subjects in the large-unit, $t(47) = 4.26$, $p < .001$, or in the fast-motion, $t(47) = 4.83$, $p < .001$, conditions. Fast-motion and large-unit condition means did not differ, $t(47) = .81$, $p < .40$.

Results on this measure in the descending series conditions indicated significantly higher estimates of future performance in the fine-unit condition than in the large-unit, $t(47) = 5.99$, $p < .001$, or in the fast-motion, $t(47) = 8.35$, $p < .001$, conditions. The difference between the large-unit and fast-motion conditions did not quite attain significance, $t(47) = 1.81$, $p < .10$.

The measure of recall of past performance produced results similar to those on the previous measure. In the ascending series conditions, mean recall of past performance was significantly lower in the fine-unit condition than in the large-unit condition, $t(47) = 2.98$, $p < .005$, or in the fast-motion condition, $t(47) = 3.22$, $p < .005$. Recall of past performance did not differ between the large-unit and fast-motion conditions, $t(47) = .41$, $p < .70$.

In the descending series conditions, recall of past performance was significantly higher in the fine-unit condition than in the large-unit, $t(47) = 1.96$, $p < .05$, or in the fast-motion, $t(47) = 5.25$, $p < .001$, conditions. The difference between the large-unit and fast-motion conditions, which was marginally significant on estimates of future performance, attained significance on this measure, $t(47) = 2.95$, $p < .005$.

Discussion

Since the same film speed and instructional conditions were maintained throughout the series in this experiment, subjects could not have responded to a demand that they weight the performance in the series unequally. This rules out a "confound" explanation of results on the cognitive measures in the first two experiments. Furthermore, results on number of breakpoints marked and estimates of future performance were comparable to those obtained in the first two experiments.

One problem in these data, however, is that variance in number of units marked was much larger than in the previous two experiments. This difference may reflect the difference between the button-press procedure used in Experiment 1 and 2 and Wilder's (1978a) pencil-and-paper procedure in this experiment.

These data also support the interpretation that level of analysis alters the cessation of information integration. Subjects analyzing the ascending series into fine units predicted future performance to be poorer than subjects analyzing the series into larger units, and subjects analyzing the descending series into fine units predicted future performance to be better than subjects analyzing that series into large units. The precise relationship between estimates of future performance and number of units, it should be noted, will depend upon a number of factors and is not a simple linear function of number of units. These factors include the underlying pattern of performance, level of analysis, and time of termination of processing. In the present studies, for example, performance did not steadily increase or decrease; the pattern of performance was more on the order of a step function, with very good performance followed by very poor performance in the "descending" series and the reverse pattern in the "ascending" series. Nevertheless, one might expect a correlation between number of breakpoints and estimates of future performance in the fine-unit and slow-motion conditions, since the point of sufficiency should occur closer to the midpoint of the series when level of analysis was very fine. These correlations were computed for each sequence for the fine-unit and slow-motion conditions combined over the three experiments. For the ascending series, number of breakpoints correlated $-.27$ ($df = 37$, $p < .05$) with estimates of future performance, whereas this value was .64 ($df = 30$, $p < .001$) for the descending series. That is, when performance was poor at outset, finer unit analysis led to lower predictions of future performance; when performance was good at the outset, finer unit analysis led to higher predictions of future performance. Similar correlations over the larger unit conditions were not significant.

There was one result in Experiment 3 that was not consistent with the previous experiment. Despite the comparable number of units marked for the descending series in large-unit and fast-motion conditions, estimates of future performance and recall of percent correct tended to be significantly lower in the fast-motion conditions. It is possible that the novelty of the fast-motion presentation raised the criterion of sufficiency in this condition, although it is not clear why this was not equally the case for the ascending series.

General Discussion

Jones and Goethals (1971) identified three types of processes that can produce primacy effects: (a) discounting processes, whereby later information is rejected if it is inconsistent with earlier information; (b) assimilation processes, whereby the interpretation of later information is altered in light of previous information; and (c) attention decrements, whereby the perceiver's sample of information is biased in that it is largely drawn from the first part of the sequence.

The explanation of primacy effects that we

have proposed shares important features with each of these explanations. That is, once a point of sufficient information is reached, later information is rejected, consistent with a discounting interpretation, although this rejection is not seen as a result of its consistency with previous information. Jones and Goethals (1971) reject the discounting interpretation, noting that "it is difficult to see how the discounting hypothesis might apply in performance evaluation when there is no evidence for motivation decrement or variations in problem difficulty" (p. 38). Our suggestion would be that later evidence in a series may be "discounted," but only in the sense that it is not used, rather than being rejected.

Similarly, once a point of subjectively sufficient information is reached, both the "sufficiency" and assimilation interpretations make the same prediction. The assimilation hypothesis assumes that later information is distorted; the sufficiency hypothesis assumes it is disregarded. Thus the sufficiency explanation is proposed as an alternative view, but the present data cannot discriminate between them.

Finally, the sufficiency explanation incorporates some features of the attention decrement interpretation in that it postulates that such effects are due to a decrease in processing effort over time. The results of Experiment 3, as argued, rule out a simple attentional weighting interpretation of these effects.

One question raised by the present data is the reason that a "recency effect" was obtained in the fast-motion and large-unit conditions. That is, if subjects were maintaining an accurate count of successes, one might anticipate convergence of the large-unit conditions to an accurate estimate of a 50% success rate. This prediction, however, would require a number of assumptions as to the nature of processing that are, so far at least, unverified. One assumption would be that if successes were randomly arranged, estimates of success would be accurate. Jones et al. (1968) used randomly arranged sequences and found that estimates of future performance ranged as high as 61% in some series. A second assumption is that subjects do indeed keep running counts of total number of successes rather than more impressionistic estimates of success rates based on the last few observed outcomes. If the latter were the case, it should be noted, the judgment of success rate would simply reflect the estimate that was current when the point of sufficiency was reached.

Another question raised by these results concerns the relationship between level of analysis and information gain. Newtson (1973) argued that level of perceptual analysis determines potential information gain from observation. In the large-unit and fast-motion conditions of the present studies, more units were marked than there were outcomes in each of the conditions, so that all subjects should have had complete information on outcomes. The sufficiency explanation proposed here postulates that subjects reached their point of sufficiency because they picked up additional information as well.

Newtson (1973), in the initial study of level of analysis and information gain, found that observers analyzing a sequence into fine units of action had more confident and differentiated impressions of the actor. Subsequent investigations (cf. Deaux & Major, 1977; Newtson, 1976; Wilder, 1978b; Ebbesen, Cohen, & Lane, Note 1) have indicated that this effect may often be obtained for particular sequences, but sometimes is not found. No studies, to our knowledge, have reversed the effect such that greater information is gained under large-unit than under fine-unit analysis.

Newtson (1973) based his argument on an analysis of perceived actions in terms of information theory. Strictly defined in those terms, the notion that more units should provide greater *potential* information is a valid one. As Garner (1962) points out, however, there is a difference between the quantity of information that is potentially available in a series of instances and the information that is transmitted by those instances. That is, humans are able to use information only when it is available in particular forms or structures. Newtson's (1973) prediction that finer unit analysis should produce more information *gain* than large-unit analysis is like the prediction that concept attainment is a simple function of the number of instances presented.

Such a generalization in concept attainment paradigms is a highly limited one (cf. Bourne, 1966; Bruner, Goodnow, & Austin, 1956; Garner, 1962), restricted to sets of instances providing usable information in small increments. Newtson's (1973) prediction thus depends upon the further assumption that the actions discriminated provide information about the actor in a usable form. The present data further qualify this relationship. Level of analysis may be seen as setting the limits of information gain, but further processing of that information may depend upon other factors.

Reference Note

1. Ebbesen, E. B., Cohen, C. E., & Lane, J. L. *Encoding and construction processes in person perception*. Paper presented at the meeting of the American Psychological Association, Chicago, 1975.

References

Bourne, L. E. *Human conceptual behavior*. Boston: Allyn & Bacon, 1966.

Bruner, J. S., Goodnow, J. J., & Austin, G. A. *A study of thinking*. New York: Wiley, 1956.

Deaux, K., & Major, B. Sex-related patterns in the unit of perception. *Personality and Social Psychology Bulletin*, 1977, *3*, 297–300.

Garner, W. R. *Uncertainty and structure as psychological concepts*. New York: Wiley, 1962.

Jones, E. E., & Davis, K. E. From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2). New York: Academic Press, 1965.

Jones, E. E., & Goethals, G. R. Order effects in impression formation: Attribution context and the nature of the entity. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior*. Morristown, N.J.: General Learning Press, 1971.

Jones, E. E., Rock, L., Shaver, K. G., Goethals, G. R., & Ward, L. M. Pattern of performance and ability attribution: An unexpected primacy effect. *Journal of Personality and Social Psychology*, 1968, *10*, 317–340.

Miller, A. C., & Strenge, W. *American cinematographer manual* (3rd ed.). Hollywood, Calif.: American Society of Cinematographers, 1969.

Newtson, D. Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology*, 1973, *28*, 28–38.

Newtson, D. Foundations of attribution: The unit of perception of ongoing behavior. In J. Harvey, W. Ickes, & R. Kidd (Eds.), *New directions in attribution research*. Hillsdale, N.J.: Erlbaum, 1976.

Newtson, D., & Engquist, G. The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology*, 1976, *12*, 436–450.

Newtson, D., Engquist, G., & Bois, J. The reliability of a measure of behavior perception. *JSAS Catalog of Selected Documents in Psychology*, 1976, *6*, 5. (Ms. 1173)

Newtson, D., Engquist, G., & Bois, J. The objective basis of behavior units. *Journal of Personality and Social Psychology*, 1977, *35*, 847–862.

Newtson, D., Rindner, R., Miller, R., & LaCross, K. Effects of availability of feature changes on behavior segmentation. *Journal of Experimental Social Psychology*, 1978, *14*, 379–388.

Wilder, D. Effects of predictability on units of perception and attribution. *Personality and Social Psychology Bulletin*, 1978, *4*, 281–284. (a)

Wilder, D. Predictability of behaviors, goals, and unit of perception. *Personality and Social Psychology Bulletin*, 1978, *4*, 604–607. (b)

Zettl, H. *Sight-sound-motion: Applied media aesthetics*. Belmont, Calif.: Wadsworth Publishing Co., 1973.

Received December 5, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 10, 1859-1870

Reassertion and Giving Up: The Interactive Role of Self-Directed Attention and Outcome Expectancy

Charles S. Carver and Paul H. Blaney
University of Miami

Michael F. Scheier
Carnegie-Mellon University

Two experiments tested a theoretical model of behavioral self-regulation, which makes predictions about the effects of failure on a person's subsequent efforts. This model holds that degree of effort will be a product of two things: expectancy of being able to redress the failure and degree of self-attention. In the experiments, a failure pretreatment was used to create large within-self discrepancies among female subjects. It was predicted (a) that negative outcome expectancies regarding a subsequent task would lead to decreased persistence on that task, (b) that positive outcome expectancies for the subsequent task would lead to increased persistence on that task, and (c) that both of these tendencies would be mediated by self-directed attention. The results of the two studies supported these predictions. Discussion centers on implications for research and theory in the areas of self-awareness, self-efficacy, and helplessness.

A good deal of attention has been devoted in recent years to the consequences that an initial failure can have for subsequent task performances. Sometimes the inability to achieve an expected goal at one task has led to performance increments at a second task (e.g., Hanusa & Schulz, 1977; Roth & Kubal, 1975). This typically has been interpreted as reflecting heightened efforts to compensate for the prior inadequacy. In other circumstances, however, failure at one task has led to performance decrements at a second task (e.g., Hiroto & Seligman, 1975; Seligman, 1975). This "giving up" phenomenon is widely believed to reflect a belief that effort on the second task will not lead to a positive outcome.

The authors are grateful to Denise Brandom, Marcia Hamilton, Ileana Rodriguez, and particularly Jody Kreitchman, for their efforts in the collection of the data. The results of this research were presented in an abbreviated form at the 1979 meeting of the American Psychological Association, New York City.

Requests for reprints should be sent to Charles S. Carver, Department of Psychology, P.O. Box 248185, University of Miami, Coral Gables, Florida 33124.

A theoretical model of behavioral self-regulation has recently been proposed (Carver, 1979; Carver, Blaney, & Scheier, 1979) that has important implications for our understanding of performance increments and decrements as products of failure. This theoretical model was developed in a somewhat different context, however. Specifically, it was designed to describe the behavioral consequences of self-directed attention.

Self-Attention and Behavioral Regulation

Self-focused attention has been shown to have several predictable effects, two of which are directly relevant to present concerns. These two effects both occur when some behavioral standard has been made salient to the person prior to the heightening of self-focus. In such a case, self-directed attention is believed to cause the person to become more aware of his or her present state or behavior and how it compares to the standard. One frequently demonstrated consequence of this awareness is an enhanced tendency for self-attentive persons to alter their behavior so that it conforms more closely to the standard of comparison (e.g., Beaman, Klentz,

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3710-1859$00.75

Diener, & Svanum, 1979; Carver, 1974, 1975; Froming, in press; Scheier, Fenigstein, & Buss, 1974; Vallacher & Solodky, 1979). On the other hand, in at least one study (Duval, Wicklund, & Fine, 1972), persons in whom a large self versus standard discrepancy had been created simply withdrew more quickly from the experimental context when self-focus was high than when it was lower.

These two divergent effects of self-attention —enhanced conformity to behavioral standards, and withdrawal—are reconciled in the following way (Carver, 1979). When a discrepancy exists between a present state or behavior and a salient standard of comparison, self-focus leads initially to an impetus to match one's behavior to the standard. This is presumed to be the case until or unless something interrupts this impetus. Interruption can occur prior to initiating the attempts to alter one's behavior if, for example, the person knows ahead of time that reducing the discrepancy between self and standard will be difficult. Alternatively, interruption can occur during the discrepancy-reduction attempt itself. Any interruption of the discrepancy-reduction impetus leads to an assessment of outcome expectancy, that is, the perceived likelihood that the discrepancy can be reduced. This expectancy judgment repreents a kind of psychological watershed. Subsequent behavior can be seen as falling into one of two classes, depending on the outcome of this judgment. If the expectancy is favorable, the result is reassertion: further effort to reduce the discrepancy, to match one's behavior or state to the standard. This tendency should be exaggerated by further self-awareness. If the expectancy is unfavorable, the result is an impetus to withdraw. The execution of the withdrawal impetus should also be enhanced by further self-focus.

Carver et al. (1979) have applied this reasoning to analyzing the approach–withdrawal decision that may occur among fearful persons in response to their recognition of rising fear. Carver et al. tested subjects who were equivalent in their degree of self-rated fear of a specific stimulus but who differed from each other in their chronic expectancies of being able to approach and handle the feared stimulus. Experimentally enhanced self-attention led the doubtful subjects to withdraw at an earlier stage of the approach sequence than was true with less self-attention. On the other hand, self-attention tended to enhance approach among the more confident subjects, though not significantly so. These findings thus provided substantial support for the theoretical model, at least as it applies to fear-based behavior.

Failure, Expectancy, and Self-Attention

This theoretical model has a great many potential applications, however, of which the fear-response decision is only one. It also makes predictions, for example, about responses to failure. Failure produces a large discrepancy between one's present state and one's desired state (i.e., a discrepancy between oneself and the standard of comparison). According to our model, one's response to failure should be dictated in part by one's expectancy about being able to reduce this discrepancy. A positive expectancy—the perception that the gap can be closed successfully—should lead to reassertion attempts. This should be reflected by increased efforts to do well on a subsequent task. A negative expectancy—the perception that the discrepancy cannot be altered—should lead to an impetus to withdraw. This should be reflected by reduced efforts at a subsequent task. Moreover, both the reassertion and the withdrawal tendencies should be exaggerated by increased self-awareness.

Although there is some indirect support for this reasoning (McDonald, in press; Steenbarger & Aderman, in press), the model has not yet been tested directly in this type of behavioral context. To do so was the purpose of the present research. All participants in this research experienced a failure, which created a substantial discrepancy between their present states and their desired states. We then manipulated subjects' expectancies about whether or not they would be able to do well at a subsequent task. The subsequent task was chosen to provide us with a measure of subjects' persistence. It was predicted that subjects in whom positive expectancies had been created would display more persistence with high than with low self-attention. And

it was predicted that subjects in whom negative expectancies had been created would display less persistence with high than with low self-attention.

Experiment 1

Method

Subjects and Experimenters

Subjects were 74 female undergraduates from the University of Miami subject pool. We used only females in this experiment because our experimenters were female and we wished to avoid introducing the potential psychological complexity associated with mixed-sex dyads. Eight other subjects were deleted prior to data analysis, for the following reasons: 1 reported having succeeded on the initial (failure manipulation) task, 3 spontaneously reported the belief that neither task was possible, and 4 either failed to understand the instructions for the second task or else failed to follow those instructions, which resulted in their claiming to have solved an insoluble puzzle. Each subject, tested in an individual session, was randomly assigned to one of two expectancy conditions and to one of two self-attention conditions (described more fully below).

There were two experimenters; each interacted alone with the subject at a different stage of the experiment. The first experimenter was responsible for manipulating the subject's outcome expectancy but remained blind to the subject's self-awareness–condition assignment. The second experimenter was responsible for manipulating self-focus but remained blind to each subject's expectancy–condition assignment. Both experimenters were ignorant of the specific hypotheses being tested.

Procedure

The first experimenter escorted the subject to the first testing room. There she explained that the research team was evaluating a set of perceptual motor tasks for eventual use as an instrument for measuring abstract reasoning skills. It was stressed that the specific skills that these items assessed were very important in a number of professional and interpersonal domains. It was further explained that a wide variety of different test-item groups were being developed by the research team, in order to make possible measurement of these abilities from several quite different perspectives. The subject was led to believe that 12 different sets of items were being tested, although she herself would only be asked to work on 2 of the sets. Each participant in the project ostensibly was being assigned 2 randomly chosen item sets, in order to control for fatigue and order effects. In reality, all subjects were asked to attempt the same 2 tasks in the same order.

The experimenter continued by saying that some of the problem sets had been tested in previous semesters. In such cases, any information obtained from the previous research would be printed along with the instructions for that task. The experimenter concluded her preliminary remarks by saying that the subject would be asked occasionally to fill out some brief self-report scales. The subject was then given a questionnaire, which included an expectancy rating for the first task: "How well do you expect to do on your first task?" answered on a 9-point scale labeled *extremely poorly* at one end and *extremely well* at the other end.

Anagram task. Each subject then was given a sheet labeled "anagrams task." The sheet included printed instructions explaining that the task was to rearrange the letters of each item to form a word. Work space was provided, but the subject was cautioned not to guess randomly in the answer blanks, because a proportion of her incorrect answers would be subtracted from her total number of correct answers.

This sheet also included a paragraph indicating that the anagrams had been partially evaluated during three previous semesters. Scores on the anagrams task ostensibly had been found to predict scores on the "design problems" task and the "digit-symbol" task but not scores on the "mazes" task. This information was included to provide some validation in the subject's eyes for the expectancy manipulation that took place later (see below). This paragraph also indicated that University of Miami students over the previous three semesters had ostensibly averaged 7.21 correct solutions, with last semester's average being 7.34. This information was included in order to ensure that all subjects would view their own performances as failures. In actuality, the anagrams, chosen from Feather and Simon (1971) and Tresselt and Mayzner (1966), were quite difficult and in some cases impossible (overall, subjects' scores averaged .77, $SD = .93$).

The experimenter told the subject that she would have 5 minutes to work on the anagrams, "although I don't suppose you'll need all that time for just 9 items." The experimenter said she would be on the other side of the room working if the subject had any questions. After 5 minutes had elapsed, the experimenter indicated that the time was up and returned to where the subject was seated. She looked at the subject's performance with mild surprise and said, "That's not very good at all. Almost everyone gets at least five or so on this task." The experimenter then delivered the statement that constituted the manipulation of outcome expectancy.

Expectancy manipulation. In the positive expectancy condition, the experimenter continued as follows:

> What did I say was going to be your other task, the design problems? [After checking her list] Yes. Well, that's a good break for you. You see, the anagrams and design problems have both been

studied quite a bit already, so we know a lot about how they go together. The evidence so far is that the two are kind of opposites of each other. They both measure the same abilities, but in such different ways that people who do poorly on this one often seem to do much better on that one, and vice versa. Probably you'll be able to do better on the design problems.

In the negative expectancy condition, the experimenter said instead:

> What did I say was going to be your other task, the design problems? [After checking her list] Yes. Well, that's unfortunate for you. You see, the anagrams and the design problems have both been studied quite a bit already, so we know a lot about how they go together. The evidence so far suggests that the two correlate very very highly with each other, which means a person who does well on one will also do well on the other—if a person does badly on one, that person will do badly on the other one, too. So I'd expect you're not going to do very well on the design problems.

The experimenter went on to say that she was working only with the anagram task and that another person was handling the design problems in a different room. The experimenter said that the person supervising the design problems would deal with the subject from that point on, which would include seeing that the subject received credit for participation. The implication of the latter information was that the subject would have no further contact with the first experimenter.

The first experimenter then escorted the subject to a second experimental room, knocked on the door, introduced the subject to the second experimenter, and said that the subject was to work on the design problems next. The second experimenter guided the subject to a seat that faced the experimenter's desk and closed the door. She recorded on a form that the subject was undertaking the design problems as her second task, asked the subject what her first task had been, and recorded the subject's response. She then handed the subject a printed form that was labeled "instructions for design problems," and asked the subject to read these instructions carefully.

Design problem. The design problem used in the study was patterned after those developed by Feather (1961), which have been used a number of times subsequently to measure persistence (e.g., Glass & Singer, 1972). The goal of the task is to trace over all the lines of a geometric figure, according to two rules: first, the line must be continuous —the pen cannot be lifted from the page; second, whereas it is permissible to cross a line traced previously, each line segment must be traced once and only once. The instruction sheet included a simplified sample problem with illustrations of a correct solution and solution attempts that violated each of the two rules.

The instruction sheet continued by saying that the subject would be given four separate design problems. She was to work on the first one either until she solved it or until she decided to go on to the next one. Once she went ahead, she could not go back to prior puzzles. The subject was allowed to attempt to trace the design as many times as she wished (on separate pieces of tracing paper), provided that a previous attempt was discarded before a new attempt was begun (a wastebasket was provided for this purpose). It was implicit that neither the time taken to reach a solution nor the number of attempts required to reach a solution mattered in the scoring—only whether the problem was solved or not.

After briefly recapitulating the instructions, the second experimenter asked the subject to seat herself at a nearby table, on which were a felt-tipped pen and a stack of tracing paper. Also on the table were several pieces of nondescript equipment. The experimenter indicated that she herself was just going to keep working at her own desk while the subject worked at the design problems. The experimenter gestured at the clutter on the subject's desk, and said, "Don't pay any attention to this junk over here; it's part of another experiment." She then asked the subject to fill out a second self-report blank, which included a rating of how well the subject had done on her first task and a rating of how well the subject expected to do on the design problems (both on 9-point scales). The experimenter then placed the first design problem in front of the subject, saying, "Here's the first puzzle. By the way, this first one is the hardest one. Hardly anyone gets it, so don't be too surprised if you don't.[1] You can go ahead and start; just let me know when you want the next one."

The experimenter then turned and walked back to her own desk. As she did so, she unobtrusively looked at a clock on her desk, which had been shielded from the subject's view. Using that clock, she began timing the subject's attempt on the first design problem. In reality, the subject was given only one problem, for which there was no solution. The amount of time that the subject spent attempting that problem was the study's dependent variable.

Self-attention manipulation. Self-directed attention was varied during the second task by the presence or absence of a wall mirror (.75m × .50m) resting on the table at which the subject was working. (Evidence of the self-focusing properties of mirror presence is discussed in detail by Carver & Scheier, 1978.) In the mirror present condition, the mirror was situated directly in front of the subject's chair, leaning against the wall at an angle at which the subject could view her own reflection but not that of the experimenter. In the mirror

[1] This statement was included because pilot data had indicated that most subjects were highly persistent on the design problem in the absence of such a statement. This comment had the effect of reducing subjects' persistence slightly overall.

absent condition, the mirror's reflective surface was turned toward the wall. In both conditions, the mirror was implicitly included in the experimenter's reference to the "junk" that was part of another experiment. The instructions for the task had been given before seating the subject at the table so that there would be relatively little time for her to habituate to the mirror's presence before beginning the experimental task.

When the subject asked the experimenter for the second design problem, or after 10 minutes (whichever came sooner), the experimenter recorded the elapsed time, gave the subject a final brief questionnaire, and began to probe for signs of suspicion. She then took the subject to a third experimental room, where the general purpose of the study was explained. After being assured that her poor performances on the two tasks had been manipulated and were not her own fault, the subject was given credit for her participation and was dismissed.

Results

Manipulation Checks

It was important for an adequate test of our predictions that all subjects perceive a within-self discrepancy and that the positive and negative expectancy groups differ from each other in their expectancies of being able to reduce the discrepancy. To verify that this was the case, each subject was asked to complete two self-report items that served as checks on the failure and expectancy manipulations, respectively, just before beginning the design problems task.

The first item asked the subject to rate her performance on the first task (anagrams) by circling a number on a 9-point scale with endpoints labeled *very successful* and *very unsuccessful*. All subjects included in the data analysis circled numbers on the "unsuccessful" side of the neutral point (overall $M = 1.84$, $SD = 1.19$). Thus every subject viewed her performance on the first task as a failure. Just as important, a 2×2 factorial analysis of variance of subjects' responses to this item revealed that the experimental groups did not differ from each other in the degree to which they perceived the outcome of the first task as having been unsuccessful (all Fs < 1). Nor, in fact, had the groups differed reliably in actual performances on the anagrams task (Fs < 1.5) or in the expectations for success that they had expressed with regard to the anagrams task (Fs < 1).

On the other hand, subjects' ratings of how well they expected to do on the second task—the design problems—did differ reliably between expectancy conditions, $F(1, 70) = 5.45$, $p < .03$. Moreover, no effect other than this predicted main effect approached significance (other Fs < 1). Although the expectancy conditions differed from each other as had been anticipated, closer examination of the means revealed that the two expectancy manipulations had apparently not been equally effective. That is, the average rating among subjects in the negative expectancy condition was on the appropriate side of the neutral point of the scale ($M = 4.16$, $SD = 1.53$), but the average among subjects in the positive expectancy condition was almost exactly at the scale's neutral point ($M = 4.92$, $SD = 1.23$). We will return to this point below.

Persistence

The dependent measure of greatest interest was the amount of time that subjects had spent attempting to solve the design problem. An analysis of variance of these values (see Table 1) revealed a significant interaction between expectancy condition and self-awareness condition, $F(1, 70) = 4.87$, $p < .03$, as well as main effects approaching significance for both variables (ps $< .10$). Individual group comparisons revealed that all of these statistical effects depended primarily on the low persistence values among subjects who had been provided with poor outcome expectancies and for whom self-attention had been enhanced. These people withdrew from their attempts to solve the design problem more quickly than did either subjects with negative expectancies who were less self-attentive or subjects in whom self-attention had been enhanced but who had more positive expectancies, ps $< .05$ by Scheffé test. Contrary to our hypothesis, self-focus in the positive expectancy condition did not increase subjects' persistence. Nor, somewhat surprisingly, did negative outcome expectancy reliably reduce persistence in the absence of self-focus.[2]

[2] A separate analysis was also performed on the number of times the subject attempted to solve the problem, though prediction on that measure was

Table 1
Persistence (in minutes) at an Insoluble Task, as a Function of Outcome Expectancy and Self-Attention, Experiment 1

	Mirror absent			Mirror present		
Expectancy	Persistence	*SD*	*n*	Persistence	*SD*	*n*
Positive	8.15	2.38	17	8.43	2.20	19
Negative	8.51	1.88	17	6.55	2.21	21

Discussion

Experiment 1 provided clear evidence that self-focus enhances the tendency of subjects to withdraw from an experimental task when their outcome expectancies are negative. Self-directed attention did not cause withdrawal among subjects with more positive expectancies; however, neither did self-focus cause these subjects to display enhanced persistence on the second task as had been predicted.

There are several possible reasons for the failure to confirm this reassertion prediction. The most obvious possibility is that exaggerated persistence was prevented by our arbitrary imposition of a 10-minute time limit on the second task. The persistence means of three of the four groups were, in fact, relatively near this artificial "ceiling" on effort. Another possibility is suggested by the finding that subjects in the positive expectancy condition did not really have particularly positive expectancies. Although we can do little more at this stage than speculate as to where the expectancy "watershed" point actually is, it may be that these persons' outcome perceptions were not sufficiently positive to cause a discernible increase in reassertion under self-focus. A third possibility, of course, is that self-attention does not lead to reassertion under conditions of positive expectancy.

Because the reassertion prediction is an important part of our theoretical model, and because the results of at least one previous study (Carver et al., 1979) have been somewhat equivocal on the reassertion issue, we conducted a second experiment. In this second study an attempt was made to create slightly more positive expectancies in subjects, by means of slight changes in the expectancy manipulation. In addition, we minimized the effects of a potential ceiling on subjects' persistence by increasing the time limit. In Experiment 2, each subject was allowed up to 20 minutes to work on what she believed was the first of four design problems.

Experiment 2

Method

Subjects were 34 female undergraduates from the University of Miami subject pool. Two additional subjects were deleted prior to data analysis. One claimed to have solved the insoluble puzzle, and the other voiced suspicions about the experimental tasks.

All procedures in this study exactly duplicated those of the positive expectancy condition of Experiment 1, with two exceptions. First, the last portion of the positive expectancy manipulation was changed and expanded slightly, as follows: "They both measure the same ability, but in such different ways that people who do poorly on the anagrams often seem to do much better on the design problems, and vice versa. It seems funny, but almost always when people do badly on this one, they seem to do really well on that one. So from past experience I expect you'll do well on the design problems." The second procedural difference was that, as noted above, subjects were allowed up to 20 minutes to work on the "first" design problem. As in Experiment 1, subjects were randomly assigned to either the mirror present or the mirror absent condition.

Results and Discussion

As in Experiment 1, preliminary analyses revealed that the groups did not differ from

more ambiguous. That is, increased effort could result in more frequent attempts; alternatively, it could lead only to more concentration prior to making an attempt. Analysis of this measure yielded only a nonsignificant tendency toward more attempts in the mirror absent than the mirror present condition.

each other in first-task expectancies, number of anagrams solved ($M = .32$, $SD = .47$), ratings of their anagram performances ($M = 1.68$, $SD = 1.27$), or their expectancies with regard to the second task (all $Fs < 1$). (It may be noted that subjects' expectancies for the second task were slightly more positive overall than those of the comparable subjects in the first experiment, $M = 5.06$, $SD = 1.54$.)

Analysis of subjects' persistence on the second task revealed greater persistence among subjects in the mirror present condition ($M = 12.47$, $SD = 5.58$) than among subjects in the mirror absent condition ($M = 8.15$, $SD = 4.63$), $F(1, 32) = 6.04$, $p < .02$.[3] This difference proved to be even more reliable ($p < .008$) when a comparison between groups was made that included data from only those subjects who reported either neutral or positive second-task expectancies. Thus our hypothesis was confirmed. When one's outcome expectancy is positive, self-focus leads to enhanced efforts.

One final point should be made about these data. Subjects' persistence in the mirror absent condition of Experiment 2 was virtually identical to that displayed by the comparable subjects (positive expectancy – mirror absent) in Experiment 1. More to the point, neither of these values differed from the persistence displayed by subjects with negative expectancies in the mirror absent condition of Experiment 1. This absence of an expectancy effect in conditions of low self-focus is consistent with previous findings (Carver & Blaney, 1977a, Experiment 3; 1977b; Carver et al., 1979). Our explanation for this finding is straightforward. In the absence of an experimental manipulation, self-focus among our subjects may have been quite low. Our theoretical model describes effects that occur under conditions of relatively high self-attention. If self-focus were low enough, the postulated reassertion/withdrawal dichotomy would not have been evoked. Thus the absence of a main effect for expectancy does not appear to pose any problem for our theoretical analysis. Indeed, our data simply suggest that expectancy effects require some self-attention before they occur at all.

General Discussion

The research presented here conceptually replicates a previous finding (Carver et al., 1979), extending that finding in three important respects. First, Carver et al. *measured* outcome expectancy rather than *manipulating* it. Thus, although their results were consistent with the expectancy-based analysis, the findings were susceptible to the criticism that it could have been another variable, correlated with expectancy, that had really caused the effect. In the present research, in contrast, we experimentally manipulated outcome expectancy, thus minimizing that potential criticism.

A second way in which these findings add strength to those of Carver et al. (1979) concerns the degree to which reassertion can be said to have occurred when a positive expectancy was combined with self-directed attention. The previous study had found only directional support for that hypothesis (as was also the case in the present Experiment 1). Presumably this was because of a ceiling on subjects' efforts that was imposed by the nature of the task being attempted. In contrast, the present Experiment 2 clearly demonstrated such a reassertion response.

A third aspect of this replication may be most important of all. Specifically, we have demonstrated here that the same principles that allowed us in the past to predict fear-based behavior can be successfully applied to a fundamentally different behavioral domain. That is, the present predictions were based on precisely the same theoretical model and used exactly the same logic that led to the previous predictions of Carver et al. However, the logic in this case was applied not to the attempt to approach a feared stimulus, but rather to the persistence with which a person tries to reduce a within-self discrepancy caused by a prior failure. The fact that the theoretical analysis is so easily applied to these divergent areas of behavior suggests that the model may be very general indeed.

---

[3] Analysis of the numbers of times subjects attempted the problem indicated that there were marginally more attempts ($p < .08$) among mirror present than among mirror absent subjects.

There is one potential concern about these findings, however, which stems from the fact that a mirror was used to vary self-focus both in the present studies and in our previous one (Carver et al., 1979). Though we have argued elsewhere (Carver & Scheier, 1978; Carver & Scheier, Note 1) that the mirror is in many ways the "purest" way to increase self-attention experimentally, it does have one important side effect: It provides subjects with visual access to their facial expressions. With regard to the present research, this raises the possibility that subjects seated before a mirror became especially aware of their "hopefulness" (in the positive expectancy condition) or their "hopelessness" (in the negative expectancy condition) by observing their facial expression, inferred their internal states through this external mediation, and then behaved in a manner consistent with that inference (see, for example, Bem, 1972).

There are two sources of evidence against this line of reasoning: First, evidence converging on the validity of the self-attention construct has been gathered in a good deal of previous research (e.g., Carver & Scheier, 1978; Carver & Scheier, in press; Scheier & Carver, 1977; Scheier, Carver, & Gibbons, 1979; Scheier, Carver, & Gibbons, in press). The strategy in that research was to replicate mirror-induced effects using individual differences in the disposition to be self-attentive, which is termed self-consciousness (Fenigstein, Scheier, & Buss, 1975). Dispositional self-consciousness is a particularly useful tool for this purpose, in that its use appears to rule out possibilities such as external mediation of any obtained effects (see also Carver & Scheier, 1978; Scheier et al., 1979). That is, being a dispositional variable, it does not require the presence of any external stimulus to mediate its effects. By implication, this evidence suggests that mirror effects are also mediated by inward focus of attention.

The second source of evidence is a very recent study by Brockner (1979), examining a phenomenon that is conceptually similar to that studied here. Brockner was interested in the effects of self-focus on persons high and low in self-esteem, following success and failure experiences. He examined these effects by measuring dispositional self-consciousness as well as by manipulating mirror presence. Brockner found that in the absence of the mirror, dispositional self-consciousness led to precisely the same effects as had mirror presence. That is, among highly self-conscious subjects, failure led to a subsequent performance decrement for those with low self-esteem but not for those with high self-esteem. In contrast, persons who differed from each other in self-esteem, but who were low in self-consciousness, did not differ from each other behaviorally following failure. Brockner's results thus show an important convergence with the present data, which suggests in turn that the present effects were internally mediated.

Two differences between Brockner's research and our own should be noted briefly, both of which appear to add further generality to our theoretical stance. First, Brockner's research investigated the phenomenon of low self-esteem, rather than the giving-up response per se. The fact that his findings are predictable from our reasoning (see Brockner, 1979, for greater detail) provides one more indication that phenomena as diverse as anxiety, helplessness, and low self-esteem may have a common core of conceptually similar processes. A second point is that Brockner measured subjects' task performances, rather than persistence, on the test-phase task. Carver has argued elsewhere (1979) that when the situation does not allow behavioral withdrawal, a kind of *cognitive* withdrawal may occur, resulting in performance decrements. Brockner's data lend important support to that assertion.

Theoretical Implications

Three additional issues are raised by the nature of the variables from which we predicted subjects' behavior and by the nature of the behavior that we examined. These issues will be briefly addressed in the following paragraphs.

Self-awareness. The present findings quite obviously have some implications for theorizing about the behavioral consequences of self-directed attention. There is some division of opinion as to what these consequences are. Wicklund (1975a) has speculated that the prepotent response to self-focus is to attempt to

avoid it, either by withdrawing from the situation or by attempting to immerse one's attention in the environment (see also Wicklund, 1975b). In part, this is because Wicklund has viewed self-attention as an inherently aversive state, particularly when there is a large within-self discrepancy. To Wicklund, discrepancy reduction occurs only when self-focus cannot be avoided.

We, on the other hand, see self-focus as a necessary component process in the normal sequence of behavioral self-regulation (cf. Diener, 1979). From this perspective, we have argued that the prepotent response to self-directed attention is to attempt to reduce within-self discrepancies and that withdrawal occurs only if that attempt is interrupted, and even then only if self-assessment leads to a negative outcome expectancy (Carver, 1979; Carver et al., 1979). In our view, even when the discrepancy is very large, the reassertion impulse will hold sway if one's expectancy of being able to reduce the discrepancy is positive.

The results of the present research seem unequivocal in supporting our position over that of Wicklund. In order to provide a strong test of our reasoning, we designed the pretreatment phase of the present study so that large within-self discrepancies were produced among all subjects by a failure on the first task. Yet only when outcome expectancies were negative did self-focus lead to withdrawal. Subjects with positive expectancies exerted greater efforts in their attempts to do well on the second task when self-focus was high than when it was lower. It might be argued that our positive-expectancy subjects were simply immersing themselves in the second task to avoid self-focus (see Duval & Wicklund, 1972). This seems very unlikely, however, because an easier way of avoiding self-focus was readily available—asking for the subsequent puzzles, finishing them quickly, and leaving. (This was also the case in a study by McDonald, in press, cited earlier as providing indirect support for our model.) Moreover, if these subjects had been burying themselves in the experimental task in order to avoid self-focus, one must ask why subjects with negative expectancies did not similarly bury themselves in the task. In short, our explanation for the dichotomy of responses to self-focus is the most parsimonious presently available.

Self-efficacy. The present findings also appear to have implications for recent theorizing about "self-efficacy" and behavior regulation. There is a good deal of similarity between some of Bandura's (1977) ideas in that regard and aspects of our own reasoning, even though our analysis emphasizes moment-to-moment behavior regulation, whereas Bandura's emphasizes long-term differences as a function of therapy. It seems useful, therefore, to consider the implications of the present findings for the two models as they apply to persistence and giving up.

We have framed our discussion to this point in terms of "outcome" expectancies, following Carver's (1979) definition of outcome expectancy as the degree to which the person views a desired outcome as attainable. In our studies, however, we manipulated outcome expectancy largely by manipulating perceptions of efficacy expectancy.[4] Efficacy expectancy is viewed by Carver (1979) as one's self-perceived level of effectiveness in executing the needed responses, a perspective that is quite comparable to Bandura's, that is, "the conviction that one can successfully execute the behavior required to produce the outcomes" (1977, p. 193).

Although viewed in this light there are clear similarities between the models, the present data join those of Carver et al. (1979) in indicating the inadequacy of Bandura's theory in at least one important respect. Bandura's model fails to take into account the role of self-focus in the self-regulatory process. As is apparent from the present data, this is a role

[4] In Carver's (1979) terminology, outcome expectancy is held to depend on several variables, of which efficacy expectancy is only one. Others are such things as time constraints, environmental constraints, and knowledge of relevant response–outcome contingencies. Carver drew this distinction between these two types of expectancies because he believed that outcome expectancies are central in determining behavior (assertion vs. withdrawal), whereas efficacy expectancies determine the nature of concomitant affective experiences. Those predictions remain to be tested.

that should not be disregarded. That is, if we had not heightened self-focus among some of the subjects in the present research, we might have concluded that efficacy expectancies were not important in this behavioral context. Instead, we have shown that they are clearly important, though only in the presence of self-directed attention.

Helplessness. The present findings also seem to have some relevance for issues raised by researchers working in the "learned helplessness" tradition. All of our subjects received what might be construed as a helplessness pretreatment, and some of them displayed reduced persistence, which is often viewed as a sign of helplessness. For this reason, it is interesting to note some important similarities between our theory (and our findings) and the kind of cognitive analysis of helplessness in humans that recent theorists in that area have favored (e.g., Abramson, Seligman, & Teasdale, 1978; Hanusa & Schulz, 1977; Wortman & Brehm, 1975). We will then consider two differences between our model and previous theories.

Abramson et al. (1978), for example, have proposed a reformulation of helplessness theory in which the impact of a pretreatment on subsequent performance depends on one's beliefs about the relevance of one's pretreatment performance for performance on a subsequent task. In simpler terms, the effect of the pretreatment is mediated by its effect on the person's "expectation of future noncontingency." Expectation of future noncontingency is similar to what Wortman and Brehm (1975) called "expectancy of no control." Both, of course, are similar to what we have termed *unfavorable outcome expectancy.*

In the Abramson et al. model, this expectancy is influenced largely by the subject's attributions regarding the reasons for the initial failure.[5] Indeed, it is this attributional aspect of their model that has received by far the greatest amount of attention thus far. It is worth emphasizing, however, that despite this focus on attributions, it is the person's *expectancy* that ultimately influences performance, not the attributions themselves. The person who finds himself or herself expecting to do poorly, *for whatever reason,* should display "helplessness" effects. Persons who expect to do well should not show these effects.

The central importance of outcome expectancy in the Abramson et al. model is quite consistent with our own reasoning (Carver, 1979; Carver et al., 1979). However, our theory adds to theirs in two important respects. First, our model assumes that giving-up phenomena reflect an impulse to withdraw from the behavioral context. Indeed, both of the studies that have supported our analysis of giving up (Carver et al., 1979; the present Experiment 1) have used dependent measures embodying behavioral withdrawal as a response option. In contrast, in most studies in the helplessness tradition, such an option has not been explicitly available. As was mentioned above, Carver (1979) has suggested that when behavioral withdrawal is prevented, a cognitive withdrawal may occur, manifested either as a focus on one's inability to do the required behavior or as a mental dissociation from the task, in which further attempts are sporadic or halfhearted. Thus, in our view, the kinds of deficits shown in helplessness studies may actually depend on an impulse to remove oneself, physically or cognitively, from the behavioral context.

It is interesting to note, in this regard, that there is indirect support for this notion from at least one previous research project in the helplessness area (Diener & Dweck, 1978). Diener and Dweck studied the verbalizations and behaviors of children, classified either as

[5] Previous research has shown that exposure to a mirror can increase the amount of causality attributed to the self for hypothetical outcomes (Duval & Wicklund, 1973; Buss & Scheier, 1976). Inasmuch as the internal–external attributional dimension plays an important role in the model of Abramson et al. (1978), it might appear that their model thus is capable of explaining the persistence differences between self-awareness conditions obtained in the present research. This does not, however, appear to be the case. The Abramson et al. model seems to suggest that attributional internality will affect only the person's level of self-esteem following failure, not the extent to which persistence occurs on subsequent tasks. Thus, any explanation involving attributional internality would seem to be inadequate to account for the present findings.

"helpless" or "mastery-oriented," who were confronted with a series of task failures. The response style of mastery-oriented children included the following: monitoring of their own behaviors; verbalization of task-oriented self-instructions; attempts to renew their concentration; statements of positive affect ("I love a challenge"); and statements reflecting a positive prognosis ("I've almost got it now"). Helpless children did not make statements reflecting monitoring of their behavior or self-instructions but were much more likely to make statements of negative affect ("This isn't fun anymore") and statements that were irrelevant to task solution ("There is a talent show this weekend, and I'm going to be Shirley Temple"). Diener and Dweck characterized the latter category of statements as reflecting psychological withdrawal from the situation, which implicitly was not physically escapable. This characterization is quite consistent with our reasoning.

The second inadequacy of the Abramson et al. model is, of course, the same as the one discussed above regarding Bandura's (1977) analysis of fear-based behavior. Specifically, it has no mechanism for taking into account the effect of self-attention. Accordingly, it is unable to account for the present pattern of results. In supporting our theoretical model (Carver, 1979; Carver et al., 1979), the present findings thus suggest that the deficits obtained in helplessness research may be due in part to a process not yet addressed by helplessness theorists: the focusing of attention on oneself.

Reference Note

1. Carver, C. S., & Scheier, M. F. *The self-attention-induced feedback loop and human motivation: A control-theory analysis of social facilitation.* Manuscript submitted for publication, 1979.

References

Abramson, L. Y., Seligman, M. E. P., & Teasdale, J. D. Learned helplessness in humans: Critique and reformulation. *Journal of Abnormal Psychology,* 1978, *87,* 49–74.

Bandura, A. Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review,* 1977, *84,* 191–215.

Beaman, A. L., Klentz, B., Diener, E., & Svanum, S. Self-awareness and transgression in children: Two field studies. *Journal of Personality and Social Psychology,* 1979, *37,* 1835–1846.

Bem, D. J. Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental Social Psychology* (Vol. 6). New York: Academic Press, 1972.

Brockner, J. The effects of self-esteem, success-failure, and self-consciousness on task performance. *Journal of Personality and Social Psychology,* 1979, *37,* 1732–1741.

Buss, D. M., & Scheier, M. F. Self-consciousness, self-awareness, and self-attribution. *Journal of Research in Personality,* 1976, *10,* 463–468.

Carver, C. S. Facilitation of physical aggression through objective self-awareness. *Journal of Experimental Social Psychology,* 1974, *10,* 365–370.

Carver, C. S. Physical aggression as a function of objective self-awareness and attitudes toward punishment. *Journal of Experimental Social Psychology,* 1975, *11,* 510–519.

Carver, C. S. A cybernetic model of self-attention processes. *Journal of Personality and Social Psychology,* 1979, *37,* 1251–1281.

Carver, C. S., & Blaney, P. H. Avoidance behavior and perceived arousal. *Motivation and Emotion,* 1977, *1,* 61–73. (a)

Carver, C. S., and Blaney, P. H. Perceived arousal, focus of attention, and avoidance behavior. *Journal of Abnormal Psychology,* 1977, *86,* 154–162. (b)

Carver, C. S., Blaney, P. H., & Scheier, M. F. Focus of attention, chronic expectancy, and responses to a feared stimulus. *Journal of Personality and Social Psychology,* 1979, *37,* 1186–1195.

Carver, C. S., & Scheier, M. F. Self-focusing effects of dispositional self-consciousness, mirror presence, and audience presence. *Journal of Personality and Social Psychology,* 1978, *36,* 324–332.

Carver, C. S., & Scheier, M. F. Self-consciousness and reactance. *Journal of Research in Personality,* in press.

Diener, C. I., & Dweck, C. S. An analysis of learned helplessness: Continuous changes in performance, strategy, and achievement cognitions following failure. *Journal of Personality and Social Psychology,* 1978, *36,* 451–462.

Diener, E. Deindividuation: The absence of self-awareness and self-regulation in group members. In P. B. Paulus (Ed.), *The psychology of group influence.* Hillsdale, N.J.: Erlbaum, 1979.

Duval, S., & Wicklund, R. A. *A theory of objective self-awareness.* New York: Academic Press, 1972.

Duval, S., & Wicklund, R. A. Effects of objective self-awareness on attribution of causality. *Journal of Experimental Social Psychology,* 1973, *9,* 17–31.

Duval, S., Wicklund, R. A., & Fine, R. L. Avoidance of objective self-awareness under conditions of high and low intra-self discrepancy. In S. Duval and R. A. Wicklund, *A theory of objective self-awareness.* New York: Academic Press, 1972.

Feather, N. The relationship of persistence at a task to expectations of success and achievement related motives. *Journal of Abnormal and Social Psychology,* 1961, *63,* 552–561.

Feather, N., & Simon, J. Attribution of responsibility and valence of outcome in relation to initial confidence and success and failure of self and other. *Journal of Personality and Social Psychology,* 1971, *18,* 173–188.

Fenigstein, A., Scheier, M. F., & Buss, A. H. Public and private self-consciousness: Assessment and theory. *Journal of Consulting and Clinical Psychology,* 1975, *45,* 522–527.

Froming, W. J. The relationship of moral judgment, self-awareness, and sex to compliance behavior. *Journal of Research in Personality,* in press.

Glass, D. C., & Singer, J. E. *Urban stress: Experiments on noise and social stressors.* New York: Academic Press, 1972.

Hanusa, B. H., & Schulz, R. Attributional mediators of learned helplessness. *Journal of Personality and Social Psychology,* 1977, *35,* 602–611.

Hiroto, D. S., & Seligman, M. E. P. Generality of learned helplessness in man. *Journal of Personality and Social Psychology,* 1975, *31,* 311–327.

McDonald, P. J. Reactions to objective self-awareness. *Journal of Research in Personality,* in press.

Roth, S., & Kubal, L. Effects of noncontingent reinforcement on tasks of differing importance: Facilitation and learned helplessness. *Journal of Personality and Social Psychology,* 1975, *32,* 680–691.

Scheier, M. F., & Carver, C. S. Self-focused attention and the experience of emotion: Attraction, repulsion, elation, and depression. *Journal of Personality and Social Psychology,* 1977, *35,* 625–636.

Scheier, M. F., Carver, C. S., & Gibbons, F. X. Self-directed attention, awareness of bodily states, and suggestibility. *Journal of Personality and Social Psychology,* 1979, *37,* 1576–1588.

Scheier, M. F., Carver, C. S., & Gibbons, F. X. Self-focused attention and reactions to fear. *Journal of Research in Personality,* in press.

Scheier, M. F., Fenigstein, A., & Buss, A. H. Self-awareness and physical aggression. *Journal of Experimental Social Psychology,* 1974, *10,* 264–273.

Seligman, M. E. P. *Helplessness: On depression, development, and death.* San Francisco: Freeman, 1975.

Steenbarger, B. N., & Aderman, D. Objective self-awareness as a non-aversive state: Effect of anticipating discrepancy reduction. *Journal of Personality,* in press.

Tresselt, M., & Mayzner, M. Normative solution times for a sample of 134 solution words and 378 associated anagrams. *Psychonomic Monograph Supplements,* 1966, *1,* 293–298.

Vallacher, R. R., & Solodky, M. Objective self-awareness, standards of evaluation, and moral behavior. *Journal of Experimental Social Psychology,* 1979, *15,* 254–262.

Wicklund, R. Objective self-awareness. In L. Berkowitz (Ed.), *Advances in experimental social psychology.* New York: Academic Press, 1975. (a)

Wicklund, R. A. Discrepancy reduction or attempted distraction? A reply to Liebling, Seiler, and Shaver. *Journal of Experimental Social Psychology,* 1975, *11,* 78–81. (b)

Wortman, C. B., & Brehm, J. W. Responses to uncontrollable outcomes: An integration of reactance theory and the learned helplessness model. In L. Berkowitz (Ed.), *Advances in experimental social psychology.* New York: Academic Press, 1975.

Received December 6, 1978

rnal of Personality and Social Psychology
79, Vol. 37, No. 10, 1871-1876

Consistency and Bias in the Attribution of Attitudes

Icek Ajzen, Carol Ann Dalto, and Daniel P. Blyth
University of Massachusetts—Amherst

This article is based on the proposition that people attribute a disposition to an actor by evaluating its consistency with other information about the actor or the situation. This strategy is assumed to be accompanied by a cognitive set or bias to view ambiguous information as consistent with the hypothesized disposition. Respondents were told that a student had chosen or had been assigned to write a proabortion or antiabortion paper, and they were or were not given an ambiguous description of the author's personality. In support of predictions, under choice conditions attitudes were always attributed in accordance with the paper's position, but under assignment conditions such attributions occurred only when respondents received the ambiguous personality description.

A central focus of attribution theory is the tent to which an actor's behavior provides formation about his or her stable underlying spositions (cf. Ajzen, 1971; Jones & Davis, 965; Steiner, 1970). Most analyses of the tribution process (e.g., Heider, 1958; Jones Davis, 1965; Kelley, 1967) have likened servers to naive scientists who, except for a w motivational biases, make systematic use the information available to them. It has cently been suggested, however, that instead using sophisticated information-processing rategies, people rely on rather simple intuive heuristics (Tversky and Kahneman, 974) or cognitive scripts (Abelson, 1976) in aking their judgments (see Slovic, Fischhoff, Lichtenstein, 1977, for a review). Although ese intuitive strategies often produce reanable judgments, they are said to be capable leading to systematic biases and errors Ajzen, 1977; Kahneman & Tversky, 1973; isbett & Borgida, 1975).

This article deals with what Ross (1977) s termed the *fundamental attribution error,* e assumed tendency for observers to overtimate the importance of dispositional factors relative to environmental factors. We will try to identify a cognitive strategy that may be in part responsible for any such bias and to specify some of the conditions under which this bias is likely to materialize.

One area in which the fundamental attribution error is assumed to operate is the attribution of attitudes on the basis of observed (or reported) behavior. Jones and Harris (1967) reported that observers attribute attitudes to the author of an essay in accordance with the position advocated in the essay. Although weaker, this tendency also emerged under conditions in which situational constraints, in the form of an assigned position on the issue, appeared sufficient to explain the actor's behavior. This and other studies along the same lines (e.g., Jones, Worchel, Goethals, & Grumet, 1971; Miller, 1976; Miller, Mayerson, Pogue, & Whitehouse, 1977; Snyder & Jones, 1974) are usually taken as evidence for a pervasive bias on the part of observers to attribute an actor's behavior to internal dispositions rather than to environmental factors.

Note, however, that even under conditions of little choice, attribution of behavior to internal factors does not necessarily imply a bias in favor of dispositional explanations. The presence of a viable situational cause (such as assignment by a course instructor) should produce some discounting of the dis-

The authors thank Ronnie Bulman for her comnts on an earlier draft of this paper.
Requests for reprints should be sent to Icek Ajzen, partment of Psychology, University of Massachuts, Amherst, Massachusetts 01003.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3710-1871$00.75

positional explanation (Kelley, 1973), but it need not completely discredit it. We might thus reasonably expect that under situational constraints, the position expressed in an essay would have a significant, though weaker, effect on the author's perceived attitude.

Before considering potential biases, we must explore the cognitive strategies people employ when asked to judge the likelihood that the author of an essay actually holds an attitude in accordance with the expressed position. This article is based on the proposition that in attributing an attitude or any other disposition to an actor, observers evaluate the extent to which the disposition in question would be consistent with other available information. A similar strategy has been suggested by Ajzen and Fishbein (1975) in their application of Bayes' theorem to causal attribution, and by Kruglanski, Hamel, Maides, and Schwartz (1978) in their analysis of lay epistemology.

Evaluation of consistency in the validation of specific hypotheses concerning an actor's dispositions can be described as follows. When judging the validity of a given dispositional attribution, the observer examines the extent to which the stipulated disposition is consistent with other information about the actor and about the context in which the behavior occurred. The greater the number of informational items consistent with the disposition, the more confident the observer will be that the actor in fact has the disposition in question. Information inconsistent with the stipulated disposition should lower the observer's confidence.[1]

The interpretation of information as consistent or inconsistent with a given disposition is, however, not made in a vacuum. It is here proposed that these judgments are influenced, among other things, by prior beliefs concerning the validity of the hypothesized disposition. Once a person has formed a given dispositional hypothesis, he or she will tend to interpret new information as consistent with the disposition in question. A prior hypothesis is thus assumed to provide a "cognitive set" for the processing of new information. As is true of a perceptual set, this tendency to view new information as consistent with a prior hypothesis should be particularly pronounce in the case of ambiguous information that is and of itself has little bearing on the validity of the hypothesis under consideration. Th end result of such a tendency would be a overestimation of the amount of evidence sup porting the hypothesis, in this instance th disposition that is being attributed.

A bias toward perceiving ambiguous infor mation as consistent with prior hypothese could help explain attribution of attitude under situational constraints. In previou studies, subjects were given a considerabl amount of information that may have bee interpreted as consistent with the hypoth esized attitude. For example, in their first ex periment, Jones and Harris (1967) describe the author of the essay as "a student at th University of North Carolina, a resident o the state, and the son of an automobile sales man." Furthermore, the respondents read th 200-word essay defending or criticizing Ca tro's Cuba. The direction of the essay pro vided the initial dispositional hypothesis. Th description of the author and the style an organization of the material in the essay, bo ing largely ambiguous with respect to the di positional hypothesis in question, may hav been interpreted as consistent with the infe ence that the author's attitude correspond to his expressed position in favor of or again Castro's Cuba. Providing ambiguous back ground information may thus have attenuate any tendency to discount the disposition hypothesis under conditions of situation constraints.

One implication of this analysis is that the ambiguous background information we removed, apparent support for the hypoth esized disposition would be greatly reduced eliminated, thus resulting in little dispos tional attribution under no choice condition The study reported below was designed part to test this prediction. In addition, it a tempted to provide a more direct test of th operation of the cognitive set assumed t accompany use of the consistency strateg

[1] The person is also assumed to take into accou each item's diagnosticity or informational value f the disposition under consideration.

Information known to be ambiguous with respect to the attribution in question was provided for some subjects. This information, in the form of a general personality description, was assumed to permit emergence of the consistency bias and hence to increase dispositional attributions in the presence of situational constraints.

Method

Subjects

A total of 256 male and female[2] undergraduate psychology students received experimental credit for their participation in the study. They were assigned at random and in equal numbers to the 12 experimental and 4 control conditions (16 per cell).

Procedure

The participants were scheduled in small groups and were handed a self-contained questionnaire that described the experiment as a study of the perception and interpretation of social events. They then read about a student who had written a paper in favor of or against abortion. The stand taken on the issue either had been assigned to the student or had been a matter for his or her choice. For one-third of the participants, the freedom-of-choice information was followed by an ambiguous description of the student's personality. The order of presentation was reversed for another third of the participants. Here, the personality description appeared prior to the choice information. The remaining participants received no information about the student's personality. Specifically, the two conditions containing personality descriptions provided the following information.

> Peter G. is a student at Hampshire College. Last semester he wrote a paper for a social science class in which he took a stand favoring [opposing] abortion on demand. The course instructor had given the following instructions for the paper.
>
> *Free choice condition.* Based on the past week's readings and lectures, write a short essay either defending or criticizing abortion on demand.
>
> *No choice condition.* Based on the past week's readings and lectures, write a short essay defending [criticizing] abortion on demand.

The personality description appeared either prior to or after this choice manipulation.

> The following paragraph gives you a short description of Peter's personality. It was prepared by a psychologist on the basis of a personality inventory administered to all students in the class.
>
> Peter G. prides himself as being an independent thinker and he does not accept others' opinions without satisfactory proof. He has a great deal of unused capacity that he has not turned to his advantage. Peter has a tendency to be critical of himself. He has a strong need for other people to like and admire him. At times he has serious doubts as to whether he has made the right decision or done the right thing. Disciplined and controlled on the outside, he tends to be worrisome and insecure on the inside. Some of his aspirations tend to be quite unrealistic. Peter has found it unwise to be too frank in revealing himself to others. He prefers a certain amount of change and variety and becomes dissatisfied when hemmed in by restrictions and limitations. While he has some personality weaknesses, he is generally able to compensate for them.

This description contains 10 of the 12 statements used previously by Ulrich, Stachnik, and Stainton (1963), who showed that students readily accept such general statements as accurate descriptions of their personalities (see also Carrier, 1963; Forer, 1949).

At this point (or immediately after the choice information in the no personality conditions), attribution of attitude was assessed. "What do you think is Peter G.'s true position on the question of abortion on demand? The probability that Peter G. favors (opposes) abortion on demand is _____%."

To test our assumption that the personality description was ambiguous or neutral with respect to this judgment, control subjects were given no information about the stand Peter G. had taken on the issue of abortion or about his freedom of choice. They were simply told that he had written a paper for a social science class concerning the issue of abortion on demand, and one-half of the control subjects were given the description of Peter G.'s personality. On the basis of this information, they were asked to judge the probability that Peter G. favored (or opposed) abortion on demand.

Results

Ambiguity of Personality Description

A 2 × 2 analysis of variance was performed on the control data. The factors in this analysis were personality description (yes or no) and question format (in favor of or opposed to abortion). Responses were scored, with high numbers indicating a proabortion stand. That is, when subjects judged the probability that Peter G. was opposed to abortion, their responses were subtracted from 100. Neither

---

[2] Since there were no significant differences between males and females, the participants' sex will not be further considered.

Table 1
Mean Estimates of Proabortion Attitudes Under Different Experimental Conditions

Direction of essay	No personality description		Personality description	
	Choice	No choice	Choice	No choice
Pro-abortion	80.44	57.19	68.50	63.59
Anti-abortion	20.94	50.00	35.68	44.47

Note. $n = 16$ per cell in the no personality conditions and $n = 32$ per cell in the personality conditions.

the main effects nor the interaction were statistically significant ($F < 1.0$ in each case). Disregarding question format, the mean probability of a proabortion attitude was judged to be 57% when the personality description was provided and 59.5% without the personality description. Clearly, then, in and of itself the personality description had no appreciable effect on attitude attributions.

Attribution of Attitudes

The data of the experimental conditions were submitted to a 2 × 2 × 3 analysis of variance, with direction of essay (pro vs. con), freedom of choice (choice vs. no choice), and personality description (before choice information vs. after choice information vs. no personality description) as design factors. Since there were no significant differences due to the positioning of the personality description (before or after the choice information), the data for these two conditions were pooled, resulting in a 2 × 2 × 2 design. For ease of presentation, the data are discussed only in terms of this latter design.

Table 1 presents mean probability judgments, scored with high numbers again indicating attributions of proabortion attitudes. Table 2 shows the results of the analysis of variance. The significant main effect for direction of essay indicates that, predictably, a student who wrote a proabortion essay was viewed as more likely to be in favor of abortion than a student who wrote an antiabortion essay. More interesting, the significant interaction between essay direction and fre dom of choice shows that this effect w stronger in the choice condition than in t no choice condition.

Most important, however, is the significa three-way interaction. It indicates that t presence or absence of an ambiguous pe sonality description influenced the extent which the effect of essay direction was co tingent upon freedom of choice. The patte of means shown in Table 1 supports our h potheses. Consider first the results obtain when no personality description was provide It can be seen that as in previous studi essay direction had a strong effect on attri uted attitudes under free choice condition Contrary to previous findings, however, the was virtually no difference due to direction the essays under no choice conditions. Po hoc comparisons revealed that the effect d to direction of essay was significantly grea for choice than for no choice subjects, $F($ 184) = 20.97, $p < .001$. Moreover, the di ference between 57.19 and 50.00 in the n choice condition was not statistically signi icant ($F = 1.21$).

Very different results emerged when t participants were given the ambiguous d scription of Peter G.'s personality. Here, su jects attributed attitudes in accordance wi the position expressed in the essay even in t presence of situational constraints. Essay d rection had a significant effect on attributio under choice conditions, $F(1, 184) = 25.2$ $p < .001$, by post hoc comparison, as well under no choice conditions, $F(1, 184) = 8.$ $p < .005$. Further, the effects due to the

Table 2
Analysis of Variance

Source	*df*	*M*	*F*
Direction of essay (A)	1	38,817.19	113.
Freedom of choice (B)	1	247.52	<1.0
Personality description (C)	1	35.65	<1.0
A × B	1	8,480.08	24.
A × C	1	577.71	1.
B × C	1	9.69	<1.0
A × B × C	1	3,971.94	11.
Error	184	341.50	

* $p < .001$.

rection of the essays (proabortion or antiabortion) did not differ significantly under the two freedom-of-choice conditions ($F = 2.20$).

Post hoc comparisons also revealed that the ambiguous personality description *reduced* the effect of essay direction in the free choice situation. The difference between proabortion and antiabortion essays, given a personality description ($M = 32.82$), was significantly smaller, $F(1, 184) = 8.34$, $p < .005$, than the difference between these essays when no personality description was provided ($M = 59.50$).

Discussion

The findings reported in this article are consistent with our analysis of dispositional attribution. When no personality description of the essay's author was made available, respondents' judgments were strongly influenced by the presence or absence of situational constraints. Attitudes in accordance with the position advocated in the essay were attributed to authors who were free to support or oppose abortion on demand, while respondents were uncertain as to the attitude of an author who had written his essay under no choice conditions. In fact, given situational constraints and no personality description, dispositional attributions were very close to those made by control subjects who were given no information about the position advocated in the essay. These results provide little support for the argument that observers have a pervasive tendency to attribute an actor's behavior to internal dispositions. When situational constraints are salient and no other information is available, there seems to be no inclination to make dispositional attributions.

However, the availability of other information in the form of an ambiguous description of the actor's personality had a profound effect on dispositional attributions. Even though by itself the description was found to be completely uninformative with respect to attitudes toward abortion, it increased dispositional attributions in the no choice conditions. At the same time it was found to weaken such attributions in the choice conditions. Only the former effect was predicted, but both effects may in retrospect be viewed as reflecting the same underlying process. When no personality description was provided, all information available to the respondent was either consistent with the hypothesized disposition (in the choice conditions) or largely discredited by situational constraints (in the no choice conditions). As a result, freedom of choice had a strong effect on attribution of attitudes.

In contrast, even given the operation of a cognitive set to interpret ambiguous information as consistent with the hypothesized attitude, some of the information contained in the personality description is likely to have been viewed as contradicting the dispositional attribution. Taken as a whole, the available information would still favor attribution of an attitude corresponding to the behavior, but the few items of contradictory evidence would tend to lower confidence in the attribution. The situational constraints in the no choice conditions would again act to discredit the behavioral information and thus to some extent further reduce the strength of the dispositional attributions. Consequently, freedom of choice was found to have an effect on attribution of attitudes, but its effect was not significant and was much less pronounced than in the conditions where no personality description was provided.

This finding appears to contradict previous research (e.g., Jones & Harris, 1967), which has reported significant interactions between essay direction and freedom of choice. Note, however, that although previous research also provided some ambiguous background information, the personality conditions of the present study contained a much more elaborate set of such information explicitly designed to be interpretable as consistent with a proabortion or antiabortion attitude. As a result, only a little discounting of the dispositional hypothesis was found to occur under no choice conditions when the personality description was provided.

The tendency revealed in the present study to view ambiguous information as consistent with a prior hypothesis could theoretically function in different ways. One possibility is

that a dispositional attribution in accordance with the essay's position becomes anchored in the personality description, and as a result subsequent refutational evidence (in the form of information about situational constraints) is rendered relatively ineffective (cf. Ross, 1977). This account was not supported by the present study, since providing the personality description had the same effect whether it appeared prior to or after the freedom-of-choice information. Instead, information about the essay appears to have led respondents to hypothesize a corresponding disposition, and evidence consistent with this hypothesis was then sought in other available information. The cognitive set created by such an orientation seems to produce a tendency to interpret ambiguous information as consistent with the hypothesis in question.

References

Abelson, R. P. Script processing in attitude formation and decision-making. In J. S. Carroll and J. W. Payne (Eds.), *Cognition and social behavior*. Hillsdale, N.J.: Erlbaum, 1976.

Ajzen, I. Attribution of disposition to an actor: Effects of perceived decision freedom and behavioral utilities. *Journal of Personality and Social Psychology*, 1971, *18*, 144–156.

Ajzen, I. Intuitive theories of events and the effects of base rate information on prediction. *Journal of Personality and Social Psychology*, 1977, *35*, 303–314.

Ajzen, I., & Fishbein, M. A. Bayesian analysis of attribution processes. *Psychological Bulletin*, 1975, *82*, 261–277.

Carrier, N. A. Need correlates of "gullibility." *Journal of Abnormal and Social Psychology*, 1963, *66*, 84–86.

Forer, B. R. The fallacy of personal validation: A classroom demonstration of gullibility. *Journal of Abnormal and Social Psychology*, 1949, *44*, 118–123.

Heider, F. *The psychology of interpersonal relations*. New York: Wiley, 1958.

Jones, E. E., & Davis, K. E. From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2). New York: Academic Press, 1965.

Jones, E. E., & Harris, V. A. The attribution of attitudes. *Journal of Experimental Social Psychology*, 1967, *3*, 1–24.

Jones, E. E., Worchel, S., Goethals, G. R., & Grumet, J. F. Prior expectancy and behavioral extremity as determinants of attitude attribution. *Journal of Experimental Social Psychology*, 1971, *7*, 59–80.

Kahneman, D., & Tversky, A. On the psychology of prediction. *Psychological Review*, 1973, *80*, 237–251.

Kelley, H. H. Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 15). Lincoln: University of Nebraska Press, 1967.

Kelley, H. H. The processes of causal attribution. *American Psychologist*, 1973, *28*, 107–128.

Kruglanski, A. W., Hamel, I. A., Maides, S. A., Schwartz, J. M. Attribution theory as a special case of lay epistemology. In J. H. Harvey, W. Ickes, & R. F. Kidd (Eds.), *Advances in attribution research* (Vol. 2). Hillsdale, N.J.: Erlbaum, 1978.

Miller, A. G. Constraint and target effects in the attribution of attitudes. *Journal of Experimental Social Psychology*, 1976, *12*, 325–339.

Miller, A. G., Mayerson, N., Pogue, M., & Whitehouse, D. Perceivers' explanations of their attributions of attitude. *Personality and Social Psychology Bulletin*, 1977, *3*, 111–114.

Nisbett, R. E., & Borgida, E. Attribution and the psychology of prediction. *Journal of Personality and Social Psychology*, 1975, *32*, 932–943.

Ross, L. The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10). New York: Academic Press, 1977.

Slovic, P., Fischhoff, B., & Lichtenstein, S. Behavioral decision theory. *Annual Review of Psychology*, 1977, *28*, 1–39.

Snyder, M., & Jones, E. E. Attitude attribution when behavior is constrained. *Journal of Experimental Social Psychology*, 1974, *10*, 585–600.

Steiner, I. D. Perceived freedom. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 5). New York: Academic Press, 1970.

Tversky, A., & Kahneman, D. Judgment under uncertainty: Heuristics and biases. *Science*, 1974, *185*, 1124–1131.

Ulrich, R. E., Stachnik, T. J., & Stainton, N. R. Student acceptance of generalized personality interpretations. *Psychological Reports*, 1963, *13*, 831–834.

Received December 11, 1978

Journal of Personality and Social Psychology
1979, Vol. 37, No. 10, 1877-1887

On the Requirements of Proof: The Timing of Judicial Instruction and Mock Juror Verdicts

Saul M. Kassin and Lawrence S. Wrightsman
University of Kansas

At the close of a trial, the judge instructs the jury that the defendant is presumed innocent, that the burden of proof is on the prosecution, and that guilt must be established beyond a reasonable doubt. In view of criticisms that the judge's charge has no effect on jury decisions, the present study examined whether the timing of judicial instruction mediates its efficacy. One hundred seven mock jurors watched a 1-hour videotape of a trial and were instructed by the judge either before the evidence, after the evidence, or not at all. Results for posttrial measures indicated that although the timing manipulation had no significant effect on the standards of reasonable doubt adopted by subjects, those who were instructed before the evidence viewed the defendant as less likely to have committed the crime and demonstrated a lower conviction rate than subjects in the instructions-after and no-instructions groups. Results for a series of mid-trial judgments of 54 subjects further indicated that pre-instructed subjects were less likely to convict throughout the trial. These findings are discussed, and their implications for procedural reform in the courtroom are noted.

In trials by jury, the judge is obligated to instruct jurors in both general and case-specific matters of law. These instructions serve a number of functions: The judge orients jurors in their task, outlines the undisputed facts and issues of the case, explains the relevant law, and informs jurors about procedural matters (McBride, 1969). From the defendant's perspective, perhaps the most crucial of the mandatory instructions is that concerning the "requirements of proof" (LaBuy, 1963). Specifically, the accused is entitled to the instruction that he or she is presumed innocent, that the burden of proof is on the prosecution, and that all elements of the crime must be proven to a constitutional standard of "beyond a reasonable doubt."

Although the potential importance of the judge's charge is widely recognized (McCart, 1964), its actual effectiveness is a subject of controversy. On the one hand, the courts assume that jurors understand their instructions, use them in making decisions, and are acutely sensitive to even minor variations in wording. Thus, attorneys often request specific instructions, appellate courts have occasionally reversed verdicts on the basis of an improperly worded instruction, and many states currently favor the practice of having judges recite a preapproved pattern instruction in order to ensure standardization. Experimental support for the effect of variations in instructional content has recently been obtained. Kerr et al. (1976) presented mock jurors with one of three definitions of reasonable doubt: one that described a lax criterion of reasonableness (i.e., "you need not be absolutely sure that the defendant is guilty to find him guilty," p. 286), one that described an extremely stringent criterion (i.e., "if you are not sure and certain of his guilt, you must

This research was conducted while the first author was on a postdoctoral fellowship at the University of Kansas and was supported in full by Faculty Research Grant 3401 to the second author. We would also like to thank Martin F. Kaplan and two anonymous reviewers for their helpful comments on the manuscript.

Requests for reprints should be addressed to Saul M. Kassin, who is now at the Department of Psychological Sciences, Purdue University, West Lafayette, Indiana 47907.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3710-1877$00.75

find him not guilty," p. 286), and one in which reasonable doubt was not defined. As it turned out, these varying definitions significantly influenced both individual and group verdicts—a lax criterion resulted in a high conviction rate, whereas a stringent one produced a low rate of conviction. Subjects for whom reasonable doubt was not explicated fell between these extremes. This study thus demonstrated that verdicts are indeed influenced by the reasonable doubt element of the requirements-of-proof instruction.

On the other hand, it has been suggested by legal scholars (Frank, 1949) and researchers (Sealy & Cornish, 1973) that these instructions have little effect on jurors' verdicts. One common criticism is that because they are written in statutory language, the instructions are often confusing to laypersons untrained in the law (Elwork, Sales, & Alfini, 1977). In fact, one study revealed that 40% of 375 sampled jurors reported that they did not understand their judge's instruction (Hervey, 1947). A second criticism is aimed at the timing of the judge's charge. Although the procedure is not fixed by law, the jury is typically instructed at the close of the trial presentation, that is, after the evidence has been presented. Although it is possible that this sequence increases the salience of the instruction and its availability for recall during deliberation,[1] a number of sources (e.g., McBride, 1969) have questioned the utility of an instruction that is given at a stage where jurors might have already decided on a verdict. Kalven and Zeisel (1966) noted that jurors often form very definite opinions about a defendant's guilt or innocence before the close of the trial. Accordingly, Judge E. Barrett Prettyman (1960) argued the following:

> It makes no sense to have a juror listen to days of testimony only then to be told that he and his conferees are the sole judges of the facts, that the accused is presumed to be innocent, that the government must prove guilt beyond a reasonable doubt, etc. What manner of mind can go back over a stream of conflicting statements and alleged facts, recall the intonations, the demeanor, or even the existence of the witnesses, and retrospectively fit all these recollections into a pattern of evaluation and judgment given him for the first time after the events; the human mind cannot do so. . . . Why should not the judge, when the jury is sworn, then and there tell them the rules of the game. (p. 1066)

In view of the flexibility in courtroom procedure and the potential importance of the judge's charge to the jury, the absence of research on temporal factors associated with the instructions is surprising (Elwork, Sales, & Alfini, 1977, is an exception). After all, order effects in the perception of unfolding behavior sequences have repeatedly been noted in both person perception (Jones et al., 1968) and jury contexts (Walker, Thibaut, & Andreoli, 1972). In the latter study, Walker et al. (1972) varied the order in which the prosecution and defense presented their cases but did not address the issue of judicial instruction.

The present study was designed to investigate the relationship between the timing of the judge's instructions and mock juror verdicts for a criminal case. In particular, the aims of the research were twofold. First, we sought to determine whether or not a prototypical, ecologically valid instruction on the requirements of proof influences jurors' decisions and whether the timing of that instruction mediates its efficacy. In order to achieve these goals, requirements-of-proof instructions were gleaned from those currently employed (De Vitt & Blackmar, 1977; LaBuy, 1963; McBride, 1969) and were introduced to subjects before testimony, after testimony, or not at all. At the conclusion of the trial presentation, subjects rendered their individual verdicts and answered a number of other case-related questions. Since the requirements of proof are defendant oriented, their effectiveness should be manifested in a lowered rate of conviction. Based on jurors' tendencies to make early decisions about guilt or innocence and Judge Prettyman's (1960) reasoning that instructions can have their intended effects only when they are delivered before jurors have made up their minds, a kind of primacy or inoculation effect was predicted. That is, the present instructions should produce fewer guilty verdicts when presented before the evidence than when presented after it.

The second general aim of the present study was to explore the mechanism or cog-

[1] Jones and Goethals (1971) have termed this the "recall readiness" hypothesis for recency effects.

nitive process that underlies the proposed order effect. Requirements-of-proof instructions may operate either by decreasing the belief that the defendant committed the crime (i.e., a lowered probability-of-commission estimate) or by increasing the standard or threshold to which that likelihood is compared (i.e., a stringent interpretation of reasonable doubt). Consequently, both were assessed. A collateral issue addressed here was whether or not subjects who are instructed prior to the presentation of evidence evaluate evidentiary information differently as it unfolds from those who do not receive prior instruction. Accordingly, half the subjects in each instruction group made judgments of guilt–innocence at various points during the trial. It was hypothesized that instructed subjects would set a higher standard by which to evaluate the case against the defendant and would therefore be less influenced by the prosecutor's case throughout the trial.

Method

Subjects and Design

A total of 107 introductory psychology students (47 male, 60 female) participated in the study. The experiment was conducted in 25 small groups ranging in size from 3 to 7 and took 1 hour and 20 minutes to complete. Each group was randomly assigned to one of six cells produced by the 3 (Judge's Instructions Before Evidence, Instructions After Evidence, No Instructions) × 2 (Multiple Judgments vs. Single Judgment Only) factorial design.

Procedure

Upon entering, subjects were told that they would observe a videotaped trial simulation, after which they would be asked to render a verdict. They were further instructed that as jurors they should be attentive but should not take notes and should not converse while the trial presentation was in progress.

All subjects were then told that in order to get the entire trial presentation on one reel of tape, a few pauses and meaningless exchanges had been deleted, but that all the testimony remained intact. Those groups who were to receive the judge's instruction, however, were informed that this instruction had inadvertently been deleted but would be read to them at the appropriate time from the original transcript. Finally, subjects in the multiple-judgment cells were informed of the fact that at certain points the tape would be stopped and their judgments would be assessed. The trial was then presented.

The Trial

The stimulus trial was one that had previously been employed (Juhnke et al., 1979). Stylistically, the simulation was presented on a 1-hour (black and white) videotape in which a number of law students retried a case in a realistic courtroom setting. It was filmed from a juror's perspective: the judge, attorneys, and witness stand were all in view.

Substantively, the trial was based on an actual criminal case in which the defendant, Ronald Oliver, was charged with stealing a car and with transporting it across state lines.[2] Though continuously presented, the trial consisted of the following three distinct phases: (a) opening statements by prosecutor and defense; (b) direct examination, cross-examination, and redirect examination of two prosecution witnesses (the salesman from whom the vehicle was stolen and the arresting officer) and one defense witness (the defendant); and (c) closing arguments of counsel (prosecution, defense, prosecution). The judge's instructions on the requirements of proof represented a fourth phase whose presence–absence and timing were varied.

In one condition, these instructions appeared prior to the introduction of evidence (i.e., between the first and second phases). In a second condition, they appeared after the closing arguments (i.e., after the third phase). In a third condition, no instruction was given. The specific instruction employed was neither strong nor weak. Rather, it was patterned after the approved instructions designed to convey each element of the requirements of proof: presumption of innocence, burden of proof, and reasonable doubt (see DeVitt & Blackmar, 1977; McBride, 1969). The instruction read as follows:

> Ladies and gentlemen of the jury—at this point I want to emphasize that the law presumes the defendant, Ronald Oliver, to be innocent unless proven otherwise. A defendant begins the trial with a "clean slate" with no evidence against him.
>
> This presumption places the burden *not* upon the defendant to prove his innocence but, on the contrary, the burden is on the prosecution to convince you beyond any reasonable doubt that the defendant, Ronald Oliver, committed the crime. That burden never shifts at any stage of the proceeding to the defendant. Ronald Oliver has no obligations of any kind to go forward and prove that he is innocent.
>
> You have now heard the term "reasonable doubt."

---

[2] A transcript taken verbatim from the videotape and a summary of the trial are available upon request.

Table 1
Pattern of Final Verdicts in Each of the Six Cells

Instruction and verdict	Multiple judgments		Single judgments	
	No.	%	No.	%
Before				
Guilty	7	39	6	35
Not guilty	11	61	11	65
After				
Guilty	12	63	10	56
Not guilty	7	37	8	44
None				
Guilty	12	71	10	56
Not guilty	5	29	8	44

What is it? It is a doubt based upon reason and common sense—the kind of doubt that would make a reasonable person hesitate to act in important matters. To summarize, the defendant is presumed innocent, so the prosecution must prove to your satisfaction beyond any reasonable doubt that Ronald Oliver is guilty. If two conclusions can reasonably be drawn from the evidence—one of innocence and one of guilt—the jury should adopt the one of innocence.

Dependent Measures

At the close of the trial, all subjects responded individually and without deliberation to a two-page questionnaire in which they first rendered a dichotomous judgment (guilty–not guilty) and indicated their confidence in that verdict on a 0–8-point scale. They then rated the strength of the evidence as well as their interest and involvement in the case (all on 0–8-point scales).

Since verdicts are a function of the perceived probability that the defendant committed the crime and of the standard of proof deemed necessary for conviction, both of these variables were also assessed. All subjects were thus asked, "What is the likelihood that the defendant committed the crime?" to which they responded by circling a number from 0 to 100 (in multiples of 5), and "A defendant should be found guilty if there is at least a _____% chance that he committed the crime." Finally, subjects answered 16 short-answer recall questions that pertained to the major facts of the case (e.g., "On what highway was Ron Oliver stopped?"). The total number of correctly recalled items served as a measure of fact recall.

In addition to providing these outcome data, half the subjects indicated their judgments (guilty–not guilty), confidence values, and probability-of-commission estimates at six distinct points during the trial—after both the direct examination and cross-examination of each witness. Specifically, they were asked, "If the trial ended now, would you vo that the defendant is *guilty* or *not guilty*?" "Ho confident are you in this judgment?" (0–8) "Wh is the likelihood that the defendant committed t crime?" (0–100). It was thus possible to trace b liefs as the trial unfolded and to examine wheth subjects who were already instructed on the r quirements of proof (instructions before) evaluat the evidence differently from those who had n yet been instructed (instructions after and no i structions). Previous research employing the multiple judgments (e.g., Weld & Danzig, 1940) h been criticized on the ground that such a pro cedure may bias final verdicts (Davis, Bray, Holt, 1977). For that reason, only half the sub jects in the present study made these on-li responses. A comparison of their posttrial respons with those of single-judgment subjects thus pro vided a test for the (non)reactivity of the pro cedure.

Results

Outcome Measures

Overall, 53% of the subjects voted guilt and 47% voted not guilty. The pattern o verdicts in each cell appears in Table 1.

In order to test for all main and interac tion effects and to determine which model best fits these categorical data, these dichoto mous judgments were analyzed with a likeli hood ratio (goodness of fit) chi-square (Fien berg, 1977). Results indicated that th simplest model that described the observe frequencies was the two-way interaction be tween instructional set and verdicts, $\chi^2(6) = 1.21$, $p > .97$. None of the more complicate models (i.e., those involving the multiple vs single judgments factor) contributed an explanatory power to this Instructional Set × Verdict model. Put another way, the pattern of verdicts was accounted for by the main effect for timing of instruction. Table 1 shows that the instructions-before condition pro duced 37% guilty verdicts, compared to 59% in the instructions-after and 63% in the no instructions conditions.

A scalar variable was defined by combining subjects' verdicts with their confidence rat ings (confidence itself was unaffected by the independent variables). Specifically, positive confidence values were assigned to guilty ver dicts and negative values to verdicts of not guilty. Scores could thus range from −8 (maximum confidence in not-guilty verdict

to 8 (maximum confidence in guilty verdict). A 2 × 3 analysis of variance on this measure revealed one marginally significant effect for timing of instruction, $F(2, 101) = 2.90$, $p < .06$. Duncan's multiple-range test further indicated that subjects in the instructions-before condition were less likely ($p < .05$) to convict the defendant than were noninstructed subjects (means of −.94 and 2.35, respectively). Instructions-after ($M = 1.79$) and noninstructed conditions did not so differ. An analysis of variance on the probability-of-commission estimates yielded results that closely paralleled those for the verdict–confidence measure—the only significant effect was for timing of the judge's instruction, $F(2, 101) = 2.92$, $p < .06$. Only subjects who received instructions before the evidence viewed the defendant as less likely to have committed the crime ($p < .05$) than did the noninstructed subjects (mean percentage estimates of 64% and 77.6%, respectively); those who were instructed after the reception of evidence ($M = 73.1\%$) did not lower their probability-of-commission estimates.

Contrary to predictions, no significant differences were obtained on evaluations of evidence strength, interpretations of reasonable doubt, or self-ratings of interest and involvement. Interestingly, a close look at the reasonable-doubt data indicates that the overall estimate of reasonable doubt ($M = 86.07$) was almost identical to that previously reported for college students (Simon & Mahan, 1971, obtained a general estimate of 87% and an estimate of 85% when the crime was auto theft). Subjects thus believed that there should be at least an 86% chance that the defendant committed the crime in order to vote for conviction.

Finally, subjects demonstrated a moderately high level of recall, averaging 12.15 correctly recalled items from a total of 16.[3] Further analysis indicated that a main effect was obtained for timing of instruction on the number of case-related facts recalled, $F(2, 101) = 3.26$, $p < .05$. Surprisingly, subjects in the instructions-after condition recalled fewer items ($M = 11.30$) than those in either the instructions-before ($M = 12.51$) or noninstructed ($M = 12.69$) conditions. The presence of the instruction between the evidence and response assessment apparently inhibited recall.

Table 2
Pattern of Mid-Trial Verdicts for the Three Instruction Groups at Each of the Six Decision Points

Instruction and verdict	Decision point					
	1	2	3	4	5	6
Before						
Guilty	8	8	8	7	5	8
Not guilty	10	10	10	11	13	10
After						
Guilty	16	16	17	16	13	15
Not guilty	3	3	2	3	6	4
None						
Guilty	15	14	15	14	14	12
Not guilty	2	3	2	3	3	5

Note. Decision points 1–6 immediately follow the direct examination and cross-examination of the prosecution's two witnesses and the defendant.

In sum, the mere timing of instructions had an impact on jurors' final judgments. Results for the verdict–confidence measure tended to support the major hypothesis that instructions on the requirements of proof would reduce subjects' tendencies to convict (i.e., relative to no instructions) only when delivered prior to the introduction of testimony. Although the timing variable had no significant effect on interpretations of reasonable doubt, preinstructed subjects actually viewed the defendant as less likely to have committed the crime than either the instructions-after or no-instructions subjects. Those who received instructions at the end of the trial did not respond differently on the verdict and probability-of-commission measures from those who were never instructed. In fact, they even recalled fewer of the case-related

[3] Only the major facts in the case were tested. Moreover, these facts were typically repeated throughout the trial proceedings (e.g., the highway on which Ron Oliver was stopped was first mentioned during the prosecutor's opening statement and was reiterated during the examination of both the arresting officer and the defendant and during the closing arguments). As a result, the facts tested could not be classified for further analysis as pro-prosecution or prodefense, nor could they be located at a single point in the trial.

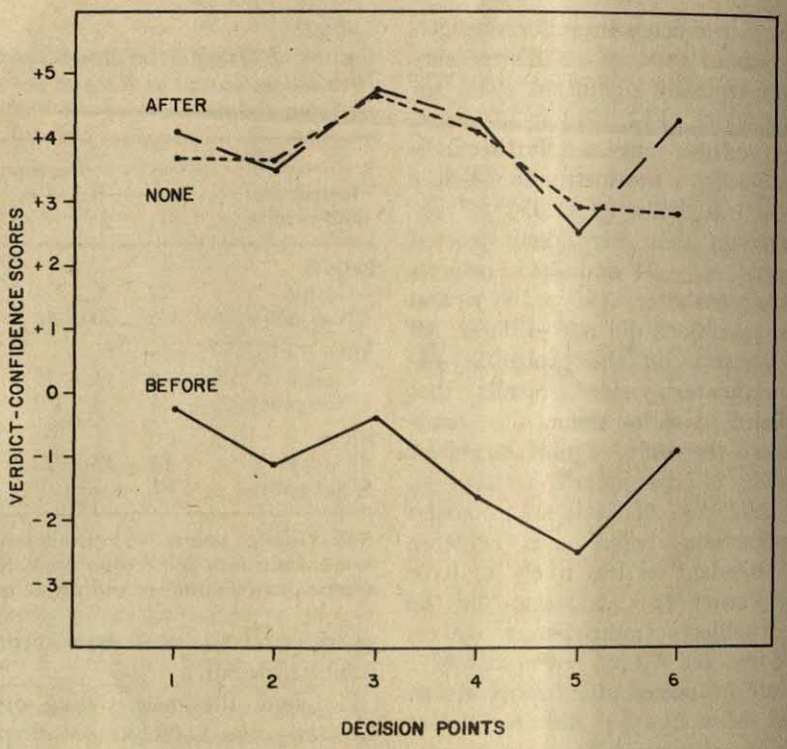


Figure 1. The pattern of verdict–confidence scores for the three groups at each of six decision points. (Higher scores indicate confidence in guilt.)

facts than all other subjects did. Finally, whether or not subjects made mid-trial judgments did not affect any of the above variables. The nonreactive nature of this multiple-judgment procedure was thus confirmed.

Process (Mid-Trial) Measures

Overall, 68% of the mid-trial responses were guilty verdicts. For the 54 subjects who made these judgments, the pattern of verdicts is presented in Table 2. As with the outcome data, these verdicts were analyzed with a likelihood ratio goodness-of-fit chi-square. Again, the simplest model that explained the observed frequencies was the two-way interaction between instructional set and verdicts, $\chi^2(30) = 6.50$, $p = 1.0$. Although a higher order model comprised of both Instruction × Verdict and Decision Point × Verdict interactions also fits the data, $\chi^2(20) = 3.13$, $p = 1.0$, it did not contribute significantly to the predictions made by the simpler model. Mid-trial verdicts were thus accounted for by th effect for instructional set. Table 2 shows th across all decision points the instruction before group yielded 41% guilty verdict compared to 82% in both the instruction after and no-instructions groups.

A verdict–confidence measure was agai created and was this time submitted to 3 (Instructions) × 6 (Decision Point) analy sis of variance. Figure 1 illustrates the stron main effect for the instruction factor, $F($ 51) = 10.15, $p < .001$, on these verdi scores. As before, this analysis clearly indi cated that subjects who were instructed to fore the evidence ($M = -1.14$) were con sistently less likely to convict the accused (p .001) than were either the instruction after or noninstructed subjects ($Ms = 3.0$ and 3.65, respectively). These latter group did not differ. An additional main effect f the decision point factor, $F(5, 255) = 4.$ $p < .001$, indicated, as expected, that verdi scores fluctuated widely as the trial pr

Table 3
Correlation Coefficients of Mid-Trial and Final Verdict–Confidence Scores in Each Group and Across All Groups

Group	Verdict–confidence score					
	1	2	3	4	5	6
Before	.14	.42	.57**	.46**	.42	.54*
After	.22	.22	.18	.15	−.03	−.15
None	.31	.30	.37	.42	.45	.86***
Overall	.28*	.41**	.47***	.44**	.36**	.52***

Note. Correlations are based on *n*s of 18, 19, and 17 ($N = 54$).
* $p < .05$. ** $p < .01$. *** $p < .001$.

gressed. As might be anticipated, the conviction rate was highest at the third point—right after the prosecutor examined his second witness ($p < .01$), and lowest at the fifth point—right after the direct examination of the defendant ($p < .01$). The interaction between instructional set and decision point did not approach significance ($F < .1$).

On the question of whether jurors' early impressions were predictive of their final predeliberation verdicts, Table 3 presents the correlations between subjects' mid-trial and final verdict–confidence scores. It can be seen that over all multiple-judgment groups the average correlation between mid-trial and posttrial verdict scores was .44 ($p < .01$). In fact, the average correlation between the verdict scores rendered at even the first decision point and those given at the end of the trial was .28 ($p < .05$). To some extent, then, subjects' decisions were substantially formed very early in the trial presentation.

Results for the repeated probability-of-commission measures followed a similar pattern (see Figure 2). A main effect for timing of instruction, $F(2, 5) = 6.58$, $p < .005$, re-

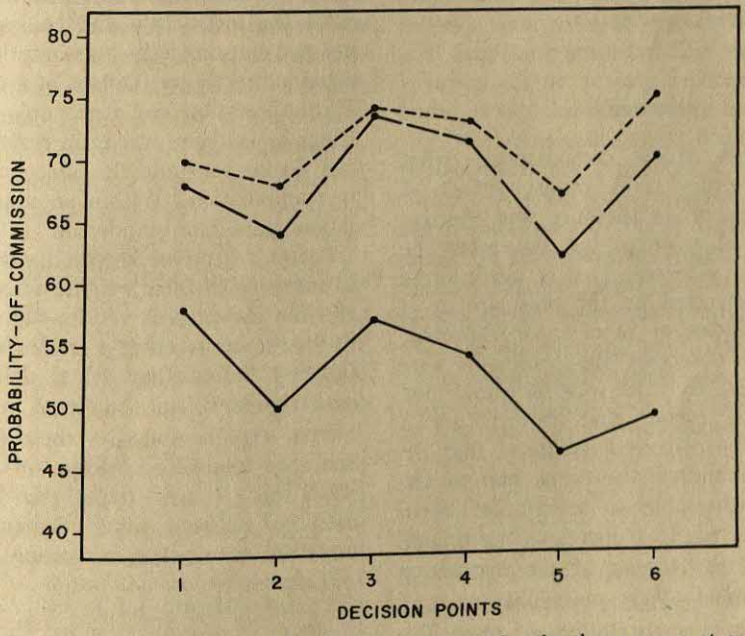


Figure 2. The pattern of probability-of-commission estimates for the three groups at each of six decision points.

vealed that during the trial, the instructions-before group ($M = 52.5\%$) viewed the defendant as less likely to have committed the crime ($p < .01$) than either the instructions-after or no-instructions groups (68.16% and 71.67%, respectively), who again did not differ from each other. Moreover, a main effect for the decision point factor, $F(5, 255) = 3.08$, $p < .01$, indicated that the perceived probability of commission was lowest at the fifth point—right after the defendant testified in his own behalf.

In sum, subjects' mid-trial judgments fluctuated in the predicted directions (i.e., guilty judgments after the prosecutor's examination and not-guilty judgments after the defense's examination). More important, these judgments were influenced largely by whether or not subjects had been instructed on the requirements of proof. That is, although instructional set and decision point did not interact (i.e., their curves were almost perfectly parallel, thereby indicating that pre-instructed subjects were neither less influenced by the prosecutor's testimony nor more influenced by the defendant's testimony), subjects who had been instructed before the evidence immediately and continually viewed the defendant as less likely to have committed the crime.

Discussion

The present study demonstrated what might be described as a primacy effect. A judge's instruction to the jury was effective when delivered prior to but not after the presentation of evidence. That is, mock jurors who were instructed on the presumption of innocence, burden of proof, and reasonable doubt before observing the testimony were ultimately less likely to vote for conviction. It was noted earlier that verdicts are a function of the perceived likelihood that the defendant committed the crime and of the standard or threshold to which that likelihood is compared. In the present experiment, variations in the timing of the instruction affected estimates of the probability of commission but not interpretations of reasonable doubt. Subjects who received the instructions before the evidence thus demonstrated a low rate of conviction because they actually viewed the defendant as less likely to have committed the crime.

Although the present experiment was not theoretically guided, it now appears that an information integration model of juror judgments (Kaplan & Kemmerick, 1974) might well describe the results. Briefly, each piece of evidentiary and nonevidentiary information possesses some scale value along a dimension of guilt–innocence and a weight that determines the importance of that information. A trial judgment is then formed on the basis of a weighted-average combination of stimulus components. In the context of the present study, subjects received two global categories of information: the trial presentation and judicial instruction. From the probability-of-commission responses of noninstructed subjects, the scale value of the trial information may be estimated at .78 (i.e., .78 is the marginal mean for noninstructed subjects). Although the scale value of the requirements of proof instruction was not assessed in the present study, it should—by legal ideal—have implied a scale value or initial probability of commission of 0 (Ostrom, Werner, & Saks, 1978). On the assumption that the scale values of the instruction and trial presentation remained constant (these assumptions are supported indirectly by the lack of a timing effect on interpretations of reasonable doubt and ratings of evidence strength, respectively), to what do we attribute the effects for the timing manipulation? Within an impression formation paradigm, Anderson (1965) found that later adjectives keep a fixed scale value but decrease in their weight. This mechanism describes the present results—the prodefendant instruction received a greater weight when delivered before than when delivered after the evidence.[4]

What *process* underlies these findings? At least two plausible mechanisms are worth evaluating. It was hypothesized that pre-instructed subjects would demand a greater burden of proof when evaluating the strength

[4] Note that although the present analysis treats judges' instructions as information that has a scale

of the prosecutor's case as it unfolds during the trial. This critical, evaluative "schema" was expected to manifest itself in different fluctuating patterns in the responses of the instructed (instructions-before) and noninstructed (instructions after and no instructions) mid-trial judgment groups. However, the absence of an Instruction × Decision Point interaction on verdicts and probability-of-commission ratings ruled out this hypothesis. In fact, the three instruction groups demonstrated remarkably similar mid-trial shifts in judgment (see Figure 1). Those who had been instructed were neither more influenced by the defense nor less influenced by the prosecutor than the others were. Instructed subjects thus did not reject or distort the prosecutor's evidence to fit an initial impression or presumption of innocence. And why should they? Luchins (1957) has demonstrated that by forewarning observers of the imminence of additional information, the proactive effects of early information on subsequent information are suppressed. Subjects in the present study and jurors in general *expect* to be confronted with inconsistent information. Early information in this setting (i.e., judicial instruction) is thus unlikely to produce the strong bias that would stimulate the processes of discounting or assimilation.

The process that did appear to operate was considerably less complex. Subjects who were instructed before the evidence were more likely than the others were to vote for acquittal and indicated a lower probability of commission *even from the first decision point*. Although they responded similarly to testimony that followed, their initial leanings resulted in fewer guilty verdicts at the conclusion of the trial. The impact of this initial reaction is reflected in the significant correlation between first and final verdict scores. Simply put, most preinstructed subjects "presumed innocent," whereas the others "presumed guilty."

The *practical* implications of the present study are straightforward. Mandatory code provisions do not forbid or discourage judges from using their discretion in instructing the jury before the trial. A number of states (e.g., Indiana, 1966), individual judges (e.g., Prettyman, 1960), and legal scholars (e.g., McBride, 1969) have thus adopted or advocated delivery of preliminary instructions prior to the presentation of evidence. In Missouri (1964), for example, the Supreme Court Committee on Jury Instructions recommended that "the jury be instructed before the trial begins. . . . The committee believes that it is better to draw the jury's attention to these matters before the trial rather than waiting until after the jurors may have reached a decision" (quoted in McBride, 1969, p. 62). This latter example is unfortunately an exception rather than the rule, which is that preliminary instructions are vastly underutilized (DeVitt & Blackmar, 1977). Viewed in this context, the present study provides firm support for proponents of procedural reform in the courtroom. The Anglo-American system of justice has traditionally favored the accused on the philosophy that acquitting a truly guilty person is better than convicting a truly innocent one (Ploscowe, 1935). Yet the present results suggest that the accused may not benefit from this protective instruction unless it is delivered to jurors *before* the trial. In this regard, perhaps the requirements of proof might well be included in the currently popular "juror handbooks" that are distributed to jurors as part of a pretrial orientation (National Institute of Law Enforcement and Criminal Justice, 1975) and might perhaps be presented at both the beginning and the end of the proceedings (Elwork et al., 1977).

From a methodological standpoint, the multiple-judgment procedure initially employed by Weld and Danzig (1940) merits increased consideration in future research. The present study confirmed that it does not bias jurors' ultimate verdicts or any other posttrial measures (also see Pyszczynski, Note

---

value and weight, Kaplan and his colleagues are quick to point out that strictly speaking, because instructions do not pertain to the specific defendant or crime, they do not relate to information scale value (S_i). Instead, instructions should affect an initial impression of the defendant that exists prior to jurors' receiving information about him or her (S_o), that is, the impression of defendants in general (see Kaplan & Miller, 1978).

1). Moreover, it appears to be valuable for examining jurors' reactions to the trial as it unfolds. Questions concerning the relative impact of different trial phases (e.g., opening statements and closing arguments, direct examination vs. cross-examination) and different kinds of evidence (e.g., tangible exhibits, expert or eyewitness testimony) may be fruitfully investigated through the mid-trial assessment of verdicts.

Finally, some of the limitations of the present results deserve mention. As with other nonevidentiary factors, the timing of an instruction cannot be expected to influence the outcome of a one-sided case (i.e., an extremely strong or weak case against the defendant). Rather, the effects should be considered limited to close, ambiguous cases. Note also that the present experiment assessed the verdicts of nondeliberating jurors. Whether or not deliberation provides enough of a corrective strategy to erase the timing effect upon juries remains open to question. Finally, the trial used in the present study was much shorter than the average criminal case. Whether or not the timing manipulation would affect jurors' responses in realistically longer trials remains to be seen.

Reference Note

1. Pyszczynski, T. A. *The effects of opening statements on liability judgments at various points in a simulated jury trial.* Unpublished master's thesis, University of Kansas, 1978.

References

Anderson, N. H. Primacy effects in personality impression formation using a generalized order effect paradigm. *Journal of Personality and Social Psychology,* 1965, *2,* 1–9.

Davis, J. H., Bray, R. M., & Holt, R. W. The empirical study of decision processes in juries: A critical review. In J. L. Tapp & F. J. Levine (Eds.), *Law, justice, and the individual in society: Psychological and legal issues.* New York: Holt, Rinehart & Winston, 1977.

DeVitt, E. J., & Blackmar, C. B. *Federal jury practice and instructions* (Vols. 1–3). St. Paul, Minn.: West Publishing, 1977.

Elwork, A., Sales, B. D., & Alfini, J. J. Juridic decisions: In ignorance of the law or in light of it? *Law and Human Behavior,* 1977, *1,* 163–189.

Fienberg, S. E. *The analysis of cross-classified data.* Cambridge, Mass.: MIT Press, 1977.

Frank, J. *Courts on trial: Myth and reality in American justice.* Princeton, N.J.: Princeton University Press, 1949.

Hervey, J. C. Jurors look at our judges. *Oklahoma Bar Association Journal,* 1947, *25,* 1508–1513.

Indiana pattern jury instructions. Indianapolis, Ind.: Bobbs-Merrill, 1966.

Jones, E. E., & Goethals, G. R. *Order effects in impression formation: Attribution context and the nature of the entity.* Morristown, N.J.: General Learning Press, 1971.

Jones, E. E., Rock, L., Shaver, K. G., Goethals, G. R., & Ward, L. M. Pattern of performance and ability attribution: An unexpected primacy effect. *Journal of Personality and Social Psychology,* 1968, *10,* 317–340.

Juhnke, R., Vought, C., Pyszczynski, T. A., Dane, F. C., Losure, B. D., & Wrightsman, L. S. Effects of presentation mode upon mock jurors' reactions to a trial. *Personality and Social Psychology Bulletin,* 1979, *5,* 36–39.

Kalven, H., & Zeisel, H. *The American jury.* Boston: Little, Brown, 1966.

Kaplan, M. F., & Kemmerick, E. D. Juror judgment as information integration: Combining evidential and nonevidential information. *Journal of Personality and Social Psychology,* 1974, *30,* 493–499.

Kaplan, M. F., & Miller, L. E. Reducing the effects of juror bias. *Journal of Personality and Social Psychology,* 1978, *36,* 1443–1455.

Kerr, N. L., Atkin, R. S., Stasser, G., Meek, D., Holt, R. W., & Davis, J. H. Guilt beyond a reasonable doubt: Effects of concept definition and assigned decision rule on the judgments of mock jurors. *Journal of Personality and Social Psychology,* 1976, *34,* 282–294.

LaBuy, W. J. *Jury instructions in federal criminal cases.* St. Paul, Minn.: West Publishing, 1963.

Luchins, A. S. Primacy-recency in impression formation. In C. I. Hovland (Ed.), *The order of presentation and persuasion.* New Haven, Conn.: Yale University Press, 1957.

McBride, R. L. *The art of instructing the jury.* Cincinnati, Ohio: W. H. Anderson, 1969.

McCart, S. W. *Trial by jury.* Philadelphia: Chilton Books, 1964.

Missouri Supreme Court Committee on Jury Instructions. *Missouri approved jury instructions.* New York: Vernon Law Book Company, 1964.

National Institute of Law Enforcement and Criminal Justice. *A guide to jury system management.* Washington, D.C.: U.S. Department of Justice, 1975.

Ostrom, T. M., Werner, C., & Saks, M. J. An integration theory analysis of jurors' presumptions of guilt or innocence. *Journal of Personality and Social Psychology,* 1978, *36,* 436–450.

Ploscowe, M. The development of present-day criminal procedures in Europe and America. *Harvard Law Review,* 1935, *48,* 433–473.

Prettyman, E. B. Jury instructions—First or last? *American Bar Association Journal,* 1960, *46,* 1066.

Sealy, A. P., & Cornish, W. R. Juries and the rules of evidence. *Criminal Law Review,* April 1973, 208–223.

Simon, R. J., & Mahan, L. Quantifying burdens of proof: A view from the bench, the jury, and the classroom. *Law and Society Review,* February 1971, 319–330.

Walker, L., Thibaut, J., & Andreoli, V. Order of presentation at trial. *Yale Law Review,* 1972, *82,* 216–226.

Weld, H. P., & Danzig, E. R. A study of the way in which a verdict is reached by a jury. *American Journal of Psychology,* 1940, *53,* 518–536.

Received December 11, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 10, 1888-1901

Expectancy Shifts and the Expectancy Confidence Hypothesis

Richard W. Wollert
Portland State University

Explanations of differences in the size of expectancy changes following task outcomes were considered. The *control perception* hypothesis, the most frequently proposed explanation, is that small expectancy shifts occur when task outcomes are perceived to be externally (i.e., chance) controlled. An alternative explanation, the *expectancy confidence* hypothesis, is that small shifts occur when subjects are relatively confident of the accuracy of their expectations. Two experiments examined these positions. Experiment 1 partially replicated a study often cited as supporting the control perception hypothesis. Expectancy confidence was assessed, and as predicted by the expectancy confidence hypothesis, expectancy shifts were found to be related negatively to expectancy confidence. Skill perceptions and levels of expectancy confidence were manipulated in Experiment 2, and their impact was assessed by several expectancy shift measures. Expectancy confidence was found to influence expectancy shifts as predicted for four of five measures, whereas skill perceptions did not significantly affect expectancy shifts on any measure. Expectancy confidence thus exerts a substantial impact upon expectancy shifts. The relevance of the findings for a third explanation of expectancy shifts, the *causal stability* hypothesis, is discussed, as are the implications of the expectancy confidence hypothesis for theories of personality and depression.

For the last 25 years, expectancy for reinforcement has remained a fundamental concept of Rotter's social learning theory of personality (Rotter, 1954; Rotter, Chance, & Phares, 1972). A corollary of Rotter's belief that expectancies influence personality processes is that changes in these expectancies are important psychological phenomena. In Rotter's analysis, expectancy changes are considered to reflect learning about behavioral outcome relationships and to be precursors o behavior change.

Because of their significance for learnin and behavior change, many attempts hav been made to outline the conditions underl ing changes in expectancies for success. Tw general findings serve to put these efforts i broad perspective. First, subjects typicall raise their expectations after successfully per forming a task and lower them after unsuc cessful performance (cf. Rotter, Liverant, & Crowne, 1961). Second, large differences in expectancy changes exist between subject even where initial expectancy levels have bee equated and the pattern of task outcomes ha been controlled (cf. Miller & Seligman, 1973) Taken together, such findings indicate tha recent and direct task experience exerts a powerful influence on expectancy changes bu is not sufficient to account for all between subjects differences.

Rotter has proposed that several cognitive variables, termed "generalized expectancies

Some of the research reported in this article was part of a doctoral dissertation submitted to the Graduate School of Indiana University and was supported by a research grant from the Graduate School. Portions of this article were presented at the annual conventions of the Western Psychological Association, Seattle, 1977, and San Diego, 1979. The author is indebted to Alexander M. Buchwald, chairman of his dissertation committee, and to other committee members, Leon H. Levy, Richard D. Young, and Paul Gebhard. Many thanks also are extended to Bob Knight and Michele Smith for their assistance with the pilot studies, and to Stephen West for his helpful comments on the manuscript.

Requests for reprints should be sent to Richard W. Wollert, Department of Psychology, Portland State University, Portland, Oregon 97207.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3710-1888$00.75

account for these individual differences. The most well known of these is internal versus external locus of control of reinforcement. According to Rotter (Rotter et al., 1972), small changes in verbalized expectancies occur where subjects perceive reinforcements to be externally rather than internally controlled. This view, which will hereafter be referred to as the *control perception* hypothesis, has been supported by several findings (Miller & Seligman, 1973; Phares, 1957; Rotter et al., 1961) that subjects faced with chance tasks (i.e., where task outcomes were perceived to be externally controlled) changed their expectancies of success less than subjects faced with skill tasks (i.e., where task outcomes were perceived to be internally controlled).

This hypothesis has had a significant impact on many areas of psychology. It has, as already noted, been used in social and personality psychology research to explain why subjects learn more about behavior-outcome relationships on some tasks than on others (Rotter et al., 1961). It has also been applied to the field of psychotherapy, where increases in a client's internal locus of control are often seen to precede both expectancy shifts and positive behavior changes (Rotter et al., 1972). Still a third application has come in the area of psychopathology. Seligman and his associates have found that depressives change their expectancies for success on skill tasks less than nondepressives (Klein & Seligman, 1976; Miller & Seligman, 1973). Relying on the control perception hypothesis, they have interpreted this finding to mean that depressives perceive reinforcement to be under greater external control than nondepressives do. This interpretation is consistent with Seligman's learned helplessness theory of depression (Seligman, 1975).

It should be recognized that these conclusions are justified only as long as the control perception viewpoint provides the most adequate explanation of expectancy shifts. This paper reviews and critiques some of the research supporting the control perception hypothesis and presents an alternative theory of expectancy shifts called the *expectancy confidence* hypothesis. The results of several studies supporting this alternative are then reported. Finally, the implications of the expectancy confidence hypothesis are discussed for theory and research on personality and abnormal psychology.

Review and Critique of the Control Perception Hypothesis

The two studies most frequently cited in support of the control perception hypothesis were conducted by Phares (1957) and Rotter, Liverant, and Crowne (1961). In the first study, Phares manipulated perceptions of control of reinforcement by giving subjects different pretask inductions. Phares discovered that smaller expectancy shifts were reported when a task was preceded by a chance induction rather than a skill induction.

In the second study, Rotter et al. assigned subjects to either a prediction or motor task. It was assumed that, "on the basis of previous cultural experiences" (Lefcourt, 1966, p. 208), subjects would perceive outcomes of the prediction task to be chance dependent. Outcomes were controlled so that success occurred on the first trial of each task. The prediction of Rotter et al. that chance task subjects would change their expectancies less than the other subjects was subsequently confirmed by several expectancy shift measures: (a) from the first to the second trial; (b) over all trials in which shifts in the appropriate direction occurred (i.e., upward or downward shifts in expectancies following success or failure, respectively, on the immediately preceding trial); (c) the final expectancy, recorded after the eighth trial. On the bases of these and earlier results, Rotter et al. concluded that the control perception hypothesis had been strongly supported.

The control perception interpretation of these two studies may be questioned on several grounds. A prime objection to Phares' approach is that neither initial expectancy differences nor expectations for improvement were assessed. However, the inductions placed such different emphases on task difficulty that they may have differentially affected initial expectations. Skill inductions, for example, pointed out that subjects performed at various levels, whereas chance inductions indicated that the task was "so difficult that it

did not differentiate between people . . ." (Phares, 1957, p. 340). Failure to assess initial expectancies and expectations for improvement is a serious weakness, given these inductions, since they conceivably created expectancy differences that later affected the size of expectancy changes.

Significant objections may also be raised to the methodology employed by Rotter et al. (1961). No manipulation check was conducted, for example, to determine if subjects perceived the tasks as skill or chance. It is consequently unclear whether the tasks were actually perceived as determined by skill or by chance. Equally important, the skill task differed greatly from the chance task. The possibility thus exists that task factors other than skill or chance perceptions may have led to the reported expectancy shift differences.

To these methodological drawbacks may be added the criticism that the control perception hypothesis has not received consistent empirical support. James and Rotter (1958) had subjects perform identical prediction tasks but manipulated perceptions of control by skill and chance inductions. They found that "the acquisition (i.e., raw expectancy) curves of the 100% (reinforcement) groups are fairly similar, as are the 50% groups" (p. 400). If raw expectancy curves are similar, there is no possibility that expectancy shifts may differ. This follows because differences in expectancy shifts may be defined as differences in the slopes of two expectancy curves at the same point. The findings of James and Rotter thus conflict with the findings of Phares (1957).

This lack of empirical support is apparent in correlational as well as experimental studies. Miller and Seligman (1973), found no relationship between expectancy shifts and personality measures that assess generalized perceptions of internal and external control of reinforcement. Consistent with this result, other authors have concluded that perception of control, when treated as a personality dimension, rarely has been found to be related to expectancy shifts (Weiner, Nierenberg, & Goldstein, 1976).

A final reservation applies not only to the research of Phares (1957) and Rotter et al. (1961) but to almost all studies of expectations for reinforcement. As defined by Rotter, expectancy for reinforcement is the "probability held by the individual that a particular reinforcement will occur as a function of a specific behavior on his part in a specific situation or situations" (Rotter et al., 1972, p. 12). In spite of this definition, however, subjects in expectations studies have consistently been asked to report how *sure* or how *certain* they were that they would succeed on a task. This question is not the same as asking subjects to report their *chances* of success. In fact, it may be argued that this method of assessment did not focus upon expectancies of success, but on how confident subjects were that an expectancy of success having the probability of 1 was accurate. Previous studies, therefore, may have confounded subjective expectancies of success with confidence in the accuracy of these expectations.

To summarize, in some cases adequate control has not been exercised over the effects of different tasks on expectancy shifts. Where tasks have been controlled, pretask inductions may have created initial differences in expectations for success and improvement. In addition, the results of different studies have conflicted. Finally, the assessment of subjective expectancies seems to have been confounded consistently with assessing confidence in the accuracy of these expectancies. Given these difficulties, it may be argued that the cognitive factors most importantly involved in expectancy shifts have not yet been clearly isolated.

The Expectancy Confidence Hypothesis

A possible alternative explanation of the findings of Phares (1957) and Rotter et al. (1961) stems from the expectancy confidence hypothesis. This hypothesis is based on the premise that individuals not only possess beliefs about their probabilities of success or reinforcement but also hold beliefs about the accuracy of these expectations. Expectancy confidence may be defined in terms of this certainty or, more formally, as the probability held by an individual that his or her expectancy is correct. For example, when such

jects hold that only expectations in a narrow range are plausibly accurate estimates of their chances of success, they possess high levels of expectancy confidence. Alternatively, when subjects will accept many expectations as plausible estimates, they have low levels of confidence.

If we contrast locus of control with expectancy confidence, a fundamental conceptual difference between approaches becomes apparent. While locus of control emphasizes specific beliefs, expectancy confidence emphasizes the relationships between cognitions in terms of certainty as defined by the range of plausible alternatives. Adopting Levy's (1970) terminology, locus of control may be considered to deal with the content of cognitions, whereas expectancy confidence deals more with the structure of cognitions.

The central proposition of the expectancy confidence hypothesis reflects this structural emphasis. That proposition is that the degree of change in expectancies that occurs over time or events is a function of expectancy confidence. When expectancy confidence is low, expectancies are likely to change; when expectancy confidence is high, expectancies are likely to remain stable. Applied to the findings of Phares (1957) and Rotter et al. (1961), the expectancy confidence hypothesis suggests that chance condition subjects may have been more confident of their expectancies than other subjects were. An analysis of some of the experimental procedures followed in these studies underscores the plausibility of this suggestion. For example, Rotter et al. (1961) informed chance condition subjects that their objective was to predict correctly whether slides marked with an *X* or an *O* would be projected upon a screen. In all likelihood, subjects had had some previous experience with gambling games or other situations similar to this task (e.g., predicting the toss of a coin). In the skill condition, the goal was to raise a steel ball resting on a movable platform above a certain height (Sky, 1950). Subjects in this condition probably had almost no prior experience with similar games or situations. Greater task experience presumably induced a relatively high level of confidence on the part of chance condition subjects that their initial expectations of success were correct. Since they were more confident of their expectations, subjects in the chance condition were less likely to change these expectations than were subjects in the skill condition.

In the Phares (1957) study, chance inductions stressed both task difficulty and lack of variation in task performance. It seems that this would lead chance condition subjects to expect to perform poorly and also to believe that their performance would not improve. In other words, chance subjects may have held more negative expectations than skill subjects held and may also have believed that these expectations were accurate estimates of their objective probability of success. With respect to this analysis, it should be pointed out that no mention of task difficulty or performance variation was included in the chance induction used by James and Rotter (1958). Levels of expectancy confidence may consequently have been somewhat similar for their skill and chance subjects. This could account for the lack of expectancy shift differences they found.

Data from several pilot studies conducted by the present author support the logic of this analysis. In one, the apparatus and descriptions for the tasks used by Rotter et al. (1961) were presented in a group setting to 49 undergraduates. Since a slide projector was unavailable, the prediction task was slightly modified so that sample letters were presented on 4″ × 6″ (10 × 15 cm) file cards, and subjects were told that their task was to guess correctly the letters on sets of file cards (rather than on slides). All subjects picked one of the tasks to complete this sentence: "I have more general experience with tasks similar to the [platform lifting or prediction] task." As predicted, a greater proportion of students (.76) reported more experience with the prediction task ($p < .001$).

In a second study, the color and line tasks and inductions used by Phares were presented to a group of 40 undergraduates. The color task was paired with the chance induction, and the line task with the skill induction. Subjects were asked to circle their best estimate of their chances of success on an 11-

point scale ranging from 0 to 100%. They also checked all the alternative estimates they considered plausible. This made it possible to determine their highest and lowest plausible chances of success as well as their best estimate. As predicted, paired t tests showed that all three expectancy variables were less for the chance induction task than for the skill induction task (all p's $\leq .001$). When we paired the color task with the skill induction and the line task with the chance induction in a later administration to 50 undergraduates, this finding was replicated.

In the following two experiments, the expectancy confidence hypothesis has been examined in a more precise fashion and under more carefully controlled conditions.

Experiment 1

In an effort to determine whether the expectancy confidence hypothesis provides a plausible explanation of the findings of Rotter et al. (1961), a partial replication of this study was conducted. The following predictions, based upon the expectancy confidence model, were tested:

1. Subjects presented with the chance task would report that they were more confident of their initial expectations than would subjects assigned to the skill condition. This follows from the expectancy confidence position that subjects who report small expectancy shifts should also be confident that their initial expectations are correct.

2. Expectancy shifts from Trial 1 to Trial 2 would be correlated negatively with initial measures of expectancy confidence within conditions. In other words, the negative relationship between expectancy confidence and expectancy shifts was expected to hold within conditions as well as between conditions.

3. Expectancy shifts from Trial n to Trial $n + 1$ would be correlated negatively with expectancy confidence on Trial n for the subjects within each condition; that is, the relationship between expectancy confidence and expectancy shifts was expected to hold within subjects as well as within conditions.

4. Trend analysis of expectancy confidence over all trials would show an increase in confidence within the skill condition. This hypothesis was based on the presumption that subjects are usually unfamiliar with the task used in this condition and that expectancy confidence will increase as experience increases. A similar hypothesis was not stated for subjects in the chance condition, as it was assumed that they had had more prior experience with the chance task.

Method

Subjects

Subjects were 51 female college students enrolled in introductory psychology courses. Female subjects were used because Rotter et al. used females and because this increased the homogeneity of the subject samples. Eleven of these subjects were eliminated from the data analysis: 4 because of apparatus failure, 5 because they suspected the skill task was unfair, 1 because she suspected the chance task was unfair, and 1 because she did not perform the skill task as directed.

Apparatus

Chance task. Slides were projected onto a screen 8 feet (2.46 m) in front of the subject. Subjects were informed that some of these slides were imprinted with an *X* and that others were imprinted with an *O*. Before a slide was projected, subjects were to guess whether an *X* or an *O* would appear. To be successful on a "trial," a subject had to predict correctly four out of five presentations.

Skill task. The skill task, hereafter referred to as the "Sky" task, involved an apparatus similar to that originally developed by Sky (1950) and later used by Rotter et al. (1961). Each subject was requested to raise a 4-inch (10 cm) square platform upon which a ½-inch (13 mm) steel ball bearing had been placed. The platform was to be raised by gently pulling on a string connected to the platform and threaded through a pulley at the top of the Sky apparatus. To be successful, it was necessary to lift the platform 9 cm without the ball rolling over one of the edges. The ball, positioned by the experimenter, was held in place by an electromagnet implanted on the underside of the platform. Current to this device could be interrupted by depression of a concealed, silent solenoid switch. Since the platform of the apparatus sloped forward very slightly, the ball would roll off the platform when such an interruption occurred. Outcomes could therefore be surreptitiously controlled by the experimenter.

Assessment. To report expectancies of success, subjects answered the following question using an 11-point scale running from 0 to 100% (i.e., 0%, 10%, 20%, etc.): "What are the chances that you will succeed on this task? In other words, if you did this

task over and over, what percent of the time would you be successful?"

To report expectancy confidence, subjects answered the following question on a 5-point scale ranging from "very good idea" to "almost no idea": "How good an idea do you really have of your actual chances of success?"

Subjects reported perceptions of control of reinforcement by responding to the following question on an 11-point scale ranging from 0 (success depends on chance) to 10 (success depends on skill): "How important is skill or chance in producing success on this task?"

Procedure

Twenty subjects were randomly assigned to each condition. Data were collected for each subject individually.

Subjects were seated and a prepared script[1] was read by the experimenter. The first section of the script described whichever task the subject was to receive. Subjects were then given printed forms of the questions assessing expectancies and confidence and were provided with examples of how they might answer these if they were to report high or low ratings of expectancies of success or expectancy confidence. After the experimenter answered any remaining questions, subjects recorded their answers to the questions. The question assessing perceptions of control of reinforcement was then explained, and a rating was obtained from the subject.

After completing the third question, subjects were given eight trials on either the chance or skill task. Success was controlled at a rate of 50%,[2] and all subjects were presented with an identical pattern of outcomes (success, failure, failure, success, success, failure, failure, success). In the chance task, the experimenter exerted control by either advancing or reversing the slide tray in response to the subject's predictions. In the skill task, outcomes were controlled by depression of the silent solenoid switch. The outcome of every trial (i.e., success or failure) was announced to subjects immediately after it was completed. Each subject then reported her expectancy of success and expectancy confidence for the next trial. After the last trial, a final rating of perception of control of reinforcement was obtained in addition to the other ratings.

Following the data collection period, subjects were asked what they thought was the hypothesis of the experiment, whether they fully understood the questions, and whether they believed that the task was unfair. No subjects correctly guessed any of the experimental hypotheses or indicated that they did not understand the questions. Data for subjects who believed the task was unfair (e.g., some thought that the platform was not level or believed that the experimenter had some means of controlling the outcomes) were eliminated from the final analysis.

Results

Before data analysis, indices of expectations of success and expectancy confidence were produced, as follows:

Expectations of success for Trial n = raw expectations score × 10;
expectancy confidence for Trial n = raw expectancy confidence score/4 × 10.

These transformations simply allow all responses to be presented as if they had come from 11-point scales, where 0 represented minimal ratings of skill perception, expectancy confidence, or expectations of success.

Preliminary data analysis[3] indicated that the three dependent measures were not intercorrelated significantly before Trial 1 within either condition (product–moment r values between .00 and .23). Consequently, pretask expectancies of success, levels of expectancy confidence, and perceptions of skill requirements are assumed to represent relatively separate dimensions.

Further data analysis, summarized in Table 1, pointed to the conclusion that the results of Rotter et al. (1961) were replicated in the current study. Using t tests for independent means, we found that subjects assigned to the two groups did not differ in their initial expectations of success, $t(38) = 1.77$, $p > .05$. However, skill condition subjects displayed significantly greater expectancy shifts than did chance task subjects from Trial 1 to Trial 2, $t(38) = 4.66$, $p < .001$. In addition, subjects did perceive the Sky task as requiring greater skill than the prediction task, both before the first trial, $t(38) = 5.58$, $p < .001$, and following the last trial, $t(38) = 4.25$, $p < .001$.

More important for the expectancy confidence position, Hypothesis 1 was confirmed. That is, subjects in the chance condition were

---

[1] The scripts used in both experiments have not been included because of space limitations. They are available from the author.

[2] When expectancy shifts under four different reinforcement conditions (25, 50, 75, and 100%) were observed, Rotter et al. (1961) discovered that similar shift discrepancies took place under all conditions. Therefore, to focus upon only a single condition of success was sufficient for the present study.

[3] All tests of significance were two-tailed.

Table 1
Means and Standard Deviations for Skill and Chance Condition Subjects in Experiment 1

Measure	Chance		Skill		
	M	*SD*	*M*	*SD*	*t*
Expectations of success	3.45	1.28	4.30	1.72	1.77
Expectancy shifts, Trials 1 to 2	.75	.97	2.55	1.43	4.66**
Skill perceptions, pretask	1.90	1.33	4.85	1.95	5.58**
Skill perceptions, posttask	3.30	2.20	5.90	1.61	4.25**
Expectancy confidence	3.88	2.50	1.62	2.03	3.13*

Note. All *t* tests for independent means were two-tailed. $n = 20$ in each group.
* $p < .01$. ** $p < .001$.

found to be significantly more confident of their expectancies than were subjects in the skill condition, $t(38) = 3.13$, $p < .01$. Hypothesis 2 was partially confirmed, since a significant negative product-moment correlation was found between expectancy confidence on Trial 1 and expectancy shifts from Trial 1 to Trial 2 within the skill condition, $r(18) = -.50$, $p = .02$, although not within the chance condition, $r(18) = .16$, $p > .05$. Subsequent analysis revealed that the variance of initial shifts for the chance group was significantly less than that for the skill group, $F(19, 19) = 2.68$, $p < .05$, which may account for the low and nonsignificant correlation found for the chance group. No significant relationships between initial skill perceptions and initial expectancy shifts were found within either the skill condition, $r(18) = -.21$, $p > .05$, or the chance condition, $r(18) = .35$, $p > .05$.

Table 2
Means and Standard Deviations for Chance and Skill Condition Subjects in Experiment 1 for all Experimental Trials

Trial	Outcome	Expectancy confidence			
		Chance		Skill	
		M	*SD*	*M*	*SD*
1**	Success	3.88	2.50	1.62	2.03
2	Failure	4.50	2.08	5.12	1.51
3	Failure	5.00	2.29	5.00	1.40
4	Success	4.25	2.82	5.00	1.81
5	Success	5.00	3.03	5.75	1.43
6*	Failure	5.38	2.84	7.00	1.54
7	Failure	5.25	3.13	6.62	1.86
8*	Success	4.50	3.10	6.25	2.07
9*	—	4.88	3.09	6.88	1.79

Note. All *t* tests for independent means were two-tailed. $n = 20$ in each group.
* $p < .05$. ** $p < .01$.

To test Hypothesis 3, product–moment correlations between expectancy confidence on Trial n and expectancy shifts from Trial n to Trial $n + 1$ were calculated for each subject over all eight trials. Correlations were transformed according to Fisher's r-to-z transformation, and the z values were averaged within each condition (McNemar, 1962). By testing for the significance of each z_{av}, the significance levels of the corresponding r_{av} values were determined. Hypothesis 3 was partially confirmed, since a significant negative average correlation was found within skill condition subjects, $r_{av} = -.49$, $p = .007$. A similar relationship was not found for chance condition subjects, $r_{av} = .16$, $p > .05$. Three subjects in the chance condition did not display any variance in expectancy shifts. Since the correlation is undefined in such cases, the data from these subjects did not contribute to the average correlations for the chance condition.

A test for linear trend revealed that expectancy confidence increased over the course of eight trials for subjects assigned to the skill condition, $F(1, 171) = 79.52$, $p < .001$. No such changes were noted for chance condition subjects, $F(1, 171) = 1.38$, $p > .05$. The fourth hypothesis, therefore, was also confirmed. Mean levels of expectancy confidence for each trial are shown in Table 2.

In order to analyze expectancy shifts most fully, it would be important to consider the fluctuations in expectancy shifts that occurred over the course of the experiment. An analysis of variance of a two-factor mixed design with

repeated measures on one factor revealed significant main effects for conditions, $F(1, 38) = 13.90$, $p < .001$, and trials, $F(7, 266) = 13.5$, $p < .001$, and a significant Conditions × Trials interaction, $F(7, 266) = 6.2$, $p < .001$. Inspection of the mean trial by trial expectancy shifts revealed that greater expectancy shifts occurred for skill task subjects than for chance task subjects over the first three or four trials and from Trial 1 to Trial 2 in particular. Differences in expectancy shifts decreased after the first few trials, so that they were virtually of the same magnitude for both groups by the end of eight trials.

In sum, the results of Experiment 1 confirmed, or partially confirmed, each hypothesis of the study.

Discussion

An analysis of expectancy shift research suggested that high expectancy confidence leads to small expectancy changes. This analysis also suggested that in an important expectancy shift study (Rotter et al., 1961), chance task subjects were more confident than skill task subjects that their initial expectancies were correct.

The first experiment evaluated these contentions. In accordance with the expectancy confidence hypothesis, subjects performing the prediction task in Experiment 1 were found to possess higher levels of expectancy confidence than subjects performing the platform task. Product-moment correlations also indicated that differences in expectancy confidence were related to expectancy shifts for subjects in the Sky task. Although similar differences were not found for prediction task subjects, subsequent analysis revealed that the variance of initial expectancy shifts was significantly less for subjects in this condition. McNemar (1962) has pointed out that the magnitude of a correlation coefficient is lessened when the range of either variable is curtailed. Therefore, although the relationship between expectancy confidence and expectancy shifts under chance conditions warrants examination in future studies, this explanation plausibly accounts for the correlational differences between skill and chance conditions in Experiment 1.

The first experiment also determined if subjects performing the Sky and prediction tasks perceived outcomes to be differentially dependent upon skill or chance factors as asserted by Rotter et al. (1961). Subjects did perceive the Sky task as requiring more skill than the prediction task. Skill perceptions were not, however, correlated with expectancy changes. In addition, the convergence of expectancy shifts reflected by the significant Conditions × Trials interaction is at odds with the control perception hypothesis. Since subjects rated the Sky task as requiring more skill than the prediction task, both before and after completing the tasks, the control perception hypothesis would seem to suggest that some differences in the size of expectancy shifts should have been observed throughout the experiment.

Experiment 2

Experiment 2 was an attempt to make comparative tests of predictions about expectancy shifts based on the expectancy confidence and control perception hypotheses. All subjects performed on the skill task used in Experiment 1. Expectations of success were recorded over several trials, thereby allowing for the calculation of expectancy shift measures similar to those frequently used by other researchers (cf. Klein & Seligman, 1976; Miller & Seligman, 1973; Rotter et al., 1961). Skill perceptions were manipulated by pretask inductions so that one group of subjects perceived outcomes to be skill dependent (high skill perception), whereas another group perceived outcomes to be chance controlled (low skill perception). Levels of expectancy confidence were also manipulated by pretask inductions so that one group of subjects was relatively certain that their expectancies of success were correct (high expectancy confidence), whereas another group was less certain of this (low expectancy confidence). These locus of control and confidence manipulations were combined factorially. The following hypotheses were tested:

1. Smaller expectancy shifts will be reported by subjects assigned to the low skill conditions than by those in the high skill conditions (the control perception prediction).

2. Smaller expectancy shifts will be reported by subjects assigned to the high expectancy confidence conditions than by those in the low expectancy confidence conditions (the expectancy confidence prediction).

Method

Subjects

Subjects were 97 female college students enrolled in introductory psychology courses. Seventeen subjects were eliminated from the data analysis: 1 because of apparatus failure, 2 because they did not operate the apparatus according to instructions, 7 because they suspected the task was not fair, and 5 because they reported that they either did not answer the questions properly (e.g., estimated expectancies on a trial by trial rather than on a long-term basis) or did not understand the objective of the task.

Instructions

All of the pretask instructions given to subjects performing on the skill task in Experiment 1 were presented to subjects participating in Experiment 2. In addition, instructions designed to induce particular perceptions of control of reinforcement and levels of expectancy confidence were delivered. These inductions had been found in pilot studies to be sufficiently powerful to create significant differences between groups with respect to skill perceptions and levels of expectancy confidence. The high skill subjects were informed that how they performed was determined by many factors under their control. The low skill subjects were told that how they performed was determined by many factors outside their control. The high confidence subjects were told that, of the college students who had performed the Sky task many times, 50% were successful more than half the time, and the other 50% were successful less than half the time. The low confidence subjects were told that 25% of their peers were successful on the task almost all the time, 25% were successful over half the time, 25% were successful less than half the time, and 25% were almost never successful. To clarify this information, these subjects were also presented with a figure indicating the rectangular distribution of performance scores. These confidence inductions were used on the reasoning that subjects given a performance distribution broken down into several categories would be less sure where their average performance scores might fall than other subjects given only two categories. In order to deliver the appropriate inductions in as uniform a manner as possible, one of four videotapes was presented to each subject.

The questions and scales used in Experiment 1 to assess perceptions of control of reinforcement, expectancy confidence, and expectations of success were also used in Experiment 2.

Procedures

Twenty subjects were assigned randomly to each of the four induction conditions. Data were collected for each subject individually. Since the data collection period spanned two quarters, a randomized block design (Winer, 1971) was adopted to control for fluctuations in subject characteristics.

After being seated at a table upon which the Sky apparatus had been placed, each subject was presented with videotaped instructions. After the videotaped instructions were presented, subjects performed the Sky task. The object of each trial was the same as that described in Experiment 1. Each subject participated in eight trials, with the pattern of outcomes being surreptitiously controlled by the experimenter, as in Experiment 1. Expectancies of success and expectancy confidence were reported before each trial and after the final trial. Skill perceptions were reported before the first trial and after the last trial.

Subjects were asked the debriefing questions used in Experiment 1. No subjects correctly guessed any of the hypotheses of the experiment, although several indicated that they believed the outcomes were controlled. Data for these latter subjects were eliminated from the final analysis.

Results

The analysis of variance applied to the data was a randomized block design with replications as an additional factor (Winer, 1971). Therefore, wherever analyses of variance were undertaken in Experiment 2, the following three factors were included: skill perception (two levels), expectancy confidence (two levels), and replications (four levels). Before data analysis, raw scores were transformed as in Experiment 1.

To make meaningful comparisons between groups, it was necessary first to determine whether the instructions were successful in producing the intended differences and also whether there were differences in initial expectancies that might account for later differences. When skill perceptions were analyzed, high skill subjects indicated that a higher level of skill was required for successful task performance than that indicated by low skill subjects, $F(1, 64) = 8.20$, $p = .006$. On measures of expectancy confidence, high confidence subjects were more confident than low confidence subjects that their initial expectancies were correct, $F(1, 64) = 11.36$, $p = .002$. All other main effects and interactions in these two analyses had F values less than 1.0.

Table 3
Means and Standard Deviations of Dependent Measures for Experiment 2

Measure	Experimental condition			
	Low skill–low confidence	Low skill–high confidence	High skill–low confidence	High skill–high confidence
Change, Trials 1 to 2				
M	2.20	1.45	2.85	1.90
SD	1.24	1.15	1.53	1.45
Total change				
M	5.70	3.70	6.70	4.50
SD	3.10	2.89	3.08	2.98
Total increase				
M	3.55	2.30	4.05	2.80
SD	1.96	1.75	2.21	1.99
Total decrease				
M	2.15	1.40	2.65	1.70
SD	1.39	1.31	1.63	1.26
Final expectancy				
M	5.80	5.45	5.15	5.60
SD	1.28	1.28	1.09	1.88

Note. $n = 20$ for all conditions.

Initial expectations of success were not influenced significantly by either the skill, $F(1, 64) = .42$, $p > .05$, or confidence, $F(1, 64) = 1.18$, $p > .05$, inductions, and a significant interaction was not found between inductions, $F(1, 64) = .13$, $p > .05$. Taken together, the preliminary analyses suggest that the expectancy shifts observed in Experiment 2 may be attributed to the manipulations rather than to some uncontrolled factor.

Measures employed to evaluate expectancy shifts included the following: (a) change in expectancies between Trials 1 and 2; (b) total increases in expectancies, defined as the absolute sum of the changes that occurred immediately following successful outcomes; (c) total decrease in expectancies, defined as the absolute sum of the changes that immediately followed failure experiences; (d) total amount of expectancy change, defined as the sum of total increases and decreases in expectancies; and (e) the final expectancy.

Table 3 contains the means and standard deviations of the dependent measures. Significant main effects were found for the confidence induction in the direction predicted by the expectancy confidence hypothesis for changes in expectancies from Trial 1 to Trial 2, $F(1, 64) = 8.91$, $p = .004$; total amount of expectancy change, $F(1, 64) = 8.95$, $p = .004$; total increase in expectancies, $F(1, 64) = 7.36$, $p = .008$; and total decrease in expectancies, $F(1, 64) = 7.05$, $p = .01$. The only measure not reflecting such differences was the final expectancy reported by subjects, $F(1, 64) = .02$, $p > .05$. A main effect for the skill induction approached the .05 level for changes in expectancies from Trial 1 to Trial 2, $F(1, 64) = 3.33$, $p = .07$. However, no significant main effects for the skill induction were found for any of the other dependent measures ($p > .20$ in all cases). In addition, no significant interactions were found ($p > .20$ in all cases).

Discussion

In Experiment 2, high and low levels of expectancy confidence were factorially combined with high and low skill perceptions. Comparative predictions based on the expectancy confidence and control perception hypotheses were tested. The prediction based upon the expectancy confidence hypothesis was supported by the finding that subjects with high levels of confidence changed their expectancies

less than subjects with low levels. Although high skill subjects reported that the Sky task required more skill than the low skill subjects said it did, skill perception differences did not significantly affect expectancy shifts. The prediction based upon the control perception hypothesis was therefore not supported.

General Discussion

Considering the contributions of each experiment separately, Experiment 1 pointed up naturally occurring relationships between variations in expectancy confidence and variations in expectancy shifts. Experiment 2 showed similar effects for manipulated levels of expectancy confidence. Taken together, the results provide strong support for the expectancy confidence hypothesis but not for the control perception hypothesis.

Generality of Applicability

The clear predominance of support given the expectancy confidence hypothesis raises the question whether all expectancy shift differences previously explained by the control perception hypothesis could now be accounted for by the expectancy confidence hypothesis. For example, could the findings of Rotter et al. (1961) be considered, on the basis of the results of Experiment 2, to be an artifact produced by dissimilar tasks? This probably is too strong a statement in view of the fact that skill perception differences were larger in Experiment 1, the replication of Rotter et al., than in Experiment 2. This observation suggests that the size of expectancy shifts could be affected differentially where large skill perception differences exist.

Another consideration that argues against premature closure on this issue is that a main effect for the skill induction in Experiment 2 approached the .05 level of significance for one expectancy shift measure. Several explanations of this finding are plausible. One straightforward account is that it may be understood in terms of random fluctuation. This view is supported by the fact that the use of five expectancy shift measures greatly increased the chances that one of the several measures would have a low p value. Another obvious possible explanation is that, under certain circumstances, perceptions of control may influence expectancy shifts independently of expectancy confidence. Confirmation of this possibility would limit the generality of the expectancy confidence hypothesis. This would not, however, negate the conclusion that the concept of expectancy confidence has a useful "range of convenience" (Kelly, 1955) for the comprehension and prediction of some behaviors.

In evaluating the generality of the expectancy confidence hypothesis, some attention should be given to other expectancy shift theories. Weiner's attributional theory is particularly relevant in this regard, since it has been explicitly offered to account for the results obtained by Rotter and his colleagues (Weiner, Frieze, Kukla, Reed, Rest, & Rosenbaum, 1971; Weiner, Heckhausen, Meyer, & Cook, 1972; Weiner, Nierenberg, & Goldstein, 1976). Weiner (Weiner et al. 1976) reviewed several empirical studies of the relationship between expectancy shifts and causal attributions that were stable and internal (skill), stable and external (task difficulty), unstable and internal (effort), and unstable and external (luck). He concluded that the greatest shifts occurred for subjects who attributed their outcomes to stable factors rather than to internal factors. Causal stability, as opposed to control perception, is thus considered by Weiner to be the major dimension underlying expectancy shifts.

While powerful tests of the expectancy confidence and causal stability hypotheses must await future research efforts, it is possible to make some comparative statements about these theories on the basis of the present findings. Weiner has argued that the skill and chance tasks used in Experiment 1 lead subjects to make stable and unstable attributions, respectively (cf. Weiner et al., 1972). He would presumably advance a similar interpretation of the effects of the skill and chance inductions used in Experiment 2. Such assertions would suggest many of the same hypotheses as those based on the control perception hypothesis:

1. Skill perceptions in Experiment 1 should

be positively correlated with the size of expectancy shifts.

2. Differences in expectancy shifts should be observed for skill and chance subjects in Experiment 1 on later as well as earlier trials.

3. A significant main effect for skill perceptions should be present in Experiment 2.

Since these hypotheses were not confirmed, whereas those derived from the expectancy confidence hypothesis were confirmed, the results of Experiments 1 and 2 support the expectancy confidence hypothesis but not the causal stability hypothesis.

A final theoretical dimension that should be related to the generality of the expectancy confidence hypothesis is that of commitment. Mischel (1958) and Watt (1965) have asserted that verbalization and discussion of initial expectancies increase the level of commitment to the specific values of these expectancies. Expectancy changes may therefore be small where commitment is high. Applying this formulation to the present research, it could be argued that reporting high levels of confidence induced high levels of expectancy commitment and that this led to small expectancy changes.

Although this explanation could account for the results of Experiment 2 and the correlational findings of Experiment 1, it should be pointed out that the relationship between commitment and expressions of confidence was not empirically determined in the present studies. Furthermore, by itself, a commitment hypothesis does not provide a clear explanation of the shift differences that have been found when confidence was not assessed and when commitment was presumably equalized as a result (cf. Miller & Seligman, 1973; Phares, 1957; Rotter, Liverant, & Crowne, 1961).

It is clear from this discussion that many variables may be involved in the production of expectancy shifts. It is also apparent that some post hoc alternative explanations may be formulated to account for the present results. In view of these circumstances, an advisible future research course would include developing several methods of manipulating important variables, devising unobtrusive measures for assessing these manipulations, and then completing additional comparative studies.

Implications

Methodological. Applications of the expectancy confidence hypothesis may be made to several areas. Perhaps the most straightforward of these falls on a methodological level. The nonsignificant correlations between initial expectancies and expectancy confidence in both experiments and the lack of initial expectancy differences between confidence conditions in Experiment 2 indicate that expectancies were independent of expectancy confidence. Unfortunately, the measurement of expectancies has consistently been confounded with expectancy confidence in previous research. This makes it difficult to interpret clearly many previously reported findings. Furthermore, from the standpoint of future research, it would seem extremely important to assess expectancy confidence separately whenever expectancies are assessed.

Theoretical. The expectancy confidence hypothesis also holds implications for personality and psychotherapy theories using expectancy constructs (e.g., Bandura, 1977; Rotter, 1954), in that it provides a new explanation of changes in expectancies and behavior. Rotter, for example, has proposed that behavior is a function of expectancies for reinforcement and reinforcement value (Rotter, 1954; Rotter et al., 1972). Where reinforcement value remains constant, changes in expectancies for reinforcement have typically been considered necessary preconditions for behavior change. A formulation of behavior change that takes into account the importance of expectancy confidence suggests that low levels of expectancy confidence are antecedents of changes in expectancies and behavior.

Another application of the expectancy confidence hypothesis lies in the reinterpretations it provides of previous research based on the control perception hypothesis. As pointed out in an earlier section, it not only accounts for the findings of Phares (1957) and Rotter et al. (1961) but reconciles these findings with those of James and Rotter (1958). It also holds implications for research on the learned-

helplessness model of depression. Seligman's (1975) interpretation of the low expectancy shifts of depressed undergraduates as evidence for this model rests not only on the explicit assumption that expectancy shifts are greater for skill tasks than chance tasks but also on the implicit assumption that there are no other sources of differences in the size of expectancy shifts. The present results contradict these assumptions by indicating that expectancy confidence is an important source of differences that has not previously been taken into consideration. The present results thus call into question evidence that has been thought of as providing strong support for the learned-helplessness model.

An alternative explanation of the association between depression and small expectancy shifts is that depressives shift their expectations less than nondepressives because they are more confident of their expectancies. At first, this might seem illogical. However, it appears more reasonable when the nature of the certainty associated with depression is considered. The findings of Friedman (1964) and Loeb, Beck, and Diggory (1971) indicate that one of the hallmarks of clinical depression is extreme negativity regarding self-evaluation and expectations of successful performance. Beck (1967) has also observed that depressives cling to such beliefs with remarkable tenacity. Negative expectations of success and negative evaluations of performance have also been reported for populations of depressed undergraduates (Wollert & Buchwald, 1979). On the bases of these reports and the results of the present investigation, it is proposed that depressives shift their expectations less than nondepressives because they are relatively certain that they will neither succeed nor improve when confronted with a skill task.

The expectancy confidence hypothesis provides another explanation of expectancy shift differences and points to several important reinterpretations of previous results. It also holds implications for theories of personality, behavior change, and depression. This hypothesis emphasizes the structure of cognitions and seemingly provides additional breadth to personality theories based on expectancy constructs. To realize this potential, however, it will be necessary to integrate the construct of expectancy confidence with existing theories. Consideration of those factors possibly affecting expectancy confidence would also be required. Such factors include experience, task perceptions, the degree of internal consistency within an expectancy network, the number of plausible alternative expectancies, the availability of meaningful information, and the operation of interpersonal influence factors. In view of the numerous applications of the expectancy confidence hypothesis, further research along these lines would certainly appear warranted.

References

Bandura, A. Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 1977, *84*, 191–215.

Beck, A. T. *Depression*. Philadelphia: University of Pennsylvania Press, 1967.

Friedman, A. S. Minimal effects of severe depression on cognitive functioning. *Journal of Abnormal and Social Psychology*, 1964, *69*, 237–243.

James, W., & Rotter, J. B. Partial and 100% reinforcement under chance and skill conditions. *Journal of Experimental Psychology*, 1958, *55*, 397–403.

Kelly, G. A. *The psychology of personal constructs*. New York: Norton, 1955.

Klein, D. C., & Seligman, M. E. P. Reversal of performance deficits in learned helplessness and depression. *Journal of Abnormal Psychology*, 1976, *85*, 11–26.

Lefcourt, H. M. Internal versus external control of reinforcement: A review. *Psychological Bulletin*, 1966, *65*, 206–220.

Levy, L. H. *Conceptions of personality*. New York: Random House, 1970.

Loeb, A., Beck, A. T., & Diggory, J. Differential effects of success and failure on depressed and nondepressed patients. *Journal of Nervous and Mental Disease*, 1971, *152*, 106–114.

McNemar, Q. *Psychological statistics*. New York: Wiley, 1962.

Miller, W. R., & Seligman, M. E. P. Depression and the perception of reinforcement. *Journal of Abnormal Psychology*, 1973, *82*, 62–73.

Mischel, W. The effect of the commitment situation on the generalization of expectancies. *Journal of Personality*, 1958, *26*, 508–516.

Phares, E. J. Expectancy change in skill and chance situations. *Journal of Abnormal and Social Psychology*, 1957, *54*, 339–342.

Rotter, J. B. *Social learning and clinical psychology*. New York: Prentice-Hall, 1954.

Rotter, J. B., Chance, J. E., & Phares, E. J. *Ap-*

plications of a social learning theory of personality. New York: Holt, Rinehart & Winston, 1972.

Rotter, J. B., Liverant, S., & Crowne, D. P. The growth and extinction of expectancies in chance controlled and skilled tasks. *Journal of Psychology,* 1961, *52,* 161–177.

Seligman, M. E. P. *Helplessness: On depression, development, and death.* San Francisco: Freeman, 1975.

Sky, A. W. An apparatus for a frustration task. *Australian Journal of Psychology,* 1950, *2,* 116–120.

Watt, N. F. Public commitment, delay after commitment, and change in verbalized expectancies. *Journal of Personality,* 1965, *33,* 285–299.

Weiner, B., Frieze, I., Kukla, A., Reed, L., Rest, S., & Rosenbaum, R. M. *Perceiving the causes of success and failure.* Morristown N.J.: General Learning Press, 1971.

Weiner, B., Heckhausen, H., Meyer, W., & Cook, R. E. Causal ascriptions and achievement behavior: A conceptual analysis of effort and reanalysis of locus of control. *Journal of Personality and Social Psychology,* 1972, *21,* 239–248.

Weiner, B., Nierenberg, R., & Goldstein, M. Social learning (locus of control) versus attributional (causal stability) interpretations of expectancy of success. *Journal of Personality,* 1976, *44,* 52–68.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1971.

Wollert, R. W., & Buchwald, A. M. Subclinical depression and performance expectations, evaluations of performance, and actual performance. *Journal of Nervous and Mental Disease,* 1979, *167,* 237–242.

Received December 20, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 10, 1902-1914

Components of Aggression in Chickens and Conceptualizations of Aggression in General

D. W. Rajecki, David R. Nerenz, Terry G. Freedenberg, and Patricia J. McCarthy
University of Wisconsin—Madison

A refined analysis of the peck order in chickens was offered as a test of the notion that for this species, different responses such as leaping and various types of pecking need not be interchangeable indexes of aggression. Indeed, tests showed that particular response types of the birds were differentially mediated by organismic or environmental factors. In large cages pecking at the body was most frequent by birds that had a home-cage advantage. Contrarily, rates of aggressive leaping were independent of this environmental influence, with males having an advantage over females. Males showed more head pecking than females, but the profile for this sex difference did not resemble the profile for leaping. Correlational analyses revealed that whereas head pecking between testmates was not matched in frequency, leaping was positively related. Finally, the behavior of birds tested in small cages differed from that of the large-cage subjects. Although there was more head pecking in the small cages, males did not have an edge, and leaping was infrequent. Such results indicate that these responses cannot be viewed as interchangeable indicators of aggression in fowl.

Our thesis is that certain indexes of presumably aggressive behavior cannot be used interchangeably in an uncritical manner. Where they exist, separate influences on different measures will require identification. Indeed, there is controversy over procedures and standards for classifying particular responses as aggressive in both the human (cf. Baron, 1977; Tedeschi, Smith, & Brown, 1974) and nonhuman (cf. Deag, 1977; Rowell, 1974; Syme, 1974) research domains. In terms of the latter literature, evidence is beginning to accumulate that in analyzing the social psychology of animals certain "components," "forms," or "types" of aggression (labels used by different writers) should not be used interchangeably.

To take the literature on rodents as an example, in intact rats there was no relationship between success in intraspecies fighting (in dominance encounters) and mouse killing (Baenninger & Baenninger, 1970), and although spontaneous dominance orders and limited access (competitive) dominance orders were reliable within themselves, they showed very low interorder correlation (Baenninger, 1970). Further, rats that would not normally engage in fighting (females and males familiar with one another) could be induced to fight in competition for food (Zook & Adams, 1975). Sensory alterations and other surgical preparations have also revealed differences in components of rat aggression. Whereas blinded rats fought at control levels in a shock situation, anosmic and devibrissaed pairs fought less, and olfactory bulbectomized animals showed an increase in mouse killing (cf. Bugbee & Eichelman, 1972; Ghiselli & Thor, 1975; Thor, 1976). Moreover, it has been demonstrated that shock-induced aggression

This research was supported by Grant 190245 from the Graduate School Research Committee of the University of Wisconsin. The authors are indebted to Marlowe Embree, Kurt Steuck, Therese Durkin, and Carol Sawyer for assistance in the research. In particular, the authors thank Mr. and Mrs. George Fujioka for determining the sex of day-old chicks.

Requests for reprints should be sent to D. W. Rajecki, Department of Psychology, University of Northern Iowa, Cedar Falls, Iowa, 50613.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3710-1902$00.75

is separable from territorial aggression in rats in terms of the differential influence of hypothalamic lesions on these two response types (Adams, 1971) and the differing behavioral profiles of intact animals across the two situations (Blanchard, Blanchard, & Takahashi, 1978).

Studies with mice parallel the rat literature (Fredericson, 1952; Rowe & Edwards, 1971), and a set of researches indicates that the components or patterns of aggressive displays in certain fishes are not uniformly influenced by experiential and environmental factors (Clayton & Hinde, 1968; Davis, 1975; Shapiro & Schuckman, 1971). Further, success in peck exchanges between young chicks does not correlate with their success in limited access tests (Rajecki, Nerenz, Barnes, Ivins, & Rein, 1977). Taken together, this growing body of evidence suggests that in the study of any species it may be necessary to forego the assumption that there is some unitary factor in all modes of aggression. This consideration brought us to an assessment of components of aggression in young domestic chickens.

In fact, a detailed analysis of aggression in chickens is quite warranted. Although the term *peck order* is widely used by both scientist and layman, there is surprisingly little agreement as to just what sort of pecking or other form of behavior might constitute an aggressive response in young precocial birds. Some investigators have regarded any intersubject pecking as aggressive (ducklings and quail chicks: Eiserer, Emerling, Scardina, & Hoffman, 1976); others have restricted attention to pecks that cause the withdrawal of the opponent (ducklings: Hoffman, Boskoff, Eiserer, & Klein, 1975), whereas still others have concentrated on the locus of pecking (chicks: Rajecki, Ivins, & Rein, 1976).

However, it may be premature to categorize any activity on the part of developing birds as aggressive without a better knowledge of the role of organismic and environmental factors in such behavior or a fuller understanding of the mutual influence between birds engaged in particular response types. Further, although seminal sources have suggested that certain patterns of agonistic behavior emerge at different points in the development of fowl (e.g., Guhl, 1958; Kruijt, 1964), detailed information concerning the ontogeny of such responses is necessary before a taxonomy of components can be established. Therefore, the research reported here was undertaken to identify certain biological, maturational, and experiential influences on the capacity for aggressive social behavior among broodmates. The aim was to compare and contrast particular response types as possible components of aggressive behavior in young fowl.

Method

Design

The design of this longitudinal study permitted an assessment of the influence of particular factors on discrete components of chicks' social behavior. Earlier reports indicated that sex is a strong determinant of differential status between birds (Wood-Gush, 1971), so pairs consisting of a female and a male were tested. In supplemental tests, same-sex pairs were observed for aggression. Other studies of young chickens indicated that in encounters involving social pecking, there is a kind of "prior resident" or "home-cage" advantage. That is, birds tested in the home cage (or in a situation containing some feature of the home cage) generally have an edge over birds introduced to that home cage (Rajecki, Grams, Stursa, & Nerenz, 1978; Rajecki, Lamb, & Suomi, 1978; Rajecki, Nerenz, & Rein, 1978). Therefore, a "home–away" distinction was included in the current design. Further, it is clear that the physical dimensions of the test area have an influence on the likelihood of aggression in fowl, with less aggression seen in more cramped places (Hughes & Wood-Gush, 1977). Accordingly, chicks were reared in relatively large or small cages.

Finally, a recent report shows that intersubject pecking between pairmates can be promoted by repeatedly separating and reuniting cagemates for day-long periods (Rajecki, Lamb, & Soomi, 1978), so the current subjects were tested in at least 10 different reunions (following a period of physical separation) over the first 3 weeks after hatching. In sum, the basic design was a completely crossed $2 \times 2 \times 2$ ($\times 10$) factorial representing sex, the home–away distinction, cage size, and a number of repeated measures. Records were kept of a variety of responses.

In addition to observing the influence of the experimental conditions on social responses of developing chicks, the covariance of given components was assessed to provide further information on the similarities and differences between components.

Subjects

Subjects were 84 cockerels and 84 pullets of the White Leghorn strain, received on the day of hatching from the Sunnyside Hatchery in Oregon,

Wisconsin. The sex of the chicks had been determined at the hatchery by professional poultrymen using a method of cloacal extrusion, and sex was reestimated in the laboratory when the animals were 3 to 4 weeks of age by an examination of comb size and color. The two methods of determining sex agreed in 92% of the cases of 3-week-old birds and in 97% of the cases of 4-week-old birds, so the original method was used to classify subjects. Birds were housed in mixed-sex pairs and were marked with a felt pen for identification.

Large cages. Thirty-six pairs of chicks were assigned to 46 × 46 × 61 cm metal cages, each continuously illuminated and heated to 32–35 °C by a shaded 150-W lamp installed outside the wire mesh door. After several days the original bulbs were replaced by 40-W lamps that produced a midcage temperature of 24–26 °C. Each cage was also equipped with a plastic or metal trough for food (Purina chick starter) and a watering station consisting of an inverted quart jar and a plastic basin. A 30 × 30 cm sheet of absorbent paper (Cagesorb) was placed on the floor of the cage near the lamp and was liberally sprinkled with food prior to the arrival of the residents. The purpose of the paper was to insure that chicks would begin to peck at food particles, and to provide a warm substrate for the very young birds. Regardless of its physical condition, the original sheet of paper remained in the large cages for the duration of the experiment. These cages were arranged in racks of six in several laboratory rooms.

Small cages. Forty-eight pairs of chicks were assigned to 22.5 × 17.5 × 17.5 cm metal cages which contained two 5-cm petri dishes covered with wire netting, one for water and one for chick starter. Each cage was continuously illuminated and was heated to 34 °C by a 40-W bulb outside the mesh wall. After several days the original lamps were replaced with 15-W bulbs that reduced the midcage temperature to 24 °C. The small cages were arranged in racks of 12 in a single laboratory room.

Procedure

Experimental conditions. The chicks were delivered to the laboratory in large groups in separate cartons of cockerels and pullets. They were assigned to cages in pairs comprised of a male and a female.[1] Because the birds were to be repeatedly separated and reunited over the course of the experiment, one chick was designated the home bird, the other the away bird, and each was marked accordingly. The home bird remained in the home cage during separations, whereas the away bird was installed in a similar cage in an adjacent rack. Each away bird had its own away cage. The same bird was always the away bird, and birds were always reunited with their cagemate in the home cage. Across pairs an equal number of males and females were designated as away birds in both the large- and small-cage conditions.

Test sequence. The subjects were tested in two large batches during late spring and early summer. For a given test of a pair, the animals were separated for 24 hours and were then reunited. Since the act of returning the away bird involved handling the animal, the home bird was also handled. Both birds were removed from their respective cages at the same moment, and were simultaneously replaced in the home cage. Observations began immediately upon reunion, and continued for 5 minutes (after Rajecki, Lamb, & Suomi, 1978). Following the observation period the pair remained together for 24 hours and then underwent a series of such separation–reunion episodes until the completion of testing.

Since a given pair could be tested only every other day because of the separation period, a convenient schedule was adopted whereby half the sample was tested each and every day. That is, on the day each batch arrived, half the sample was placed in pairs in home cages, whereas the other half of the sample was immediately "separated," with the home bird placed in the home cage and the away bird placed in the away cage. These separated chicks were reunited and observed 24 hours later, at which point in time the birds that had been originally paired were separated for 24 hours. Thereafter these two subsets of the sample were alternately separated and reunited such that half the birds were observed in reunions on odd days of age (1, 3, 5, 7, and so on for those immediately separated) and half were observed on even days of age (2, 4, 6, 8, and so on for those initially paired). All observations were made between 9:00 a.m. and 1:00 p.m. local time. Testing continued in this fashion for about 3 weeks in the case of the subjects in the small cages and for about 4 weeks for the large-cage subjects. This procedure resulted in a within-subjects series of 10 and 14 observations per pair in the small and large cages, respectively. Although we wished to maximize the number of tests in order to detect effects for the variables under scrutiny, the testing in the small cages was terminated when the birds outgrew the units, and large-cage testing ceased when these cages were needed for a new batch of chicks that had been scheduled for delivery on a certain date.

Dependent Measures

Responses during a test were recorded by two observers, each responsible for a single bird. Eight

[1] It should be noted that the physical assignment of the birds to conditions was done by volunteers who had no other connection with the project, and the record of those assignments was not viewed by any observer until the completion of all testing for a given batch. Accordingly, at least during the early phase of this longitudinal study, an observer was blind to the sex of his or her subject (although observers were obviously aware of the birds' cage size and of whether their subject was a home or away bird). By about two weeks of age birds began to show signs of sexual dimorphism.

dependent measures were employed. Four of these measures were based on the locus of subjects' pecks and included pecks at the head, body, or feet of the testmate, or at any feature of the test unit. Discrete pecks were recorded by pressing corresponding buttons on hand-held panels that were connected to a multichannel event recorder.

A fifth response was aggressive leaping which consisted of a hop or a very rapid charging motion toward the testmate, with the leaping bird's wings flapping (cf. Kruijt, 1964, Figures 1 and 7; and the descriptions of sparring by Dawson & Siegel, 1967; and Ratner, 1965). These responses very often ended with the subject colliding with the testmate or landing with one or both feet on some part of the testmate. Each discrete leap was recorded with the press of a button.

Two other responses were recorded to measure states of disturbance on the part of the subjects (after Rajecki, Suomi, Scott, & Campbell, 1977). The first of these was the distress call, which is a long, loud note that is accompanied by the opening of the mandibles. Discrete calls were recorded. The other disturbance reaction was labeled *cage pushing*, in which a bird appeared to be motivated to escape the test situation by pushing with its beak at the walls of the cage (as distinct from pecking) or by trying to wriggle through the gaps in the mesh front. The button representing this response was pressed for the duration of an episode of cage pushing. The recording chart on which the responses were recorded was calibrated in 2-sec intervals, thus cage pushing will be discussed in terms of a time sample of 2-sec intervals per test in which the response was observed. On this index, maximum score per test is 150.

An eighth response was termed *immobility*, in which a bird rested on its breast feathers and held its eyes closed for at least several seconds. For this response, as for cage pushing, a time sample of 2-second intervals was obtained.[2]

Supplemental Tests in the Large Cages

As testing proceeded, casual observations indicated a decline in head pecking over days, and an inverted-U pattern for aggressive leaping (and see Results). These changes could be interpreted in a number of ways, so an additional test was proposed that might help evaluate competing accounts for the time-dependent shifts in responding. Supplemental tests of 14 pairs of birds[3] housed in the large cages took place after the final routine reunion. These birds were tested with strangers to determine if they would react in a xenophobic fashion (cf. Rajecki, Kidd, & Ivins, 1976; Rajecki, Ivins, & Kidd, 1977). Subsequent to the regular reunion on Days 27–28, normal procedures were followed whereby cagemates were allowed to remain together for 24 hours and were then separated for 24 hours. A 15th test of these birds was conducted at the end of this final day-long separation, but birds were not reunited with their cagemates. Rather, they were confronted with a same-sex bird from some other cage. A male bird that was an away bird was placed in the home cage of a home male, and the same procedure was followed for females. Seven pairs of male strangers and 7 pairs of female strangers were observed. Test procedures were identical in every way to those employed routinely.

Results

Subject Losses

Thirty-six pairs of chicks were assigned to the large cages, but 3 pairs did not survive the experiment. Of the remaining 33 pairs, 17 had males as the home bird and the rest had females as the home bird. Of the 48 pairs of birds assigned to the small cages, 2 pairs did not survive, thus in 24 of the remaining small-cage pairs the female was the home bird.

The small cages provided suboptimal conditions for the locomotor development of chicks. By the end of the third week, 16 of the 46 cages contained lame birds. Of these, 11 were male and 5 were female. The analyses of variance reported below were computed twice, once with the lame birds' scores included and a second time with these pairs

---

[2] Interobserver agreement using push-button panels is not problematic for experienced observers (see Rajecki, Ivins, & Rein, 1976), but new personnel were involved in this project, so reliability was checked. Observations seemed more difficult in the smaller cages, so all reliability estimates were made with extraexperimental birds in the small units. On several occasions pairs of observers watched the same bird and recorded its responses covertly. Average agreement on these occasions generally proved satisfactory for head pecks ($M = .95$, range $= .88$ to 1.00), body pecks ($M = .84$, range $= .67$ to 1.00), cage pecks ($M = .95$, range $= .88$ to 1.00), distress calls ($M = .91$, range $= .79$ to 1.00), and aggressive leaps ($M = .86$, range $= .78$ to .93). Cage pushing and foot pecking never occurred during reliability checks, but we have no reason to believe that observers were any less reliable on these measures than on the others. As a final step to insure accuracy during tests, observers instructed one another to record responses on occasions when a particular observer had his or her view of a subject blocked by obstructions such as the large water fount in the larger cages.

[3] The decision to conduct the supplemental tests was made at a time when the 14 pairs noted were the only birds remaining in the laboratory.

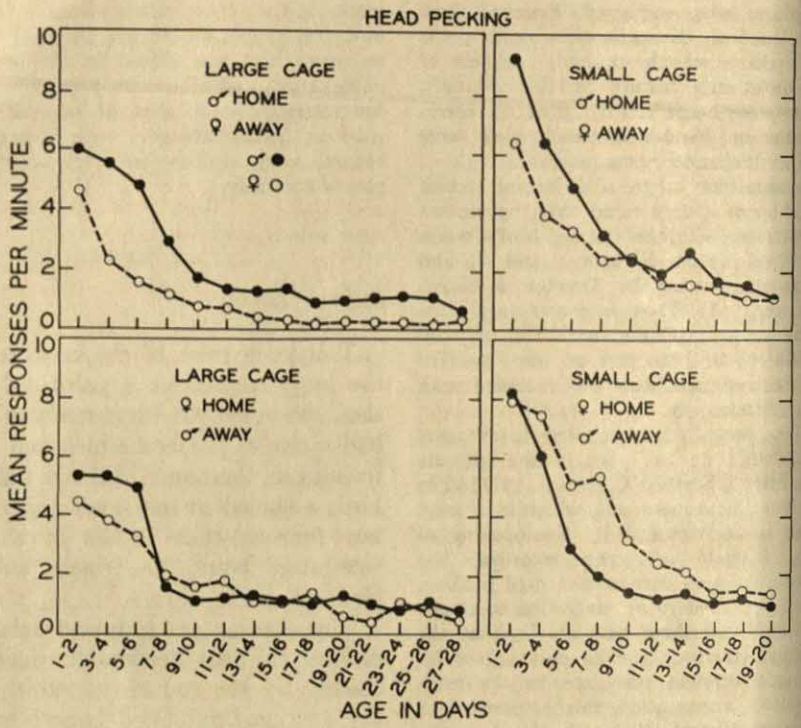


Figure 1. Head pecks per minute as a function of experimental conditions.

excluded. The two versions of each analysis were identical with respect to the direction and significance levels of results, so in this section the data from the complete sample are reported.

Analyses of Variance

Each response type was analyzed separately in terms of the $2 \times 2 \times 2$ ($\times 10$) design outlined in the method section. The birds in the large cages were tested on several more occasions than were the small-cage birds, but for purposes of comparison, only the first 10 test days of large-cage birds were contrasted with the 10 tests of birds in small cages. However, for clarity, all 14 tests of the former subjects are plotted in the figures to follow, where it can be seen that there were no substantial changes in trends established by the tenth test.

In these (and subsequent) analyses the scores for birds tested on Day 1 and Day 2 were combined, since these days represented the first test for both subsets of subjects. This combinaiton procedure was carried out for Day 3 and Day 4 scores (second test, respectively), Day 5 and Day 6 scores (third test, respectively), and so on for the remaining test days. The average scores on these combined test days are plotted in terms of entries for Day 1–2, Day 3–4, and so on, in the figures below.

Finally, the score of the individual bird was considered the unit of analysis, as this was the only way to preserve the distinction between the sexes, and the home–away distinction. However, mutual influence within pairs is of interest and will be considered in the section on correlational analyses.

Head pecking. The average numbers of head pecks per minute are shown in Figure 1. Overall, there were more head pecks in the small cages than in the large cages, $F(1, 150) = 17.55$, $p < .01$, but this difference attenuated later in testing as indicated by a Cage Size × Days interaction, $F(9, 1350) = 3.10$, $p < .01$.

Another clear finding is that home birds gave more head pecks than away birds, $F(1, 150) = 10.94$, $p < .01$. This home–away effect was not influenced by cage size, as indi-

cated by an insignificant interaction between these variables, $F(1, 150) = .92$, *ns*, and sex had little impact on the home–away distinctions, as revealed by an insignificant Sex × Home–Away interaction, $F(1, 150) = 1.02$, *ns*.

On head pecking the sexes were not clearly different overall, $F(1, 150) = 2.87$, $p > .05$. There was, however, an interesting Sex × Cage Size interaction, $F(1, 150) = 4.02$, $p < .05$, indicating that males enjoyed more of an advantage in the large cages than in the small cages. Further, holding everything else equal, a Sex × Days interaction indicated that males had an advantage early in the test sequence, but less so later on, $F(9, 1350) = 2.37$, $p < .05$.

Finally, there was a general decline in head pecking over days, $F(9, 1350) = 79.02$, $p < .01$.

Body pecking. Body pecks per minute are shown in Figure 2. Neither cage size, $F(1, 150) = .51$, *ns*, nor sex, $F(1, 150) = .14$, *ns*, had an influence on this response type. However, as in the case of head pecking, there was a clear influence for the home–away distinction, with home birds of either sex having an advantage, $F(1, 150) = 8.39$, $p < .01$. There were no interactions between any of the between-subjects factors of cage size, sex, and home–away status.

There were marked changes in the rate of body pecking over days, $F(9, 1350) = 21.46$, $p < .01$, and a Cage Size × Days interaction indicated that this response type peaked sooner in the large cages than in the small ones, $F(9, 1350) = 2.11$, $p < .05$. Finally, a Home–Away × Days interaction supports the impression that there was a decline in the home-cage advantage over the days of testing, $F(9, 1350) = 2.47$, $p < .05$.

Foot pecking. Foot pecks were rare, relative to pecks at other places. The grand per-minute means for head, body, and cage pecks were 2.61, 1.85, and 8.88, respectively, compared to a grand mean of .35 for foot pecks.

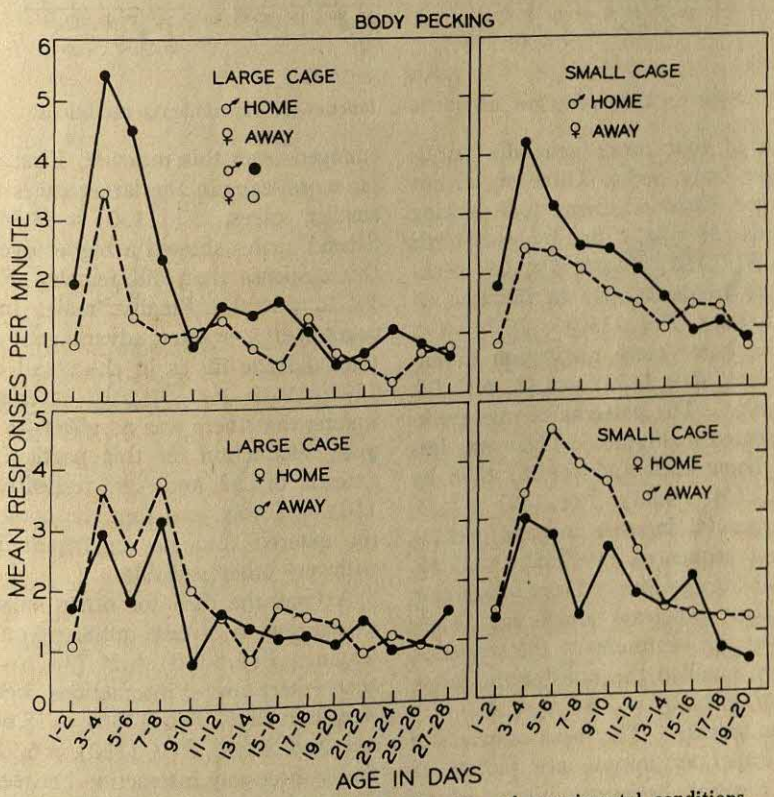


Figure 2. Body pecks per minute as a function of experimental conditions.

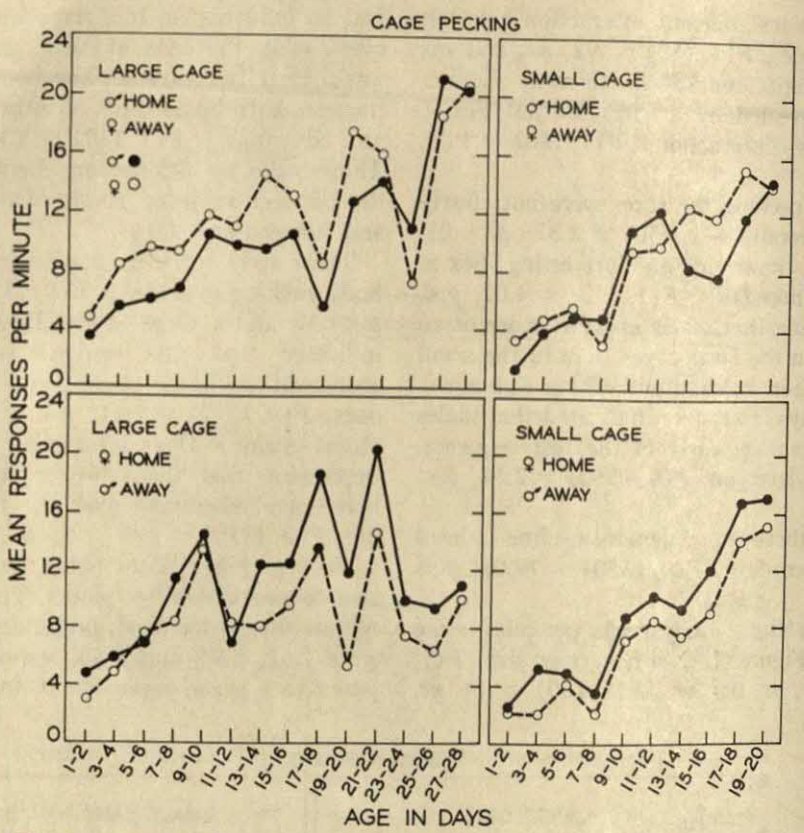


Figure 3. Cage pecks per minute as a function of experimental conditions.

The pattern of foot pecks generally paralleled that for body pecks. Three significant effects emerged. There was more foot pecking by home birds ($M = .41$) than by away birds ($M = .29$), $F(1, 150) = 5.70$, $p < .05$; pecking was more frequent early in the test sequence than later, $F(9, 1350) = 20.74$, $p < .01$; and the home–away distinction diminished over days, $F(9, 1350) = 3.18$, $p < .05$.

Cage pecking. The patterns of cage pecking are presented in Figure 3. There was less pecking by home birds ($M = 7.88$) than by away birds, ($M = 9.88$), $F(1, 150) = 4.33$, $p < .05$; a general increase in cage pecking over the test sequence, $F(9, 1350) = 17.35$, $p < .01$; and a Cage Size × Days interaction indicated that whereas small-cage birds pecked less at the beginning of the sequence, later on they equalled the rates of the large-cage birds, $F(9, 1350) = 2.84$, $p < .01$.

Aggressive leaping. The mean numbers of aggressive leaps per minute are plotted in Figure 4. Several between-subjects effects emerged from this measure. First, there were far more leaps in the larger cages than in the smaller cages, $F(1, 150) = 83.34$, $p < .01$. Second, males showed a higher overall rate of this response than did females, $F(1, 150) = 15.70$, $p < .01$. Finally, males in the large cages had a greater advantage over females than did the males in the small cages, $F(1, 150) = 9.32$, $p < .01$. Also interesting is the finding that there was no effect for the home–away distinction on this particular measure (means of .32 and .29, respectively), $F(1, 150) = .50$, *ns*, and that the home–away factor entered into no significant interactions with any other variable.

As was the case for other measures, days of age had a strong influence on aggressive leaping, $F(9, 1350) = 29.73$, $p < .01$. Further, there were interactions between cage size and days, $F(9, 1350) = 18.60$, $p < .01$, sex and days, $F(9, 1350) = 6.91$, $p < .01$, and a three-way interaction between cage size, sex, and days, $F(9, 1350) = 3.83$, $p < .01$.

These interactions mean that the differences between the sexes were more evident later in the test sequence than in early tests and that large-cage males had a longer lasting advantage over females than did small-cage males.

Other response types. The remaining responses were rarely observed. Scores on the immobility measure could range up to 150 per test, but near-zero rates emerged, with a grand mean for immobility of .12. The cage push reaction was also quite infrequent, with a grand mean of 1.35 out of a possible 150 per test. Therefore, whereas some significant effects emerged from these two measures, they do not bring much clarification to the issues at hand, and we hesitate to place strong interpretation on these contrasts because of their basis in such small numbers.

Distress vocalizations were also relatively infrequent, with a grand mean of .18. That is, during any given minute of testing, an observer was 76 times more likely to see some kind of peck than to hear a distress call. Although we have reservations about strong interpretations based on this overall low rate, it is perhaps worth noting that there was a general decline in distress calls per minute from the first day ($M = .50$) to the last day of testing ($M = .02$), $F(9, 1350) = 2.64$, $p < .01$.

Correlational Analyses

Correlations between pairmates were computed for selected measures on each day of testing. Large and small cage data were treated separately. The results of these analyses are shown in Figure 5, where conven-

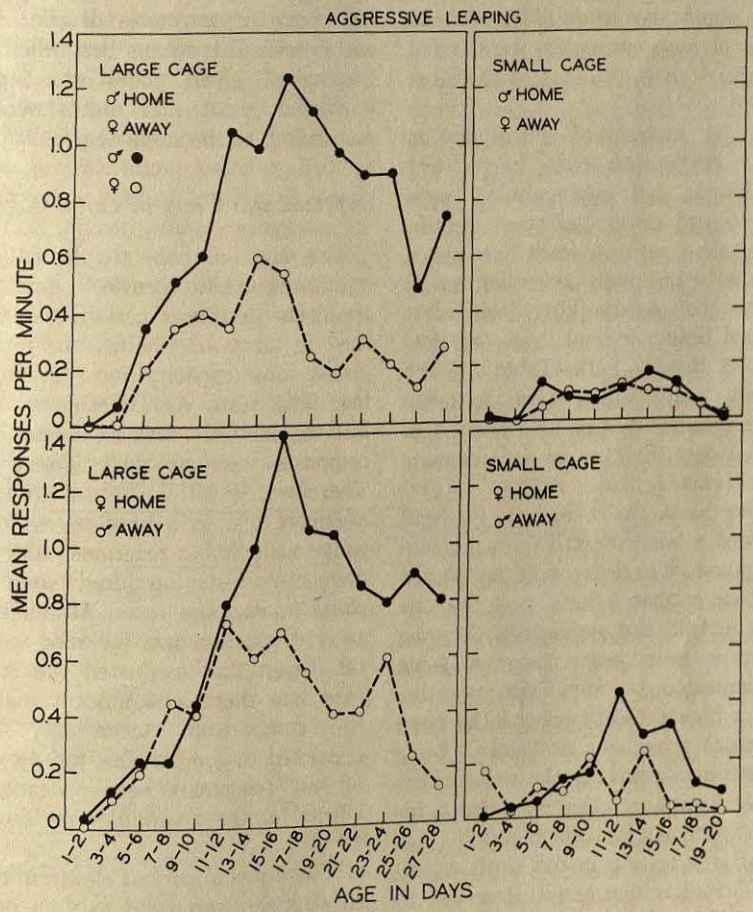


Figure 4. Aggressive leaps per minute as a function of experimental conditions.

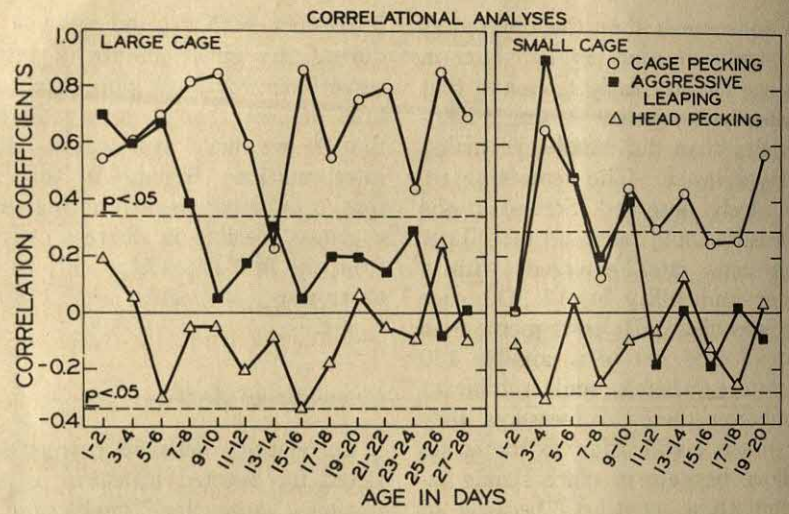


Figure 5. Correlation coefficients showing the relationship between pairs of birds on selected response types.

tional (.05) significance levels are also depicted. These significance levels differ over the two cage sizes because more pairs were tested in the small cages ($n = 46$) than in the large cages ($n = 33$).

There is a fair amount of consistency in the patterns of coefficients across the two test units. Correlations for cage pecking were mostly positive and stable, and were significantly greater than zero on many test days. This pattern is in line with an earlier report on the matching of cage pecking during this part of chicks' developmental span (cf. Rajecki, Lamb, & Suomi, 1978, Table 2). We interpret these positive correlations to mean that the cage pecking by one bird elicited or facilitated the cage pecking of its testmate (cf. Rajecki, Ivins, & Kidd, 1977).

On the other hand, the coefficients for head pecking showed a weak negative relationship for this response. The interpretation placed on this pattern is that a head peck by one bird at another did not necessarily elicit or facilitate return head pecks, and on some occasions (indicated by significant negative coefficients) a high rate of pecks to the head precluded return pecks to that locus.

In further contrast to both the patterns for cage and head pecking, the coefficients for aggressive leaping suggest that initially (or relatively early in testing in the small cages) such leaps elicited return leaps. However, as reference to Figure 4 shows, as leaps per test increase, the positive relationship between the leap rates of testmates deteriorated to zero and below. This means that when the rate of leaping of given males was highest, their particular female test mates were no longer responding to the same degree.

Supplemental Tests in Large Cages

The purpose of the supplemental tests with strangers was to provide some information about the decline in certain response rates, at least in terms of whether the birds still possessed some capacity for aggressive responding. The tests with strangers followed the last regular tests with cagemates, when many responses were at their lowest recent ebb. Therefore, it will be useful to use the final cagemate test as a baseline against which to gauge xenophobic reactions. Indeed, the supplemental tests produced some interesting shifts in response rates. Although the home-away distinction was included in the statistical design that evaluated the results, in no case was there a significant main effect for this comparison. Accordingly, attention is restricted to comparisons for sex and for type of test (cagemate versus stranger). Certain effects for these variables are shown in Table 1.

There was a marked elevation of head pecks at strangers compared to such pecks at cagemates during the last regular tests, $F(1, 24)$

$= 9.00$, $p < .01$. The sexes did not differ on this measure, $F(1, 24) = 2.18$, *ns*, nor was there an interaction between sex and type of test, $F(1, 24) = 1.54$, *ns*.

Effects for body or foot pecking did not emerge, but there was a strong effect for tests on the cage peck measure. Table 1 shows that cage pecking was much less frequent among strangers than among cagemates, $F(1, 24) = 12.15$, $p < .01$, with males and females affected across tests about equally, $F(1, 24) = .77$, *ns*.

On the other hand, sex had a very strong influence on aggressive leaping, with males engaging in this behavior more than females overall, $F(1, 24) = 28.19$, $p < .01$. As was the case with head pecking, there was more leaping during tests of strangers than during tests of cagemates, $F(1, 24) = 11.19$, $p < .01$, but interestingly enough, it can be seen in Table 1 that the cagemate–stranger difference is due entirely to the behavior of the males, $F(1, 24) = 10.56$, $p < .01$. In fact, comparisons between Table 1 and Figure 4 show that males' average rate of leaping at unfamiliar males was about twice as high as the highest average rate of leaping at female cagemates. This indicates that leaping was indeed an aggressive response and not strictly a sexual response directed at the opposite sex.

The distress call and immobility measures revealed no significant effects, but there was an interesting interaction between sex and test on the cage push measure. When tested with their familiar female cagemates, males showed 0 scoring units (see Method) on this index, but their scores rose to an average of 1.14 when tested with unfamiliar males. To the contrary, females showed an average score of 4.21 when tested with familiar males and only .43 such units when confronted with unfamiliar females, $F(1, 24) = 5.03$, $p < .05$.

Table 1
Average Responses per Minute in the Last Regular Tests With Cagemates (Mixed-Sex Pairs) and the Supplemental Tests With Strangers (Same-Sex Pairs)

Sex	Test	
	Cagemate	Stranger
	Head pecking	
Male	.31	1.07
Female	.10	.41
	Cage pecking	
Male	17.86	.84
Female	17.46	7.29
	Aggressive leaping	
Male	.56	2.54
Female	.14	.17

Discussion

Note should be made of the fact that the behavioral profiles just reported emerged under conditions of multiple separations and reunions of female and male pairmates. There is little doubt that birds reared under other circumstances, sexual combinations, or numbers might exhibit very different patterns. Previous evidence indicated that, compared to animals never separated, the multiple separation procedure enhances social behavior over this age range (see Rajecki, Lamb, & Suomi, 1978), and this was precisely why the technique was used here. The separation technique proved quite useful for demonstrating the capacity and forms of social behavior in developing fowl, which was the aim of the study.

In some cases, patterns of behavior across the large and small cages were vastly different. In order to bring organization to these and other complicated results, we will begin with the findings from the large cages, and then take up the small-cage data.

Behavior in Large Cages

The behavior of the birds in the large cages clearly indicates that in studying the development of aggressiveness in such species, different forms of physical contact between individuals cannot be viewed as equivalent. That is, response types such as head pecking, body pecking, foot pecking, and leaping are not interchangeable indexes of "aggression" because they seem governed by separable influences or factors. By this we do not mean merely that certain response types are more or less likely at given points in maturation. Rather, we mean that when they emerge or

reach peak rates, given responses seem differentially mediated by internal or external factors.

At one extreme, aggressive leaping in the large cages appears to have been independent of the environmental factor represented by the home–away distinction. Regardless of whether males were home or away, they began leaping (on average) at the same point in the developmental span and reached the same peak rate on the same test day. Also interesting in this respect are the leaping responses of the females. On average, both home and away females initially matched the male rates but then peaked at the same level on the same day.

At the other extreme, sex had little effect on body pecking compared with its influence on leaping. When rates of body pecking peaked, it was the home bird that showed the highest rate, regardless of sex (and a similar pattern emerged for foot pecking).

Further, it seems possible to draw distinctions between leaping and head pecking. Males had a general advantage over females in head pecking at the point of peak response rate, but this sex difference did not conform to the profile of the sex difference for leaping. In leaping, males established an advantage and then kept it (so to speak), whereas in head pecking the males had an advantage, then lost it. Finally, head pecking is different from body pecking, since at the point of maximum head peck rates in the large cages, the biological factor of sex outweighed the environmental factor that was based on the home–away distinction.

The components of social behavior are also distinguished from one another by the correlations within response types. The correlational analyses presented in Figure 5 demonstrate pronounced differences across response categories. These analyses indicate that cage pecks were matched between birds, but that head pecks were not. Interestingly, leaps were matched early in the test sequence but not later. The interpretation we place on these different patterns is that different motivational factors mediate such responses. A peck to the head may or may not reflect an aggressive disposition on the part of the pecking bird, but it does not seem to serve as an occasion for retaliatory pecks on the part of the target (cf. Rajecki, Ivins, & Rein, 1976). On the other hand, an aggressive leap was likely to elicit a leap in return,[4] up to a point in testing. We suspect that that point was reached when the females became submissive. This suspicion seems borne out by the reaction of males to chickens that were not submissive; that is, their opponents in the male–male supplemental tests. These supplemental tests produced the highest rates of leaping seen in the study and revealed a very close match in leap rates between such willing birds, $r(12) = .98$, $p < .01$.

This allusion to the supplemental tests brings us to another distinction between leaping and head pecking. The tests of female strangers revealed very little leaping, and we suggest (as noted earlier) that this was so because they had learned, on the basis of their experience with males, to inhibit leaps in the presence of a testmate. However, if females were subdued with respect to leaping, they were much less so with respect to head pecking. Like their male counterparts, females showed a fourfold increase in head pecking in tests of strangers, as compared with tests of cagemates.

Behavior in Small Cages

A number of reports (cf. Craig & Bhagwat, 1974; Craig, Biswas, & Guhl, 1969; Hughes & Wood-Gush, 1977) indicate that fowl kept in close confinement are less likely to show aggressive behavior than are birds reared in more spacious units. Hughes and Wood-Gush (1977) account for this difference by hypothesizing that aggression is triggered by the event of birds coming into proximity rather than by the condition of being in continuous proximity. Indeed, the data from the small cages are in strong support of the hy-

[4] It is worth reporting that head pecks are not correlated with leaps. Correlation coefficients were computed to determine the relationship between male head pecks and female leaps and between female head pecks and male leaps. This analysis yielded 48 coefficients from the regular daily tests in the large and small cages. These coefficients ranged from −.27 to .27, with none significantly different from zero.

pothesis that close confinement dampens or otherwise influences aggressive responding. The first indication of this is that males in small cages had no advantage over females on head pecking; in these units the home bird had the edge regardless of sex. There was also a home-cage advantage in body pecking, as was seen in the larger units.

However, the most dramatic difference across the two cage sizes was on the aggressive leaping index, with leaps in the small cages being far less frequent (see Figure 4). There are two ways to account for this difference. First, the relative absence of leaping in the small units may be in line with the Hughes and Wood-Gush (1977) notion that birds continuously in proximity are not in the proper condition to be stimulated to aggress. On the other hand, at present we cannot rule out the argument that the very dimensions of the small cages may have reduced leaping, whatever the disposition of the inhabitants.

The distinction between leap and head peck rates in the small cages is, however, noteworthy with respect to the motivational bases of such responses. While leaping was somehow limited by confinement, head pecking under these conditions was initially higher than in the large cages. Of course, the peak rates for these responses occurred at different points in development, but the differential effect of the small units on pecking and leaping further points to distinctions between components of agonistic behavior in fowl.

References

Adams, D. B. Defence and territorial behaviour dissociated by hypothalamic lesions in the rat. *Nature*, 1971, *232*, 573–574.

Baenninger, L. P. Social dominance orders in the rat: "Spontaneous," food, and water competition. *Journal of Comparative and Physiological Psychology*, 1970, *71*, 202–209.

Baenninger, L. P., & Baenninger, R. "Spontaneous" fighting and mouse-killing by rats. *Psychonomic Science*, 1970, *19*, 161.

Baron, R. A. *Human aggression*. New York: Plenum, 1977.

Blanchard, R. J., Blanchard, D. C., & Takahashi, L. K. Pain and aggression in the rat. *Behavioral Biology*, 1978, *23*, 291–305.

Bugbee, N. M., & Eichelman, B. S., Jr. Sensory alterations and aggressive behavior in the rat. *Physiology and Behavior*, 1972, *8*, 981–985.

Clayton, F. L., & Hinde, R. A. The habituation and recovery of aggressive display in *Betta splendens*. *Behaviour*, 1968, *30*, 96–106.

Craig, J. V., & Bhagwat, A. L. Agonistic and mating behavior of adult chickens modified by social and physical environments. *Applied Animal Ethology*, 1974, *1*, 57–65.

Craig, J. V., Biswas, D. K., & Guhl, A. M. Agonistic behavior influenced by strangeness, crowding and heredity in female domestic fowl. *Animal Behaviour*, 1969, *17*, 498–506.

Davis, R. E. Readiness to display in the paradise fish *Macropodus opercularis*, L., *Belontiidae*: The problems of general and specific effects of social isolation. *Behavioral Biology*, 1975, *15*, 419–433.

Dawson, J. S., & Siegel, P. B. Behavior patterns of chickens to ten weeks of age. *Poultry Science*, 1967, *46*, 615–622.

Deag, J. M. Aggression and submission in monkey societies. *Animal Behaviour*, 1977, *25*, 465–474.

Eiserer, L. A., Emerling, M. R., Scardina, S. J., & Hoffman, H. S. Stimulus factors in the isolation-induced aggression of bobwhite quail and Khaki Campbell ducklings. *Aggressive Behavior*, 1976, *2*, 285–293.

Fredericson, E. Aggressiveness in female mice. *Journal of Comparative and Physiological Psychology*, 1952, *45*, 254–257.

Ghiselli, W. B., & Thor, D. H. Visual, tactual, and olfactory deprivation effects on irritable fighting behavior of male hooded rats. *Physiological Psychology*, 1975, *3*, 47–50.

Guhl, A. M. The development of social organization in the domestic chick. *Animal Behaviour*, 1958, *6*, 92–111.

Hoffman, H. S., Boskoff, K. J., Eiserer, L. A., & Klein, S. H. Isolation-induced aggression in newly hatched ducklings. *Journal of Comparative and Physiological Psychology*, 1975, *89*, 447–456.

Hughes, B. O., & Wood-Gush, D. G. M. Agonistic behaviour in domestic hens: The influence of housing method and group size. *Animal Behaviour*, 1977, *25*, 1056–1062.

Kruijt, J. P. Ontogeny of social behaviour in Burmese red jungle fowl (*Gallus gallus spadiceus*) Bonnaterre. *Behaviour*, 1964, Suppl. *12*, 1–201.

Rajecki, D. W., Grams, J. R., Stursa, S., & Nerenz, D. R. Mediating factors in chick social pecking: II. Effects of exposure to movement. *Aggressive Behavior*, 1978, *4*, 331–343.

Rajecki, D. W., Ivins, B., & Kidd, R. F. Affiliation, xenophobia, and the behavioral organization of the young domestic chicken. *Behavioral Biology*, 1977, *21*, 364–379.

Rajecki, D. W., Ivins, B., & Rein, B. Social discrimination and aggressive pecking in domestic chicks. *Journal of Comparative and Physiological Psychology*, 1976, *90*, 442–452.

Rajecki, D. W., Kidd, R. F., & Irvins, B. Social facilitation in chickens: A different level of analysis. *Journal of Experimental Social Psychology*, 1976, *12*, 233–246.

Rajecki, D. W., Lamb, M. E., & Suomi, S. J. Effects

of multiple peer separation in domestic chicks. *Developmental Psychology,* 1978, *14,* 379–387.

Rajecki, D. W., Nerenz, D. R., Barnes, R. D., Ivins, B., & Rein, B. Isolation-induced social pecking in domestic chicks: Contemporary correlates and sequelae. *Aggressive Behavior,* 1977, *3,* 77–90.

Rajecki, D. W., Nerenz, D. R., & Rein, B. Mediating factors in chick social pecking. I. Effects of isolation, imprinting, and peer contact. *Aggressive Behavior,* 1978, *4,* 141–151.

Rajecki, D. W., Suomi, S. J., Scott, E. A., & Campbell, B. Effects of social isolation and social separation in domestic chicks. *Developmental Psychology,* 1977, *13,* 143–155.

Ratner, S. C. Comparisons between behaviour development of normal and isolated domestic fowl. *Animal Behaviour,* 1965, *13,* 497–503.

Rowe, F. A., & Edwards, D. A. Olfactory bulb removal: Influences on the aggressive behaviors of male mice. *Physiology and Behavior,* 1971, *7,* 889–892.

Rowell, T. The concept of social dominance. *Behavioral Biology,* 1974, *11,* 131–154.

Shapiro, S., & Schuckman, H. Habituation and covariation of the components of the threat display in *Betta splendens. Psychological Reports,* 1971, *28,* 827–837.

Syme, G. J. Competitive orders as measures of social dominance. *Animal Behaviour,* 1974, *22,* 931–940.

Tedeschi, J. T., Smith, R. B., III, & Brown, R. C., Jr. A reinterpretation of research on aggression. *Psychological Bulletin,* 1974, *81,* 540–562.

Thor, D. H. Enhanced social docility in male hooded rats by dermal cautery of the vibrissal pads. *Aggressive Behavior,* 1976, *2,* 39–53.

Wood-Gush, D. G. M. *The behaviour of the domestic fowl.* London: Heinemann, 1971.

Zook, J. M., & Adams, D. B. Competitive fighting in the rat. *Journal of Comparative and Physiological Psychology,* 1975, *88,* 418–423.

Received December 26, 1978 ▪

Journal of Personality and Social Psychology
1979, Vol. 37, No. 10, 1915-1926

Issue Involvement Can Increase or Decrease Persuasion by Enhancing Message-Relevant Cognitive Responses

Richard E. Petty
University of Missouri—Columbia

John T. Cacioppo
University of Iowa

Two experiments were conducted to test the hypothesis that high issue involvement enhances thinking about the content of a persuasive communication. Experiment 1 varied involvement and the direction of a message (proattitudinal or counterattitudinal). Increasing involvement enhanced persuasion for the proattitudinal but reduced persuasion for the counterattitudinal advocacy. Experiment 2 again varied involvement, but both messages took a counterattitudinal position. One message employed compelling arguments and elicited primarily favorable thoughts, whereas the other employed weak arguments and elicited primarily counterarguments. Increasing involvement enhanced persuasion for the strong message but reduced persuasion for the weak one. Together the experiments provide support for the view that high involvement with an issue enhances message processing and therefore can result in either increased or decreased acceptance.

Persuasion researchers have recognized for some time that it is easier to demonstrate attitude change in the laboratory than in the field. One prominently mentioned explanation for this observation is that the advocacies employed in laboratory investigations are of considerably lower "involvement" than the advocacies encountered in the real world (Hovland, 1959). Greater involvement with an issue is presumably related to greater resistance to persuasion (cf. Triandis, 1971). The major goal of the present paper is to present and test a "cognitive response" (Greenwald, 1968; Petty, Ostrom, & Brock, in press) interpretation of involvement effects holding that increasing involvement with an issue increases one's motivation to process information relevant to the issue and can lead to either increased or decreased persuasion.

We are grateful to Chuck LaJuenesse, Colleen Holt, Greg Bistline, and Michael Bryant for help in collecting and coding the data of Experiment 2, and to the University of Missouri Research Council for financial support.

Requests for reprints should be sent to Richard E. Petty, Department of Psychology, 210 McAlester Hall, University of Missouri, Columbia, Missouri 65211.

Before presenting the model in more detail, it is necessary to note that attitude researchers have distinguished between two different types of involvement that can affect susceptibility to influence. One kind of involvement concerns the extent to which the attitudinal *issue* under consideration is of personal importance, whereas a second concerns the extent to which the particular attitudinal *response* adopted is of personal importance to the individual. The first type of importance, which is the type under investigation here, has been called "issue involvement" (Kiesler, Collins, & Miller, 1969), "ego-involvement" (Rhine & Severance, 1970; Sherif, Sherif, & Nebergall, 1965), and "personal involvement" (Apsler & Sears, 1968; Sherif, Kelly, Rodgers, Sarup, & Tittler, 1973). Also, Halverson & Pallak (1978) and Madsen (1978) have argued that typical manipulations of "commitment" (cf. Kiesler, 1971) elevate involvement with an issue.

The second type of involvement is often referred to as "response involvement" (Zimbardo, 1960) or "task involvement" (Sherif & Hovland, 1961). In this second kind of involvement, the attitudinal issue is not particularly important to the person, but adopt-

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3710-1915$00.75

ing a position that will maximize the immediate situational rewards is. Thus in some cases response involvement will lead to increased influence (e.g., Zimbardo, 1960) and in some cases decreased influence (e.g., Freedman, 1964), depending upon which position the rewards favor.

The focus of this paper is on the mechanism mediating the effects of high issue involvement. High issue involvement occurs when an issue has "intrinsic importance" (Sherif & Hovland, 1961, p. 197), or "personal meaning" (Sherif et al., 1973, p. 311), when people expect the issue "to have significant consequences for their own lives" (Apsler & Sears, 1968, p. 162), and when concerns about immediate situational rewards are "dwarfed by outcomes connected with the topic itself" (Cialdini, Levy, Herman, Kozlowski, & Petty, 1976, p. 664). Most of the early research indicated that increased issue involvement was associated with increased resistance to persuasion (e.g., Miller, 1965; Sherif & Hovland, 1961). The most prominently mentioned explanation for this finding was derived from social judgment theory (Sherif et al., 1965). The notion was that on any given issue, highly involved persons should exhibit more negative evaluations of a communication because high involvement is associated with an extended "latitude of rejection" (the attitudinal positions that a person finds unacceptable). Thus, incoming messages on high involvement issues would have an enhanced probability of being rejected because they were more likely to fall within the unacceptable range of a person's implicit attitude continuum (cf. Eagly & Manis, 1966).

Contrary to social judgment theory, which seems always to predict greater resistance with increased issue involvement, some investigators have found increased involvement to be associated with greater influence. For example, Eagly (1967) presented subjects with information about either themselves (high involvement) or another person (low involvement) that was discrepant in either a favorable or an unfavorable direction from their initial attitudes. She found that although high involvement subjects changed less than low involvement subjects when unfavorable information was provided, the former exhibited more change when favorable information was provided. Similarly, Pallak, Mueller, Dollar, & Pallak (1972) presented subjects who were either publicly committed (high involvement) or privately committed (low involvement) to their initial attitudes with information that either contradicted their opinions or was consistent but more extreme. They found that involvement increased resistance to the countercommunication but facilitated change toward the position advocated by the extreme consonant message. In order to account for these results, Pallak et al. (1972) suggested a modification of the original social judgment model. The new formulation was that high involvement (or commitment) increased the probability of rejecting or contrasting counterattitudinal information but increased the probability of accepting or assimilating proattitudinal information.

These same data may be explained by an alternative view holding that involvement increases the amount of thought in which subjects engage about the stimulus information. When the stimulus information is inconsistent with subjects' original attitudes, it is likely that subjects are motivated and able to generate counterarguments to the material presented. To the extent that increased involvement is associated with more thinking, increased counterargumentation and resistance to influence would be a likely result. On the other hand, when the stimulus information is consistent with subjects' original attitudes, it is likely that they are initially biased toward generating favorable cognitions. To the extent that increased involvement is associated with more thinking here, more favorable thoughts might be generated, and increased influence would result. There is already a large body of literature supporting the view that the idiosyncratic cognitive responses elicited by a communication are an important determinant of the direction and amount of attitude change produced (e.g., Brock, 1967; Cacioppo & Petty, 1979; Cook, 1969; Insko, Turnbull, & Yandell, 1974; Petty, Wells, & Brock, 1976; Tesser, 1978; etc.).

It is also important to note that there are already some suggestions in the literature that

increased involvement with a stimulus is associated with more extensive information processing. For example, Rogers, Kuiper, and Kirker (1977) had subjects rate personality adjectives in terms of their self-relevance (high involvement) or semantic, phonemic, and structural properties (low involvement). Adjectives processed under high involvement conditions were subsequently recalled best in an incidental recall test. According to Craik and Lockhart's (1972) "depth of processing" framework, enhanced recall is thought to reflect more extensive processing of the stimulus. More relevant to persuasion situations, Cialdini et al. (1976) found that subjects who expected to engage in a discussion with an opponent generated more supportive thoughts in anticipation of the discussion when the attitude issue was of high rather than low personal relevance. Finally Chaiken (Note 1) and Petty & Cacioppo (1979) have reported that subjects' message-relevant thoughts show higher correlations with message acceptance under high than under low issue involvement conditions.

Two experiments were conducted to test the cognitive response view of the effects of issue involvement. The primary goal of Experiment 1 was to replicate conceptually the Eagly (1967) and Pallak et al. (1972) attitudinal findings; measures of subject-generated cognitive responses were included also to assess the viability of the proposed reformulation of issue-involvement effects. The second experiment was designed to evaluate the two competing explanations of issue-involvement effects, namely, thought enhancement versus the revised social judgment theory formulation. Thus, an experiment was devised for which the two formulations made competing predictions.

Experiment 1

Most of the early work on issue involvement was conducted by finding existing groups that differed in the extent to which an issue was important, and thus this work was correlational in nature (see Kiesler et al., 1969, for the interpretational problems with this approach). More recent investigators have chosen to *manipulate* involvement by varying the issue between subjects (e.g., Dean, Austin, & Watts, 1971; Rhine & Severance, 1970). In other words, some subjects would receive a highly involving issue (e.g., increasing tuition), whereas others received an issue of low involvement (e.g., increasing park acreage in a distant city). A preferable procedure that keeps the communication constant across subjects was introduced by Apsler and Sears (1968) and is the method employed here. In this procedure, subjects in both high and low involvement groups receive the same communication, but high involvement subjects are led to believe that the advocated change will affect them, whereas low involvement subjects do not believe the change will have personally relevant effects. In the present study, this was accomplished by telling the college student subjects that a proposed change in university regulations regarding mixed-sex visitation hours was being made at either their university or a distant university. Pilot testing was conducted to develop proattitudinal and counterattitudinal communications. Not surprisingly, advocating more lenient regulations regarding visitation hours was found to be highly proattitudinal, and advocating stricter regulations was highly counterattitudinal. In addition, pilot testing revealed that subjects generated predominantly favorable thoughts to the proattitudinal communication and counterarguments to the counterattitudinal appeal. In Experiment 1 subjects received either a proattitudinal or a counterattitudinal advocacy under conditions of either high or low issue involvement. We predicted that increased involvement would enhance processing of the message contents. As a result of this, we expected both cognitive and affective responses to the two messages to be more extreme under high than under low involvement conditions. More specifically, increased involvement should decrease agreement with the counterattitudinal message but increase agreement with the proattitudinal communication.

Method

Procedure

Twenty-four male undergraduates at the University of Notre Dame participated in order to earn

extra credit in an introductory psychology course. The design was a 2 (High or Low Issue Involvement) × 2 (Proattitudinal or Counterattitudinal Message Direction) factorial. Subjects were tested individually. Upon arrival, subjects were informed that students in a sound engineering course had prepared the communications that were to be employed in the investigation and that in return for the use of the tapes in other research, the investigators had agreed to provide evaluations of the sound quality of the tapes. The subjects were asked to aid in this endeavor.

Subjects were informed of the topic and position of the communication they were about to hear and 60 sec later were exposed to one of two professionally taped messages. Following exposure to a 2-minute message, subjects completed the dependent variable booklets and were debriefed, thanked, and dismissed.

Independent Variables

Message direction. All subjects were exposed to a communication concerning coed visitation hours. The *proattitudinal* advocacy contended that colleges should be more lenient in allowing mixed-sex visitation. Elaborations of the following arguments were employed: (a) More lenient hours would not interfere with education, since the "Educational Examination Foundation" found no correlation between the length of visitation hours and graduate school entrance examination scores; (b) they also found no connection between visitation hours and grade point average; (c) enforcement of the morality inherent in visitation hours fails because morality is not tied to time of day; (d) since college is a period of responsible maturation, the imposition of visitation hours can be counterproductive; and (e) students are able to judge for themselves at what time a party should end. The *counterattitudinal* advocacy contended that colleges should be more strict in their visitation policies. Elaborations of the following arguments were employed: (a) Since the main purpose of going to college is to learn, university officials feel obligated to restrict weekday partying; (b) longer visitation hours have caused a trend away from Christian morals and ideals; (c) the "Educational Examination Foundation" found that the longer the visitation hours at a university, the lower the graduate school entrance examination scores obtained by its students; (d) they also found a negative correlation between visitation hours and grade point average; and (e) young students are not mature enough to handle the social pressures that accompany an increase in visitation hours.

Issue involvement. For subjects in the *high involvement* conditions, the speaker advocated that the change in visitation rules go into effect at their own university (Notre Dame). Thus all subjects would be affected personally by the proposal. In the *low involvement* conditions, the speaker advocated that the change in visitation rules go into effect at another university (Juanita Junior College). Thus, none of the subjects would be affected personally by the proposal. The messages in both involvement conditions were identical except for the words *Notre Dame* and *Juanita Junior College*.

Dependent Variables

Attitude measures. On the first page of the dependent variable booklet, subjects read: "Because your own opinion about the position advocated on the tape may influence the way you rate the quality of the tape, we would like to obtain a measure of how you feel about the views proposed by the speaker on each scale below." Two measures of opinion about visitation hours were included. First subjects rated the advocated position on four 9-point semantic differential scales (harmful–beneficial, wise–foolish, good–bad, favorable–unfavorable) that were summed to obtain a general measure of evaluation. Next, subjects responded to an 11-point Likert-type rating scale regarding their agreement with the speaker's position. On the scale, 1 indicated that the subject "did not agree at all," and 11 indicated "agree completely." The subjects' responses to the two attitude measures were converted to standard scores and were averaged for an index of communication acceptance.

Cognitive response measures. After completion of the attitude scales, subjects were given 2½ minutes to list the thoughts they had while listening to the tape (cf. Petty & Cacioppo, 1977). Twelve 8-inch (20.32 cm) horizontal lines each about 1 inch (2.54 cm) from the one above created the boxes in which subjects were to write their ideas, one per box. After recording their thoughts, subjects were instructed to rate their ideas as either + (in favor of the advocated position), − (opposed to the advocated position), or 0 (neutral or irrelevant). No independent judges were used to score subjects' thoughts, since previous research has indicated that subjects' and judges' ratings correlate highly (cf. Petty et al., 1976). After rating their thoughts, subjects completed a page of ancillary items regarding the quality of the taped communication and how involving they felt it was.

Results and Discussion

Manipulation Checks

Subjects rated the extent to which they found the communication "involving" on an 11-point scale where 11 indicated "extremely involving." Subjects in the high involvement conditions rated the message as significantly more involving ($M = 6.67$) than subjects in the low involvement conditions ($M = 4.75$), $F(1, 20) = 5.09$, $p < .03$, providing support for the effectiveness of the involvement manipulation.

Table 1
Attitudes and Cognitive Responses in Relation to Involvement and Type of Message (Pro or Counter)

Item	Counterattitudinal message		Proattitudinal message	
	Involvement		Involvement	
	Low	High	Low	High
Attitude	$-.39_a$	-1.79_b	$.62_c$	1.57_d
Counterarguments	3.00_{ac}	4.00_a	2.00_{bc}	$.67_b$
Favorable thoughts	1.33_{ab}	$.50_b$	1.17_{ab}	3.17_a

Note. Means in any given row without a common subscript are significantly different at the .05 level by the Newman–Keuls procedure.

The mean score on the standardized attitude index for subjects hearing the counterattitudinal advocacy was −1.10, which differed significantly from the mean score of 1.10 for subjects hearing the proattitudinal message, $F(1, 20) = 35.33$, $p < .001$. An average of 3.5 counterarguments was generated to the counterattitudinal message, whereas an average of only 1.3 counterarguments was generated to the proattitudinal message, $F(1, 20) = 15.09$, $p < .001$. An average of 2.2 favorable thoughts was generated to the proattitudinal message, whereas only .9 favorable thoughts were generated to the counterattitudinal message, $F(1, 20) = 3.5$, $p < .07$. The attitude and cognitive response data provide support for the notion that the two communications differed in the agreeableness of their positions.

Tests of Hypotheses

The attitude and cognitive response data are presented in Table 1. It was expected that involvement would decrease the effectiveness of the counterattitudinal message but increase the effectiveness of the proattitudinal message. A significant Involvement × Message interaction on the attitude index provided support for this hypothesis, $F(1, 20) = 10.14$, $p < .005$. A Newman–Keuls analysis revealed that increased issue involvement increased agreement with the proattitudinal message but decreased agreement with the counterattitudinal advocacy (see Table 1).

Similar Involvement × Message interactions on the cognitive response measures provided support for the information-processing interpretation of involvement effects: counterarguments $F(1, 20) = 4.36$, $p < .05$; favorable thoughts $F(1, 20) = 4.42$, $p < .05$. Newman–Keuls analyses of these interactions (Table 1) revealed that under high involvement subjects generated more favorable thoughts and fewer counterarguments to the proattitudinal advocacy than to the counterattitudinal one; under low involvement, however, neither the number of favorable thoughts nor the number of counterarguments was affected by message direction. Surprisingly, there was a tendency for subjects hearing the proattitudinal message under low involvement to generate more counterarguments than did subjects hearing the same message under high involvement. Postexperimental interviews with subjects suggested that this effect may have resulted from subjects' jealousy over having such a desirable effect (leniency in visitation hours) occur at an institution other than their own.

The analyses of the several ancillary measures on tape quality revealed one significant effect. Subjects rated the speaker's voice quality as higher in the high involvement conditions than in the low, $F(1, 20) = 6.68$, $p < .05$. This effect was obtained regardless of whether the speaker advocated a proattitudinal or a counterattitudinal position and is puzzling, since both high and low involvement subjects heard identical tapes (except for the spliced insertion of the name of the appropriate university). A possible explanation for this finding may be that subjects wanted to

give the tapes higher quality ratings when they were prepared by students at their own institution than when they were prepared at another school.

Experiment 2

Although the cognitive response data from Experiment 1 provide support for the notion that increasing involvement increases the thinking that subjects engage in about an attitudinal issue, the attitude data could just as easily be explained by the revised social judgment formulation. Thus, subjects may have assimilated the proattitudinal information, producing acceptance, and contrasted the counterattitudinal information, producing resistance. According to the Pallak et al. (1972) formulation, the key determinant of whether involvement will facilitate or hinder persuasion is the extent to which the information provided appears to contradict the subject's initial position. When the information advocates a position opposite to that of the subject, involvement will decrease persuasion, but involvement "facilitates change toward a more extreme attitude in response to appeals which do not explicitly reject one's own position" (Pallak et al., 1972, p. 434).

According to the cognitive response view espoused here, the position advocated in the communication is not as important as the nature of the thoughts elicited by the message. Two counterattitudinal communications were constructed specifically for Experiment 2. Both messages argued that seniors be required to pass a comprehensive exam in their declared major before being granted a degree. Previous work has indicated that this position is strongly counterattitudinal for most college students (Petty & Cacioppo, 1977). The messages differed, however, in their presentation of eight key arguments. One message was designed to contain points that were logically sound, defensible, and compelling. The arguments in this message were selected from a pool that elicited predominantly favorable thoughts in a pretest. A second message was designed to be more open to refutation and skepticism. The arguments in this message were selected from a pool that elicited predominantly counterarguments in a pretest.

Given that the messages advocated an identical counterattitudinal position, but differed in the quality of the arguments used to support that position, it becomes possible to evaluate the two explanations of the effects of issue involvement. The modified social judgment formulation, as outlined by Pallak et al. (1972), predicts that increased issue involvement will produce decreased persuasion for both messages, since each message adopts an identical position in opposition to the subjects.[1]

The cognitive response hypothesis, on the other hand, predicts that increased issue involvement will be associated with decreased persuasion only for the message containing "counterarguable" arguments. For the message containing the compelling and difficult-to-counterargue arguments, increased involvement should be associated with increased persuasion. In other words, increased involvement motivates subjects to process the information contained in the communications more carefully. Thus, although high involvement may initially increase a subject's motivation to reject a counterattitudinal advocacy, subjects should ultimately better recognize the flaws in the weak communication and the virtues in the strong one.

In sum, the following predictions were made for Experiment 2. More counterarguments would be generated to the weak than to the strong communication, but more favorable thoughts would be generated to the strong message. There would be significant Arguments × Involvement interactions on the attitude and cognitive response measures. As in Experiment 1, it was predicted that both cognitive and affective responses to the two messages would be more polarized under high than under low involvement conditions.

---

[1] Alternatively, it might be predicted that involvement would increase resistance more for the message containing the compelling rather than the weak arguments, since this message might be viewed as taking a stronger stand against one's own position (Kleinhesselink & Edwards, 1975). In any case, the modified social judgment formulation expects increased involvement to be associated with decreased persuasion for both messages.

Method

Procedure

Seventy-two male and female undergraduates at the University of Missouri participated in order to earn extra credit in an introductory psychology course. The design was a 2 (High or Low Issue Involvement) × 2 (Strong or Weak Argument Quality) factorial. Subjects were run in groups of 4 to 12 in cubicles constructed so that no subject could have visual or verbal contact with any other subject. During any one session, all four experimental conditions were run. Upon arrival at the laboratory, subjects were told that each year the psychology department assists the school of journalism in evaluating radio editorials that are sent in by colleges and universities throughout the country; their task would be to provide ratings of the quality of the editorials. Following these instructions, subjects heard one of the taped communications over headphones. After listening to the appropriate 4-minute communication and completing the dependent variable booklets, subjects were debriefed, thanked, and dismissed.

Independent Variables

Argument quality. As noted previously, all subjects heard a communication advocating that seniors be required to pass a comprehensive exam in their declared major prior to graduation. In brief, the *strong version* of the message provided evidence (statistics, relevant studies, etc.) in support of the following arguments: (a) Prestigious universities have comprehensives to maintain academic excellence, (b) institution of the exams has led to a reversal in the declining scores on standardized achievement tests, (c) graduate and professional schools show a preference for undergraduates who have passed a comprehensive exam, (d) average starting salaries are higher for graduates of schools with the exams, (e) schools with the exams attract larger and more well-known corporations to recruit students for jobs, (f) the quality of undergraduate teaching has improved at schools with the exams, (g) state legislatures would increase financial support if exams were instituted, allowing a tuition increase to be avoided, and (h) the (fictitious) National Accrediting Board of Higher Education would give the university its highest rating if the exams were instituted.

The *weak version* of the message also contained 8 arguments but relied more on quotations and opinions than on statistics and data to support the following arguments: (a) Adopting the exams would allow the university to be at the forefront of a national trend, (b) graduate students have complained that since they have to take comprehensives, undergraduates should take them also, (c) by not administering the exams, a tradition dating back to the ancient Greeks was being violated, (d) parents had written to administrators in support of the plan, (e) the exams would increase student fear and anxiety enough to promote more studying, (f) the exams would help to cut costs by eliminating the necessity for other tests that varied with instructor, (g) the exams would allow students to compare their performance with students at other schools, and (h) job prospects might be improved.

Issue involvement. Involvement was manipulated in a manner analogous to that employed in Experiment 1. For subjects in the *high involvement* conditions, the speaker advocated that the comprehensive exams be instituted at the University of Missouri (the subjects' institution), whereas for *low involvement* subjects, the speaker advocated that the exams be instituted at North Carolina State University.

Dependent Variables

The same measures were available as in Experiment 1. Subjects rated the concept "comprehensive exams" on four 9-point semantic differential scales and then responded to an 11-point Likert-type scale regarding their agreement with the speaker's position. As in Experiment 1, responses to the two attitude measures were converted to standard scores and were averaged to form the measure of communication acceptance.

Next, subjects were given 2½ minutes to list the thoughts they had while listening to the tape. Subjects also rated their thoughts with a +, −, or 0, as described for Experiment 1. In addition, some ancillary questions about the quality of the tape were completed, and subjects were asked to rate the amount of thought they engaged in about the issue. Finally, subjects were given 3 minutes to attempt to list as many message arguments as they could remember. Each booklet was rated by two judges ($r = .88$) who were blind to the involvement manipulation. An argument had to correctly summarize one of the arguments that appeared in the appropriate message to be counted. Repetitions of the same argument were not counted. Disagreements between judges were resolved through discussion.

Results

Manipulation Checks

In order to determine whether our manipulation of involvement affected the amount of message processing subjects engaged in, subjects were asked to rate on an 11-point scale (1 indicated *not very much* and 11 indicated *very much*) how much thought they put into evaluating what the speaker had to say. Subjects in the high involvement conditions reported doing more thinking about the messages ($M = 8.6$) than subjects in the low involvement cells reported ($M = 7.75$), $F(1, 68) = 2.75$, $p < .05$, one-tailed. These data support the view that the involvement manip-

ulation affected the hypothesized mediating variable. In other words, increasing the personal relevance of an advocacy affected the perceived amount of thought subjects engaged in about that advocacy.

Analyses of the attitude and cognitive response measures indicated that the manipulation of message quality was also effective. The mean score on the standardized attitude index for subjects hearing the strong arguments was .45, which differed significantly from the mean score of −.45 for subjects hearing the weak arguments, $F(1, 68) = 21.65$, $p < .001$. An average of 2.19 favorable thoughts was generated to the strong arguments, whereas only .99 were generated to the weak arguments, $F(1, 68) = 15.19$, $p < .001$. An average of 2.69 counterarguments was generated to the weak message, whereas 1.69 were generated to the strong message, $F(1, 68) = 8.36$, $p < .005$. The attitude and cognitive response data provide support for the notion that the two communications differed in the strength of their arguments.

Tests of Hypotheses

The attitude and cognitive response data are presented in Table 2. A significant Arguments × Involvement interaction on the attitude index, $F(1, 68) = 5.62$, $p < .02$, and a Newman–Keuls analysis (Table 2) provided support for the hypothesis that involvement would increase the persuasiveness of the strong arguments but would decrease the persuasiveness of the weak arguments.

Involvement × Arguments interactions on the cognitive response measures also provided support for the information-processing interpretation of involvement effects: counterarguments $F(1, 68) = 7.46$, $p < .008$; favorable thoughts $F(1, 68) = 4.35$, $p < .04$. A Newman–Keuls analysis on each of these measures (Table 2) revealed that under high involvement, subjects generated more favorable thoughts and fewer counterarguments to the strong than to the weak arguments; under low involvement, neither favorable thoughts nor counterarguments were affected by the argument quality manipulation. The Newman–Keuls analyses further demonstrated that higher involvement increased the production of counterarguments to the weak arguments and increased the production of favorable thoughts to the strong arguments.

Analyses of the ancillary measures of tape quality produced no significant differences. No significant effects were obtained on the number of arguments that subjects could recall either, although—consistent with the "depth of processing" notion—there was a slight tendency for high involvement subjects to recall more arguments ($M = 3.75$) than did low involvement subjects ($M = 3.2$), $F(1, 68) = 2.07$, $p < .15$. The interaction on this measure did not approach significance ($p > .25$).

Correlational Analyses

Table 3 presents the correlations among the attitude and cognitive response measures separately for high and low involvement subjects. The pattern of correlations indicates that message-relevant cognitive responses are better predictors of attitude change under high than

Table 2
Attitudes and Cognitive Responses in Relation to Involvement and Quality of Message Arguments

Item	Weak arguments		Strong arguments	
	Involvement		Involvement	
	Low	High	Low	High
Attitude	$-.24_a$	$-.67_b$	$.20_c$	$.71_d$
Counterarguments	2.11_a	3.28_b	2.05_a	1.33_a
Favorable thoughts	$.88_a$	1.11_a	1.44_a	2.94_b

Note. Means in any given row without a common subscript are significantly different at the .05 level by the Newman–Keuls procedure.

Table 3
Correlations Among Attitude and Cognitive Responses for High and Low Involvement Conditions

Item	Low involvement			High involvement		
	Counter-arguments	Favorable thoughts	Recall	Counter-arguments	Favorable thoughts	Recall
Attitude	−.22	.35*	.01	−.73**	.64**	.19
Counterarguments	—	−.26	.09	—	−.51**	.02
Favorable thoughts	—	—	.06	—	—	.10

* $p < .05$. ** $p < .001$.

under low involvement conditions. This replicates the findings of Chaiken (Note 1) and Petty & Cacioppo (1979) and is consistent with the view that high involvement increased the importance of message processing in producing persuasion. The ability to recall message arguments did not allow any reliable prediction.[2]

Discussion

The pattern of data in both experiments provides strong support for the cognitive response view of the effects of increased issue involvement on persuasion. The results of both experiments contradicted the original social judgment theory formulation (Sherif et al., 1965), which proposed that increased involvement would invariably reduce persuasion. In addition, the results of Experiment 2 extended the Pallak et al. (1972) formulation, which limited the persuasion-facilitating effects of increased issue involvement to extreme consonant messages. Experiment 2 demonstrated that increased involvement could lead to increased persuasion for a counterattitudinal advocacy if the arguments were sufficiently compelling. This suggests that it is not the direction of the advocacy (proattitudinal or counterattitudinal) that is important, but rather the nature of the cognitive responses elicited.

Of course it is likely that proattitudinal advocacies will generally elicit favorable thoughts, whereas counterattitudinal advocacies will elicit primarily counterarguments, but this will not always be the case. Recent research has demonstrated the feasibility of constructing communications that take proattitudinal positions but employ arguments that elicit primarily counterarguments, and messages that take counterattitudinal positions but elicit primarily favorable thoughts (cf. Petty et al., 1976). The latter was also accomplished in the present investigation and allowed a strong empirical test of the hypothesis of interest.

It is also important to note that the attitudinal effects of issue involvement do not appear to be mediated by enhancing recall of the message arguments. Although high involvement did tend to increase argument recall for both the strong and weak messages in Experiment 2, high involvement increased persuasion for the strong message but decreased persuasion for the weak one. Also, consistent with previous research (e.g., Cacioppo & Petty, 1979; Insko, Lind, & LaTour, 1976; etc.), the within-cell correlations failed to substantiate a relationship between learning and persuasion.

The most compelling explanation for the present data appears to be that increased involvement enhances the importance of message content in producing persuasion. If the message content elicits primarily counterarguments, then increased involvement will tend to enhance the production of thoughts unfavorable to the advocacy and will result in decreased agreement; but if the message content elicits primarily favorable cognitions,

[2] Similar within-cell correlations among attitudes and cognitive responses were available for Experiment 1. The only significant correlation (based on an N of only 12) was between favorable thoughts and attitudes within the high involvement groups ($r = .58$, $p < .05$). Message recall was not assessed in the first study.

then involvement will tend to enhance these positive thoughts and will result in increased agreement. The correlational data from Experiment 2 further corroborated the view that increased involvement enhances the importance of message-based cognitions in producing persuasion. The correlations between attitudes and cognitive responses were substantially greater under high than under low involvement conditions.

The finding that increased involvement enhances the importance of message factors in producing persuasion may explain the previously confusing finding that sources of high and low credibility produce differential persuasion under low but not under high involvement conditions (e.g., Johnson & Scileppi, 1969; Rhine & Severance, 1970). The present analysis suggests that nonmessage cues such as the expertise or attractiveness of a source should have maximal impact when persuasion is not tied to an extensive processing of the message content, as when a message is on a topic of low involvement. On the other hand, characteristics of the message content should have maximal impact under high involvement conditions. This analysis suggests that low involvement persuasion situations may be governed by what cognitive psychologists have called "automatic processing," whereas high involvement persuasion situations may be governed more by "controlled processing" (Atkinson & Shiffrin, 1968; LaBerge, 1975; Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977). Under the former situations, message acceptance would be determined more by a well-practiced script such as "Experts are to be believed" (cf. Abelson, 1976; Langer, 1978), whereas under the latter situations message acceptance would be determined more by a subject's attention to and processing of the message content. In support of this conjecture, Chaiken (Note 1) reported that subjects were more affected by the number of arguments employed in a message under high than under low involvement conditions but were more affected by source attractiveness under low than under high involvement conditions.

At least two important questions about involvement remain unaddressed by the present research. One concerns *why* increased involvement should facilitate information processing. A possible explanation may reside in recent research indicating that information with self-relevance is processed more quickly than non-self-relevant information. For example, Markus (1977) argued that subjects have a more extensively developed "schema" or cognitive structure for self-relevant information and that this schema facilitates processing.

Another important question concerns the limitations on increased issue involvement facilitating processing. We suspect that there are circumstances where involvement may be so high, as when an issue is intimately associated with certain central values (cf. Ostrom & Brock, 1968), that processing will terminate in the interest of self-protection. This level of involvement was not reached in the present investigation, even though the message advocated a change in university policy that, if implemented, might have prevented many students from obtaining their degrees. Thus, within a normal range of involvement, increased issue importance appears to be associated with increased message processing. This enhanced processing will most likely lead to reduced persuasion when a message presents weak arguments (i.e., arguments that are open to refutation and counterargumentation), and to enhanced persuasion when a message presents particularly good arguments for which subjects have no readily available counterarguments (and thus favorable thoughts will predominate).

Reference Note

1. Chaiken, S. *Use of source and message cues in persuasion.* Paper presented at the meeting of the American Psychological Association, Toronto, Canada, August 1978.

References

Abelson, R. P. Script processing in attitude formation and decision making. In J. Carroll and J. Payne (Eds.), *Cognition and social behavior.* Hillsdale, N.J.: Erlbaum, 1976.

Apsler, R., & Sears, D. Warning, personal involvement, and attitude change. *Journal of Personality and Social Psychology,* 1968, *9,* 162–166.

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In

K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 2). New York: Academic Press, 1968.

Brock, T. C. Communication discrepancy and intent to persuade as determinants of counterargument production. *Journal of Experimental Social Psychology,* 1967, *3,* 296–309.

Cacioppo, J. T., & Petty, R. E. Effects of message repetition and position on cognitive responses, recall, and persuasion. *Journal of Personality and Social Psychology,* 1979, *37,* 97–109.

Cialdini, R., Levy, A., Herman, P., Kozlowski, L., & Petty, R. Elastic shifts of opinion: Determinants of direction and durability. *Journal of Personality and Social Psychology,* 1976, *34,* 663–672.

Cook, T. D. Competence, counterarguing, and attitude change. *Journal of Personality,* 1969, *37,* 342–358.

Craik, F. M., & Lockhart, R. S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior,* 1972, *11,* 671–684.

Dean, R., Austin, J., & Watts, W. Forewarning effects in persuasion: Field and classroom experiments. *Journal of Personality and Social Psychology,* 1971, *18,* 210–221.

Eagly, A. H. Involvement as a determinant of response to favorable and unfavorable information. *Journal of Personality and Social Psychology Monograph,* 1967, *7*(3, Pt. 2).

Eagly, A. H., & Manis, M. Evaluation of message and communication as a function of involvement. *Journal of Personality and Social Psychology,* 1966, *3,* 483–485.

Freedman, J. L. Involvement, discrepancy, and opinion change. *Journal of Abnormal and Social Psychology,* 1964, *69,* 290–295.

Greenwald, A. G. Cognitive learning, cognitive response to persuasion, and attitude change. In A. Greenwald, T. Brock, & T. Ostrom (Eds.), *Psychological foundations of attitudes.* New York: Academic Press, 1968.

Halverson, R. R., & Pallak, M. S. Commitment, ego-involvement, and resistance to attack. *Journal of Experimental Social Psychology,* 1978, *14,* 1–12.

Hovland, C. I. Reconciling conflicting results derived from experimental and survey studies of attitude change. *American Psychologist,* 1959, *4,* 8–17.

Insko, C. A., Lind, E., & LaTour, S. Persuasion, recall, and thoughts. *Representative Research in Social Psychology,* 1976, *7,* 66–78.

Insko, C. A., Turnbull, W., & Yandell, B. Facilitating and inhibiting effects of distraction on attitude change. *Sociometry,* 1974, *4,* 508–528.

Johnson, H. H., & Scileppi, J. A. Effects of ego-involvement conditions on attitude change to high and low credibility communicators. *Journal of Personality and Social Psychology,* 1969, *13,* 31–36.

Kiesler, C. *The psychology of commitment.* New York: Academic Press, 1971.

Kiesler, C., Collins, B., & Miller, N. *Attitude change.* New York: Wiley, 1969.

Kleinhesselink, R. R., & Edwards, R. E. Seeking and avoiding belief-discrepant information as a function of its perceived refutability. *Journal of Personality and Social Psychology,* 1975, *31,* 787–790.

LaBerge, D. Acquisition of automatic processing in perceptual and associative learning. In P. M. A. Rabbit & S. Dornic (Eds.), *Attention and performance* (Vol. 5). New York: Academic Press, 1975.

Langer, E. J. Rethinking the role of thought in social interaction. In J. H. Harvey, W. I. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research* (Vol. 2). Hillsdale, N.J.: Erlbaum, 1978.

Madsen, D. B. Issue importance and group choice shifts: A persuasive arguments approach. *Journal of Personality and Social Psychology,* 1978, *36,* 1118–1127.

Markus, H. Self-schemata and processing information about the self. *Journal of Personality and Social Psychology,* 1977, *35,* 63–78.

Miller, N. Involvement and dogmatism as inhibitors of attitude change. *Journal of Experimental Social Psychology,* 1965, *1,* 121–132.

Ostrom, T. M., & Brock, T. C. A cognitive model of attitudinal involvement. In R. Abelson, E. Aronson, W. McGuire, T. Newcomb, M. Rosenberg, & P. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook.* Chicago: Rand McNally, 1968.

Pallak, M. S., Mueller, M., Dollar, K., & Pallak, J. Effect of commitment on responsiveness to an extreme consonant communication. *Journal of Personality and Social Psychology,* 1972, *23,* 429–436.

Petty, R. E., & Cacioppo, J. T. Forewarning, cognitive responding, and resistance to persuasion. *Journal of Personality and Social Psychology,* 1977, *35,* 645–655.

Petty, R. E., & Cacioppo, J. T. Effects of forewarning of persuasive intent and involvement on cognitive responses and persuasion. *Personality and Social Psychology Bulletin,* 1979, *5,* 173–176.

Petty, R. E., Ostrom, T. M., & Brock, T. C. (Eds.), *Cognitive responses in persuasion.* Hillsdale, N.J.: Erlbaum, in press.

Petty, R. E., Wells, G. L., & Brock, T. C. Distraction can enhance or reduce yielding of propaganda: Thought disruption versus effort justification. *Journal of Personality and Social Psychology,* 1976, *34,* 874–884.

Rhine, R., & Severance, L. Ego-involvement, discrepancy, source credibility, and attitude change. *Journal of Personality and Social Psychology,* 1970, *16,* 175–190.

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology,* 1977, *35,* 677–688.

Schneider, W., & Shiffrin, R. M. Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review,* 1977, *84,* 1–66.

Sherif, M., & Hovland, C. *Social judgment.* New Haven, Conn.: Yale University Press, 1961.

Sherif, C. W., Kelley, M., Rodgers, H., Sarup, G., &

Tittler, B. Personal involvement, social judgment and action. *Journal of Personality and Social Psychology*, 1973, *27*, 311–328.

Sherif, C. W., Sherif, M., & Nebergall, R. E. *Attitude and attitude change*. Philadelphia: Saunders, 1965.

Shiffrin, R. M., & Schneider, W. Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review*, 1977, *84*, 127–190.

Tesser, A. Self-generated attitude change. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 11). New York: Academic Press, 1978.

Triandis, H. *Attitude and atttitude change*. New York: Wiley, 1971.

Zimbardo, P. H. Involvement and communication discrepancy as determinants of opinion conformity. *Journal of Abnormal and Social Psychology*, 1960, *60*, 86–94.

Received December 26, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 10, 1927–1931

Toward a Viable Theory of Individual Differences

Joseph R. Royce
Center for Advanced Study in Theoretical Psychology
University of Alberta, Edmonton, Canada

This article deals with two metatheoretic issues: (a) the problem of making scientific inferences and (b) the problem of the conceptual framework. Inability to overcome inductive uncertainty lies at the core of the debates concerning the interpretation of factors and how they are organized. Failure to embed factors in a viable, process-oriented conceptual framework has also obstructed progress. Theoretical structures based on the information-processing paradigm and invariant factors are the most promising extant approaches.

This article is focused on conceptual issues rather than data. In particular, I am concerned with how we can best evolve a viable theory of individual differences (Royce, 1973a,b, 1978; Royce & Powell, Notes 1, 2). Theoretical analysis and synthesis *inevitably* involve metatheoretic issues as well as substantive issues (Royce, 1976). This interplay between metatheory and substantive theory is, of course, well recognized in advanced scientific domains and in philosophy of science. The message from these quarters is clear—Progress in developing substantive theory frequently requires clarification of underlying metatheoretic issues. Two such issues are central to this article: (a) the problem of scientific inference and (b) the problem of the conceptual framework.

The Problem of Scientific Inference

The argument is that factor analysis is as subject to the vagaries of inference as any other empiricoinductive scientific method. This means it is logically impossible to be *certain* about the interpretation of a factor (i.e., scientific inference involves inductive leaps.)[1] Thus, a major point that emerges is that the problem of scientific inference is not unique to factor methodology—It poses difficulties for all scientists, and for the philosophers of science as well (Hanson, 1961; Morgan, 1973; Popper, 1959, 1963; Rozeboom, 1961). However, it does *not* follow that factor methodology has a fatal disease and must therefore be buried. Rather, what is called for is that we spell out the implications of this and other metatheoretic issues in the service of advancing substantive theory.

While it has long been recognized that differential psychology might provide the key to our understanding of personality, the Achilles' heel of this approach has been the problem of identifying the fundamental dimensions of individuality. And although factor analysis as the best available scientific method for identifying the dimensions of organized complexity has reduced the number of required personality constructs from thousands to hundreds, the current state of the art is such that there is no mechanical algorithm available for deciding which dimensions are invariant. This means it is necessary to make qualitative judgments concerning this issue. It is my view that *the problem of factor invariance under-*

This article is a product of the Individuality Project, which is supported by Canada Council Grant Number S76-0908-R1.

Requests for reprints should be sent to Joseph R. Royce, Center for Advanced Study in Theoretical Psychology, University of Alberta, Edmonton, Alberta, Canada T6G 2E9.

[1] A more complete elaboration of the philosophical problem of induction and the logical basis for making scientific inferences via the methodology of factor analysis is provided in my article for the *Nebraska Symposium* (Royce, 1976, see especially pp. 13–20).

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3710-1927$00.75

lies the substantive conflicts and confusions concerning factors and their organization.

Contrary to common belief, there are two categories of factor invariance rather than one. These have been labeled *internal* and *external* invariance (Royce, 1976). Invariance of the first kind involves the replication of factors via several analyses despite differences in subjects, variables, and occasions. This constitutes the standard attack on the problem, which although essentially qualitative in nature, is nevertheless an effective and plausible procedure. For example, in high intensity domains of investigation, such as pilot aptitude (e.g., see the research of the U.S. Air Force Aviation Psychology Program, both the 19 volumes from World War II and the postwar reports) or rodent emotionality (Royce, 1977b), "marker variables" are so well-known that they constitute reliable and valid indices for specifiable factors.

External invariance refers to the empirical and/or laboratory manipulation of previously identified factors. The most extensive research programs devoted to experimental manipulation of factors include Fleishman's (1967) analyses of the learning dynamics of psychomotor dimensions, Eysenck's (1967) massive laboratory investigations of his three third-order constructs, and Royce's (1966, 1973a; Aftanas & Royce, 1969; Mos, Lukaweski, & Royce, 1977; Royce, Holmes, & Poley, 1975; Royce & Poley, 1973; Royce, Poley, & Yeudall, 1973; Royce, Yeudall, & Bock, 1976) identification of the genetic and brain site correlates of affective and cognitive factors. Although the call for combining multivariate and experimental method and thought was sounded some time ago (e.g., Cattell, 1966a, 1966b; Cronbach, 1957; Royce, 1950), it is unfortunately true that the volume of output that combines the two modes of attack is relatively small. The best single exemplar of this research paradigm is still embodied in the Cattell (1966a) handbook. But the Society of Multivariate Experimental Psychology journal *Multivariate Behavioral Research* has not been very effective in this bridging role (e.g., see Royce, 1977a), despite the hopes of the founding fathers (Cattell, 1966b). Multivariate experimental research is very demanding, both in terms of orchestrating the relevant skills and in funding, so it will probably be some time before it hits full stride. Let me briefly cite two examples in order to make this point plainer. First example: There are no longitudinal data available as an empirical basis for a multivariate developmental psychology. The available theoretical models (e.g., Kearsley, Buss, & Royce, 1977) are based on extrapolations to the life span from cross-sectional data. Large-scale testing of an appropriate sample of subjects and factors throughout the life span can only be accomplished by establishing a massive research institute that can outlive its investigators. Second example: There is no large-scale research program devoted to identifying the brain correlates of factors. There have been sporadic factorial studies of humans with brain damage, but there has been no sustained experimental attack on the problem of the brain correlates of factors. Such a program would require experimental lesioning, microelectrode implantations, and other brain intervention techniques on appropriate animal populations (e.g., subhuman primates in the case of cognitive factors), in addition to a sophisticated and sustained factor strategy. Proper implementation of such a research program would require an interdisciplinary team that would range across factor analysis, psychometrics, animal behavior, and neuropsychology. It would also require access to a large number of humans with a broad range of brain damage and extensive laboratory and animal facilities as well, so that a variety of biological manipulations of psychological factors could be carried out. These examples bring out the cost factor on the one hand and the potential value of such investigations on the other hand. I cite these examples in order to emphasize the point that we need both internal and external invariance strategies in order to establish the "boundary conditions" (e.g., Factor A is functional at the early stages of learning a complex motor task, but nonfunctional during the later stages) or the range of effectiveness for each factorially identified unknown.

Although the intent of efforts to find a mathematical solution to the invariance prob-

em is to generate invariant factors algorithmically, this may not be possible in principle (for example, because of an insufficiency of information concerning the factors under comparison). This means that the present state of the art is such that subjective judgments regarding the interpretation and invariance of factors cannot be avoided. Thus, the hypothesis-generating character of factor theory must be taken seriously—that is, the interpretation that Factor x is concerned with x and not x' or y is a tentative hypothesis. However, the principle of *empirical tenability* says that as factor investigators increase their knowledge of the "boundary conditions" of factors, these empirical constraints will increase the probability that the nth interpretation of Factor x is a more *plausible hypothesis* (Royce, 1976).

Problems of factor invariance are compounded by methodological difficulties and a relative paucity of data when we move on to higher order factors. For example, there are ambiguities concerning causal influences between levels, there are problems concerning the partitioning of the common factor variance between levels, and there are ambiguities concerning the conceptual status of higher strata factors (e.g., see Cattell, 1965a, for the best available treatment of these issues). But despite these methodological difficulties, all factor analysts share the same goal—namely, to identify invariant factors, to specify how they are organized, and to specify how factors function as parts of complex behaviors. Thus we must transcend methodological and other shortcomings in pursuit of these common goals (e.g., via such requirements as factor invariance), if we are to overcome the historical conflicts (exemplified once again in the current exchange between Guilford, 1975, 1977, and Eysenck, 1977) that have dominated this complex and important area of investigation.

The Problem of the Conceptual Framework

A critical implication of this problem is that the mere inventorying of invariant factors *cannot* generate advanced forms of science because explanatory theory requires viable conceptual frameworks on which to hang empirical observables (Royce, 1976, 1978). The point is that factor analysis per se is incapable of providing the prerequisite conceptual framework for its substantive findings. What it can do is identify potentially unifying components of such a framework (Royce, 1963), but the subsuming model (of which factors are critical parts) must come from elsewhere. This point has not been adequately perceived by either the factor analysts or their critics. The major shortcoming of the factor analytic proponents is their belief that a theoretical structure could somehow be automatically generated as a natural consequence of inventorying invariant factors. The major implication of this view is the belief that the factor model constitutes a sufficient framework on which to hang its substantive observables. Thus its major proponents, such as Guilford (1967) and Cattell (1965b, 1973), have tended to develop theoretical structures based primarily on the factor model. A major consequence of this conceptual commitment is that personality theory too heavily based on factor analysis has not been able to go beyond structure to dynamics.

While there have been occasional noises from within the factor subcommunity concerning the need to get at process (e.g., see Messick, 1973), it is the current interaction between the two sciences (i.e., multivariate and experimental psychology) that is forcing the issue of conceptual framework into such sharp focus. For it is the experimental tradition that has kept its eye on process. But it is the importation of a new ingredient, the information-processing paradigm, that is of particular importance in this context. For this new paradigm is the conceptual framework that is bringing multivariate and experimental psychology together in an effort to understand complex psychological phenomena. The most important empirical work in this vein is being conducted in the cognitive domain (Carroll, 1976, 1978; Hunt, 1974; Hunt, Frost, & Lunneborg, 1973; Snow, 1976a,b; and Sternberg, 1977). And the most explicit and comprehensive theoretical structure to combine the information-processing paradigm with the two psychologies is Royce's multifactor system theory of individual differences (see Royce &

Buss, 1976; Royce, Kearsley, & Klare, 1978; Royce & Powell, Note 1 and Note 2, for an overview). This theory defines personality as the total psychological unit, a suprasystem that is composed of six interacting systems. The cognitive (Diamond & Royce, in press; Powell & Royce, Note 3) and affective (Royce & McDermott, 1977) systems are conceptualized as processing units that function as information transformers. The style (Wardell & Royce, 1978) and value (Schopflocher, Royce, & Meehan, Note 4) systems are also central processing units, but they function more as personality integrators. Finally, the sensory (Kearsley & Royce, 1977) and motor (Powell, Katzko, & Royce, 1978) systems are more peripheral processing units that function as input/output transducers and encoders. These six systems are organized as a multilevel, hierarchical suprasystem (Mesarovic, Macko, & Takahara, 1970) in which there is a controlled-process stratum (sensory, motor), a learning–coping stratum (cognitive, affective), and an integrative stratum (style, value). In turn, each of the six systems is also conceptualized as a multilevel, hierarchical system, where the elements of the hierarchies are identified via factor analysis.

The long-range goal of factor research is to understand variations in psychological functioning. But this kind of understanding implies the availability of a viable theory, which in turn implies the availability of an appropriate conceptual framework. The most germane candidate is the information-processing paradigm. This paradigm has the potential of generating a viable, dynamic theory of individual differences.

Reference Notes

1. Royce, J. R., & Powell, A. *A factor-system theory of individuality*. Book in preparation, 1982.
2. Royce, J. R., & Powell, A. *An overview of multifactor system theory* (Center Paper). Unpublished manuscript, University of Alberta, Edmonton, Canada, 1979.
3. Powell, A., & Royce, J. R. *Cognitive information processing: The role of individual differences in the search for invariants* (Center Paper). Unpublished manuscript, University of Alberta, Edmonton, Canada, 1979.
4. Schopflocher, D. A., Royce, J. R., & Meehan, J. *A multi-factor theory of values* (Center Paper). Unpublished manuscript, University of Alberta, Edmonton, Canada, 1979.

References

Aftanas, M. S., & Royce, J. R. A factor analysis of brain damage tests administered to normal subjects with factor score comparisons across ages. *Multivariate Behavioral Research*, 1969, *4*, 459–481.

Carroll, J. B. Psychometric tests as cognitive tasks: A new "structure of intellect." In L. B. Resnick (Ed.), *The nature of intelligence*. Hillsdale, N.J.: Erlbaum, 1976.

Carroll, J. B. How shall we study individual differences in cognitive abilities?—Methodological and theoretical perspectives. *Intelligence*, 1978, *2*, 87–115.

Cattell, R. B. Higher order factor structures: Reticular vs. hierarchical formulae for their interpretation. In C. Banks & P. L. Broadhurst, *Studies in psychology, presented to Sir Cyril Burt*. London: University of London Press, 1965. (a)

Cattell, R. B. *The scientific analysis of personality*. London: Penguin, 1965. (b)

Cattell, R. B. *Handbook of multivariate experimental psychology*. Chicago: Rand McNally, 1966. (a)

Cattell, R. B. Guest editorial: Multivariate behavioral research and the integrative challenge. *Multivariate Behavior Research*, 1966, *1*, 4–42. (b)

Cattell, R. B. *Personality and mood by questionnaire*. San Francisco: Jossey-Bass, 1973.

Cronbach, L. J. The two disciplines of scientific psychology. *American Psychologist*, 1957, *12*, 671–684.

Diamond, S., & Royce, J. R. Cognitive abilities as expressions of three ways of knowing. *Multivariate Behavioral Research*, in press.

Eysenck, H. J. *The biological basis of personality*. Springfield, Ill.: Charles C Thomas, 1967.

Eysenck, H. J. Personality and factor analysis: A reply to Guilford. *Psychological Bulletin*, 1977, *84*, 405–411.

Fleishman, E. A. Individual differences in motor learning. In R. Gagne (Ed.), *Learning and individual differences*. Columbus, Ohio: Merrill, 1967.

Guilford, J. P. *The nature of human intelligence*. New York: McGraw-Hill, 1967.

Guilford, J. P. Factors and factors of personality. *Psychological Bulletin*, 1975, *82*, 802–814.

Guilford, J. P. Will the real factor of extraversion-introversion please stand up? A reply to Eysenck. *Psychological Bulletin*, 1977, *84*, 412–416.

Hanson, N. R. *Patterns of discovery*. Cambridge, England: Cambridge University Press, 1961.

Hunt, E. B. Quote the raven? Nevermore. In L. Gregg (Ed.), *Knowledge and cognition*. New York: Halsted Press, 1974.

Hunt, E. B., Frost, N., & Lunneborg, C. Individual differences in cognition: A new approach to intelligence. In G. Bower (Ed.), *Advances in learning and motivation* (Vol. 7). New York: Academic Press, 1973, pp. 87–122.

Kearsley, G., Buss, A., & Royce, J. R. Developmental change and the cognitive system. *Intelligence*, 1977, *1*, 257–273.

Kearsley, G., & Royce, J. R. A multifactor theory of sensation: Individuality in sensory structure and sensory processing. *Perceptual and Motor Skills*, 1977, *44*, 1299–1316.

Mesarovic, M. D., Macko, D., & Takahara, Y. *Theory of hierarchical multilevel systems*. New York: Academic Press, 1970.

Messick, S. Multivariate models of cognition and personality: The need for both process and structure in psychological theory and measurement. In J. R. Royce (Ed.), *Multivariate analysis and psychological theory*. London: Academic Press, 1973.

Morgan, C. G. On the algorithmic generation of hypotheses. *Scientia*, 1973, *108*, 533–598.

Mos, L., Lukaweski, R., & Royce, J. R. The effect of septal lesions on factors of mouse emotionality. *Journal of Comparative and Physiological Psychology*, 1977, *91*, 523–532.

Popper, K. R. *The logic of scientific discovery*. New York: Basic Books, 1959.

Popper, K. R. *Conjectures and refutations*. London: Routledge & Kegan Paul, 1963.

Powell, A., Katzko, M., & Royce, J. R. A multifactor-system theory of the structure and dynamics of motor function. *Journal of Motor Behavior*, 1978, *10*, 191–120.

Royce, J. R. A synthesis of experimental designs in program research. *Journal of General Psychology*, 1950, *43*, 295–303.

Royce, J. R. Factors as theoretical constructs. *American Psychologist*, 1963, *18*, 522–528.

Royce, J. R. Concepts generated from comparative and physiological psychological observations. In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology*. Chicago: Rand McNally, 1966, 642–683.

Royce, J. R. The conceptual framework for a multifactor theory of individuality. In J. R. Royce (Ed.), *Multivariate analysis and psychological theory*. London: Academic Press, 1973. (a)

Royce, J. R. (Ed.). *Multivariate analysis and psychological theory: The third Banff conference on theoretical psychology*. London: Academic Press, 1973. (b)

Royce, J. R. Psychology is multi: Methodological, variate, epistemic, world-view, systemic, paradigmatic, theoretical, and disciplinary. In W. J. Arnold (Ed.), *Nebraska Symposium on the Conceptual Foundations of Theory & Methods in Psychology* (Vol. 23). Lincoln: University of Nebraska Press, 1976.

Royce, J. R. Guest editorial: Have we lost sight of the original vision for SMEP and MBR? *Multivariate Behavioral Research*, 1977, *12*, 135–141. (a)

Royce, J. R. On the construct validity of open field measures. *Psychological Bulletin*, 1977, *84*, 1098–1106. (b)

Royce, J. R. How we can best advance the construction of theory in psychology. *Canadian Psychological Review*, 1978, *19*, 259–276.

Royce, J. R., & Buss, A. R. The role of general systems and information theory in multi-factor individuality theory. *Canadian Psychological Review*, 1976, *17*, 1–21.

Royce, J. R., Holmes, T. M., & Poley, W. Behavior genetic analysis of mouse emotionality: III. The diallel analysis. *Behavior Genetics*, 1975, *5*, 351–372.

Royce, J. R., Kearsley, G., & Klare, W. The relationship between factors and psychological processes. In C. Brainerd & J. Scandura (Eds.), *NATO advanced study institute on structural/process theories of complex behavior*. Leyden, The Netherlands: A. W. Sijthoff International, 1978.

Royce, J. R., & McDermott, J. A multi-dimensional system dynamics model of affect. *Motivation and Emotion*, 1977, *1*, 193–224.

Royce, J. R., & Poley, W. Behavior genetic analysis of mouse emotionality: II. Stability of factors across genotypes. *Animal Learning & Behavior*, 1973, *1*, 116–120.

Royce, J. R., Poley, W., & Yeudall, L. T. Behavior genetic analysis of mouse emotionality: I. Factor analysis. *Journal of Comparative and Physiological Psychology*, 1973, *83*, 36–47.

Royce, J. R., Yeudall, L. T., & Bock, C. Factor analytic studies of human brain damage: I. First and second order factors and their brain correlates. *Multivariate Behavioral Research*, 1976, *4*, 381–418.

Rozeboom, W. W. Ontological induction and logical typology of scientific variables. *Philosophy of Science*, 1961, *28*, 337–377.

Snow, R. E. *Research on aptitudes: A progress report* (Technical Report No. 1, Aptitude Research Project). Palo Alto, Calif.: Stanford University Press, 1976. (a)

Snow, R. E. *Theory and method for research on aptitude processes: A prospectus* (Technical Report No. 2, Aptitude Research Project). Palo Alto, Calif.: Stanford University Press, 1976. (b)

Sternberg, R. J. *Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities*. Hillsdale, N.J.: Erlbaum, 1977.

Wardell, D., & Royce, J. R. Toward a multi-factor theory of styles and their relationships to cognition and affect. *Journal of Personality*, 1978, *46*, 474–505.

Received March 12, 1979 ■

New from THE FREE PRESS...

SYMLOG
A System for the Multiple Level Observation of Groups

ROBERT F. BALES and STEPHEN P. COHEN, with STEPHEN A. WILLIAMSON

This unique, theoretically sound, and highly practical system for studying group dynamics presents a step-by-step method for collecting, summarizing, and analyzing data on group behavior, and a novel feedback technique that helps members resolve their difficulties.

560 pages *ISBN 0-02-901300-3* *$19.95*
Case Study Kit *ISBN 0-02-901310-0* *$4.95*

MANAGERIAL ASSESSMENT
An International Comparison

BERNARD M. BASS and PHILIP C. BURGER

This monumental international survey of the attitudes and behavior of corporate executives in 12 nations and regions documents which managerial traits lead to success in different countries. "Far more comprehensive in scope ... than anything previously written on the subject." —Victor Vroom, Yale University

224 pages *ISBN 0-02-901960-5* *$15.95*

BETWEEN PUBLIC AND PRIVATE
Lost Boundaries of the Self

JOSEPH BENSMAN and ROBERT LILIENFELD

The authors integrate Freudian, neo-Freudian, and symbolic interactionist theories with theories or bureaucracy, mass communication, class, and politics to shed new light on the concepts of public and private as roles that mediate the relation between individual psychological needs and societal and cultural demands.

cloth 196 pages *ISBN 0-02-902690-3* *$13.95*
paper 196 pages *ISBN 0-02-902680-6* *$7.95*

INTERGROUP PROCESSES
A Micro-Macro Perspective

HUBERT M. BLALOCK and PAUL H. WILKEN

The authors provide a much needed link between micro-level theories of the motives, behavior, and interactions of individuals and macro-level theories concerned with complex organizations, nation-states, social classes and racial groups. And, they demonstrate that macro-level theories must take account of behavioral assumptions.

540 pages *ISBN 0-02-903620-8* *$22.9*

HUMAN SOCIOBIOLOGY
A Holistic Approach

DANIEL G. FREEDMAN

"Freedman is one of the first social scientists to combine a professional knowledge of human behavior with a solid understanding of evolutionary theory. *Human Sociobiology* is an excellent work, the first of its kind." —Edward O. Wilson, author of *Sociobiology: The New Synthesis*

242 pages *ISBN 0-02-910660-5* *$12.9*

BIAS IN MENTAL TESTING

ARTHUR R. JENSEN

In the most detailed and thoroughly researched study ever published on this topic, Arthur Jensen examines psychometric methods for objectively detecting bias in mental testing and challenges the claim that standardized tests are culturally biased against native-born minorities.

800 pages *ISBN 0-02-916430-3* *$29.9*

UNDERSTANDING HUMAN VALUES
Individual and Societal

Edited by MILTON ROKEACH

Elaborating on the ideas put forth in his highly acclaimed work, *The Nature of Human Values*, Rokeach brings together recent contributions that advance theoretical, methodological, and empirical knowledge of the role human values play in everyday life.

288 pages *ISBN 0-02-926760-9* *$14.95*

THE FREE PRESS
A Division of Macmillan Publishing Co., Inc.
100D BROWN STREET, RIVERSIDE, NEW JERSEY 08370

THE NEW EDITION OF THE ONE THEY ALL FOLLOWED - READY TO TAKE SECOND-SEMESTER LEADERSHIP

SOCIAL PSYCHOLOGY AND MODERN LIFE
Second Edition

Patricia Niles Middlebrook, Central Connecticut State College

SOCIAL PSYCHOLOGY AND MODERN LIFE was the first introductory text to successfully combine rigorous, research-oriented scholarship with consistently engaging material that reflected students' deepest concerns, fascinated them, and made them think. More than any other, it was the text that secured the place of social psychology as one of the richest and most exciting courses in the college curriculum. The first edition has had a long and renowned life of leadership.

Succeeding texts imitated—but never equalled—its groundbreaking achievement. And now, the one they all followed is ready to take the lead again, in a stunning page-by-page revision that speaks directly to today's college generation with:

- Expert collaborating authors, including Phillip Zimbardo, Robert Sommer, Marvin Shaw, Irene Frieze, Craig Haney, Terry Amabile, and William De Jong, all contributing to chapters in their areas of expertise
- Four important new chapters: "The Environment and Human Behavior," "Being Male and Female," "Social Psychology and Criminal Law," "Loneliness"
- Exhaustive updating and new material throughout, including coverage of learned helplessness, objective self-awareness, the influence of sex roles on sexual behavior, the "Vietnam Syndrome," television and aggression, marriage and cohabitation, battered wives, and the Moonies
- Lavish updated graphics program
- More concise text
- One of the most comprehensive Instructor's Manuals published

Alfred A. Knopf/December 1979/672 pages hardbound/Order Codes: 31248, 32316 (I.M.)

For more information, write:
Alfred A. Knopf/Random House
College Review Desk, 400 Hahn Road, Westminster, MD 21157

NEW FOR 1980

ANNOUNCING!

PsycSCAN™ CLINICAL PSYCHOLOGY

PsycSCAN™ DEVELOPMENTAL PSYCHOLOGY

Two new abstract journals each with over 1,500 full abstracts per year from relevant, selected journals — APA and non-APA — in the fields of clinical and developmental psychology. These new publications are the results of more than three years of study and analysis of the information needs of APA members.

The unique scanning format of PsycSCAN allows the reader to quickly review the specially set title, index terms, abstract, and citation for each entry.

Keep current with the ongoing literature in these important fields in psychology. Plan to order one or both PsycSCAN journals when you complete your 1980 dues statement.

Each PsycSCAN journal is:

- $8 per year for APA members ($10 for nonmembers)
- Quarterly, starting in early 1980
- Available for easy check-off ordering on the 1980 dues statement

PsycSCAN™ is produced by PsycINFO™, a service of the American Psychological Association 1200 17th Street, NW Washington, DC 20036

(Continued from front cover)

Sex Role Identity and Its Relationships to Sex Role Attributes and Sex Role Stereotypes 1779
Michael D. Storms

Acknowledgment of Handicap as a Tactic in Social Interaction 1790
Albert H. Hastorf, Jeffrey Wildfogel, and Ted Cassman

Characterological Versus Behavioral Self-Blame: Inquiries Into Depression and Rape 1798
Ronnie Janoff-Bulman

Explaining Attributions for Achievement: A Path Analytic Approach 1810
William M. Bernstein, Walter G. Stephan, and Mark H. Davis

Emergencies: What Are They and Do They Influence Bystanders to Intervene? 1822
R. Lance Shotland and Ted L. Huston

Self-Awareness and Transgression in Children: Two Field Studies 1835
Arthur L. Beaman, Bonnel Klentz, Edward Diener, and Soren Svanum

Variation in Behavior Perception and Ability Attribution 1847
Darren Newtson and Richard J. Rindner

Reassertion and Giving Up: The Interactive Role of Self-Directed Attention and Outcome Expectancy 1859
Charles S. Carver, Paul H. Blaney, and Michael F. Scheier

Consistency and Bias in the Attribution of Attitudes 1871
Icek Ajzen, Carol Ann Dalto, and Daniel P. Blyth

On the Requirements of Proof: The Timing of Judicial Instruction and Mock Juror Verdicts 1877
Saul M. Kassin and Lawrence S. Wrightsman

Expectancy Shifts and the Expectancy Confidence Hypothesis 1888
Richard W. Wollert

Components of Aggression in Chickens and Conceptualizations of Aggression in General 1902
D. W. Rajecki, David R. Nerenz, Terry G. Freedenberg, and Patricia J. McCarthy

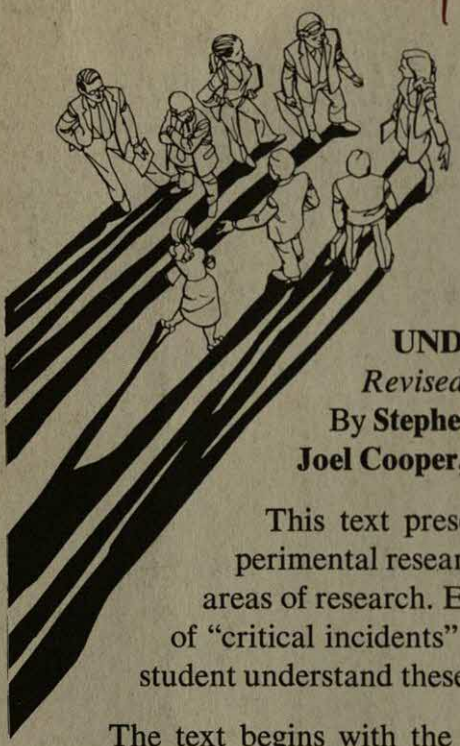
Issue Involvement Can Increase or Decrease Persuasion by Enhancing Message-Relevant Cognitive Responses 1915
Richard E. Petty and John T. Cacioppo

Toward a Viable Theory of Individual Differences 1927
Joseph R. Royce

Instructions for Submission of Manuscripts 1768

New JPSP Format Delayed Until April 1722

7/3/80



DORSEY Outshadows the Competition

UNDERSTANDING SOCIAL PSYCHOLOGY

Revised Edition

By **Stephen Worchel,** University of Virginia and **Joel Cooper,** Princeton University

This text presents both traditional and contemporary experimental research and theory with the emphasis on the active areas of research. Experimental research is placed in the context of "critical incidents" from the real world or literature to help the student understand these events.

The text begins with the reaction of individuals in the formation of attitudes, change of attitudes, and the attribution of characteristics to self and others. It then moves toward a consideration of the individual in a more macroscopic environment. **Teacher's Manual** and **Student Workbook.** 1979

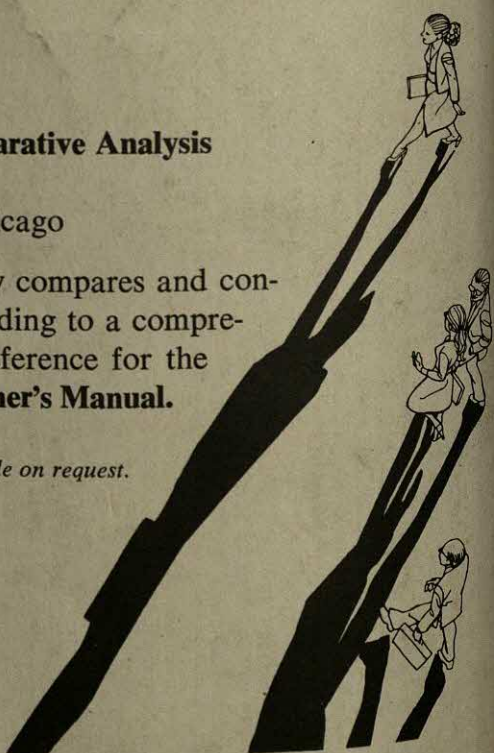
Coming in 1980

PERSONALITY THEORIES: A Comparative Analysis

Fourth Edition

By **Salvatore R. Maddi,** University of Chicago

This basic text for courses in personality compares and contrasts many theoretical viewpoints according to a comprehensive, consistent, original frame of reference for the personality field. **Teacher's Manual.**



Examination copies for adoption consideration available on request. Please indicate course title and text presently used.

The Dorsey Press
Homewood, Illinois 60430

JOURNAL OF

Personality AND Social Psychology

VOLUME 37 NUMBER 11 · NOVEMBER 1979

Coalition Bargaining in Four Games That Include a Veto Player 1933
J. Keith Murnighan and Eugene Szwajkowski

Weather, Mood, and Helping Behavior: Quasi Experiments With the Sunshine Samaritan 1947
Michael R. Cunningham

Divergence of Private Evaluations of Alternatives Prior to a Choice 1957
Paul J. Brounstein, Nancy Ostrove, and Judson Mills

Consensus Information, Prediction, and Causal Attribution: A Review of the Literature and Issues 1966
Saul M. Kassin

Learned Helplessness and Learned Effectiveness: Effects of Explicit Response Cues on Individuals Differing in Personal Control Expectancies 1982
W. Larry Gregory, George M. Chartier, and Mark H. Wright

An Empirical Analysis of the Correlations Between Leadership Status and Participation Rates Reported in the Literature 1993
R. Timothy Stein and Tamar Heller

Environmental Determinants of Memory Improvement in Late Adulthood 2003
Ellen J. Langer, Judith Rodin, Pearl Beck, Cynthia Weinman, and Lynn Spitzer

When Practice Makes Imperfect: Debilitating Effects of Overlearning 2014
Ellen J. Langer and Lois G. Imber

Idiosyncratic Weighting of Trait Information in Impression Formation 2025
Thomas M. Ostrom and Deborah Davis

The Effect of Situational Contingencies on Female Retaliative Behavior 2044
Deborah Capasso Richardson, Sandy Bernstein, and Stuart P. Taylor

(Continued on inside back cover)

Published monthly by the
AMERICAN PSYCHOLOGICAL ASSOCIATION
ISSN 0022-3514

Acting Editor

Clyde Hendrick, *University of Miami*

Action Editors

Irwin Altman, *University of Utah*
Charles S. Carver, *University of Miami*
Richard J. Harris, *University of New Mexico*
John H. Harvey, *Vanderbilt University*
William Ickes, *University of Missouri–St. Louis*
George Levinger, *University of Massachusetts—Amherst*
Brendan Gail Rule, *University of Alberta, Edmonton, Canada*
David R. Shaffer, *University of Georgia*
Cookie White Stephan, *New Mexico State University*
Barbara Strudler Wallston, *George Peabody College for Teachers*

Consulting Editors

Robert Steven Baron, *University of Iowa*
Paul A. Bell, *Colorado State University*
Ellen Berscheid, *University of Minnesota (Minneapolis)*
Michael H. Birnbaum, *University of Illinois at Urbana-Champaign*
Jack Brehm, *University of Kansas*
Philip Brickman, *Northwestern University*
J. Merrill Carlsmith, *Stanford University*
Paul Christopher Cozby, *California State University, Fullerton*
Rick Crandall, *Texas Christian University*
James M. Dabbs, Jr., *Georgia State University*
John M. Darley, *Princeton University*
Andrew R. Davidson, *University of Washington*
James H. Davis, *University of Illinois at Urbana-Champaign*
Kay Deaux, *Purdue University*
Marshall Dermer, *University of Wisconsin—Milwaukee*
Edward Donnerstein, *Iowa State University*
Ebbe B. Ebbesen, *University of California, San Diego*
Yakov M. Epstein, *Rutgers—The State University*
Irene Hanson Frieze, *University of Pittsburgh*
J. Martin Giesen, *Mississippi State University*
Lewis R. Goldberg, *University of Oregon*
William G. Graziano, *University of Georgia*
Alan E. Gross, *University of Maryland (College Park)*
Nancy Hirschberg, *University of Illinois at Chicago Circle*
Chester A. Insko, *University of North Carolina at Chapel Hill*
Alice M. Isen, *University of Maryland, Baltimore County*
Martin F. Kaplan, *Northern Illinois University*
Samuel S. Komorita, *University of Illinois at Urbana-Champaign*
Vladimir J. Konečni, *University of California, San Diego*
Ellen Langer, *Harvard University*
Mark R. Lepper, *Stanford University*
Darwyn E. Linder, *Arizona State University*
Dale T. Miller, *University of British Columbia, Vancouver, Canada*
Judson Mills, *University of Maryland (College Park)*
Warren T. Norman, *University of Michigan (Ann Arbor)*
Michael Ross, *University of Waterloo, Waterloo, Canada*
Michael F. Scheier, *Carnegie-Mellon University*
Barry R. Schlenker, *University of Florida*
Kelly G. Shaver, *College of William & Mary*
Phillip Shaver, *New York University*
Marvin E. Shaw, *University of Florida*
Steven J. Sherman, *Indiana University*
Harold Sigall, *University of Maryland (College Park)*
Mark Snyder, *University of Minnesota (Minneapolis)*
Melvin L. Snyder, *University of Texas at Austin*
Shelley E. Taylor, *Harvard University*
Stephen G. West, *Florida State University*
Steve Worchel, *University of Virginia*
Mark P. Zanna, *University of Waterloo, Waterloo, Canada*

Assistant to the Editor

Sally Miller

APA Journal Staff

Anita DeVivo, *Executive Editor*
Ann I. Mahoney, Manager, *Journal Production*
Marjorie M. Henry, *Production Supervisor*
Robert J. Hayward, *Advertising Representative*
Juanita Brodie, *Subscription Manager*

The Journal of Personality and Social Psychology publishes original reports of research, methodology, theory, criticism, and literature review in the fields of personality and social psychology of normal processes and populations. Submitted manuscripts should delineate an important question and provide a clear answer to it. In the case of research reports, this will often require demonstration that the reported results are generalizable beyond a single research setting. Consult the editorial in the January 1976 issue (Vol. 33, pp. 1–7) for a detailed statement of policies and instructions to authors.

Manuscripts: Submit manuscripts in quadruplicate, prepared according to the *Publication Manual of the American Psychological Association* (2nd ed.). Copies should be highly legible and will not be returned to authors unless they are original typewritten copies. Mail manuscripts according to instructions on p. 1439 of the August issue.

Required supplementary information: Manuscripts cannot be reviewed for publication unless the author provides supplementary information in the following categories: (a) a cover letter that indicates the relation (or the lack of relation) of the submitted manuscript to other papers by the same author(s) that are currently submitted, in press, or recently published; (b) one copy of each such related paper; (c) one copy of a statement of the procedures of the research reported in the submitted manuscript in sufficient detail to serve as a guide for conduct of an exact replication; and (d) one copy of the summary tables (e.g., correlation tables, analysis of variance summaries) for the *major* analyses reported. If it appears that the supplementary materials in categories (c) or (d) might be bulky, the author should exercise judgment and restraint in selecting what to send. It is not necessary to send supplementary material that repeats verbatim manuscript content.

After a manuscript is accepted, authors will be expected to provide assurance that their data are available in unanalyzed form for reanalyses by other investigators and that their research procedures were consistent with the American Psychological Association's published principles of research ethics except insofar as may be detailed otherwise in the manuscript.

Copyright and Permission: All rights reserved. Written permission must be obtained from the American Psychological Association for copying or reprinting text of more than 500 words, tables, or figures. Permission is normally granted contingent upon like permission of the author, inclusion of the APA copyright notice on the first page of reproduced material, and payment of a fee of $10 per page, table, or figure. Abstracting is permitted with credit to the source. Permissions and fees are waived for authors who wish to reproduce their own material for personal use; fees only are waived for authors who wish to use their own material commercially. Permission and fees are waived for the photocopying of isolated articles for nonprofit classroom or library reserve use by instructors and educational institutions. Libraries are permitted to photocopy beyond the limits of U.S. copyright law: (1) those post-1977 articles with a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through the Copyright Clearance Center, P.O. Box 765, Schenectady, NY 12301; (2) pre-1978 articles, provided that the per-copy fee stated in the Publishers' Fee List is paid through the Copyright Clearance Center, P.O. Box 765, Schenectady, NY 12301. Address requests for reprint permission to the Permissions Office, APA, 1200 Seventeenth Street, N.W., Washington, D.C. 20036.

Subscriptions: Subscriptions are available on a calendar year basis only (January through December). Nonmember rate for 1979: $60 domestic, $64 foreign, $6 single issue. APA member rate: $20. Write to Subscription Section, APA.

Change of Address: Send change of address notice and a recent mailing label to the attention of the Subscription Section, APA, 30 days prior to the actual change of address. APA will not replace undelivered copies resulting from address changes. Journals will be forwarded only if subscribers notify the local post office in writing that they will guarantee second-class forwarding postage.

Reprints: Authors may order reprints of their articles from the printer when they receive proofs.

Back Issues and Back Volumes: For information regarding back issues or back volumes write to Order Department, APA.

Microform Editions: For information regarding microform editions write to any of the following: Johnson Associates, Inc., P.O. Box 1017, Greenwich, Connecticut 06830; University Microfilms, Ann Arbor, Michigan 48106; or Princeton Microfilms, Princeton, New Jersey 08540.

Published monthly in one volume per year by the American Psychological Association, Inc., 1400 North Uhle Street, Arlington, Virginia 22201. Printed in the U.S.A. Second-class postage paid at Arlington, Va., and at additional mailing offices.

Diary No. 2...
Date 17. 4. 80
File No. Librarian
Bureau Edn. Pay Research

JOURNAL OF
Personality
AND
Social Psychology

Volume 37 November 1979 Number 11

Coalition Bargaining in Four Games That Include a Veto Player

J. Keith Murnighan and Eugene Szwajkowski
Organizational Behavior Group
Department of Business Administration
University of Illinois at Urbana-Champaign

This study tested the predictions of three models of coalition behavior in four different games, each including a veto player. Participants were master's students who played each of the four games, rotating among the five player positions between games. The games were played under one of three time pressure/default conditions: (a) no time pressure; (b) a condition such that the constant payoff to coalitions was lost if an agreement was not reached in three attempts; and (c) a condition such that the payoff for no agreement was fixed at 60 points for the veto player and 10 for the other players. The veto players' payoffs varied over games and tended to increase as play continued, at times approaching the entire payoff. Thus, the weighted probability and Roth–Shapley models were not supported; the core received some support. The default conditions had little effect. The discussion focuses on the likelihood of socially beneficial behavior in competitively motivating situations.

Gamson (1961) defined full-fledged coalition situations as those where more than two parties must make a decision, where the actors' preferences differ, and where no actor is a dictator (i.e., one who can impose a decision with no support from other parties) or veto player (i.e., one who must be included in every coalition). Most coalition research (Murnighan, 1978a) has focused on interactions that meet these requirements. However, a number of "real world" coalition situations include veto players: The oil producing countries have formed a cartel (the Organization of Petroleum Exporting Countries, or OPEC) that must be included in almost all major oil exchanges. Company presidents often must be included in all significant organizational decisions. And the members of the security council of the United Nations all hold veto power over the council's resolutions. Recently, research has also investigated coalition situations that include a veto player (e.g., Murnighan & Roth, 1977, 1978). The present study investigated four different veto games in conjunction with three time pressure/aspiration level conditions, to test both game-

The authors gratefully acknowledge the considerable assistance provided by three reviewers.

Eugene Szwajkowski is now at the University of Illinois at Chicago Circle.

Requests for reprints should be sent to J. Keith Murnighan, Department of Business Administration, 104 Commerce West, University of Illinois, Urbana, Illinois 61801.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3711-1933$00.75

Table 1
The Games, the Minimum Winning Coalitions, and the Models' Predictions for the Players' Mean Payoffs (Over All Trials)

Game	Minimum winning coalition	Models' predictions		
		Weighted probability	Core	Roth–Shapley
20(17–6–5–4–3)	AB, AC, AD, AE	80–5[a]	100–0[a,b]	80–5[a]
38(27–13–12–6–5)	AB, AC, ADE	71–11–11–3–3	100–0[a,b]	70–12–12–3–3
30(21–10–7–5–4)	AB, ACD, ACE, ADE	62–11–9–9–9	100–0[a,b]	65–15–7–7–7
28(17–8–7–6–5)	ABC, ABD, ABE, ACD, ACE, ADE	50–12.5[a]	100–0[a,b]	60–10[a]

[a] Payoff predictions that are not listed with an entry for each of the five players indicate that players B, C, D, and E are predicted to have equal payoffs.
[b] The figures listed for the core's prediction are the limits that payoffs are predicted to approach. In sequential games like those in this study, less extreme payoffs that continuously approach this limit can be taken as support for the model.

theoretic and social psychological models of coalition behavior and to establish a more general data base for veto games.

One impetus for this study is the research by Michener, Fleishman, and Vaske (1976) that investigated eight veto and eight nonveto games. The eight veto games in the study included the same set of minimum winning coalitions (i.e., those that will no longer be winning if one player is left out of the coalition). For instance, in the 14(9–6–3–2) game, where 14 refers to the number of votes needed to form a winning coalition and the numbers within the parentheses refer to the votes assigned to players designated A, B, C, and D, respectively, five different coalitions (AB, ABC, ABD, ACD, and ABCD) are winning. Only AB and ACD, however, are minimal winning. Each of the eight veto games in the Michener et al. study used this set of winning and minimum winning coalitions, with different resource distributions.

In an attempt to extend these results, the present study investigated four different five-person games that varied not only the players' resources (votes) but also the games' underlying structures (i.e., the sets of minimum winning coalitions; see Table 1). Notice that in the 20(17–6–5–4–3) game the veto player can form a winning coalition with any one of the other four players. In the 28(17–8–7–6–5) game, the opposite extreme in this set, the veto player needs two other players to form a winning coalition. The other two games fall between these two, allowing the veto player to form either one or two two-player winning coalitions. Three models of coalition behavior—the weighted probability model (Komorita, 1974), the core (Luce & Raiffa, 1957), and the Roth–Shapley value (Roth, 1977a,b; Shapley, 1953)—all make different predictions for these four games.[1] This study, then, allows for theoretical tests of three models by using four veto games that vary in their underlying structure.

Prior to discussing the other major independent variable, the time pressure/aspiration level conditions, each of the three models to be tested here will be briefly summarized, using the 30(21–10–7–5–4) game as an example.

The Weighted Probability Model

The basic assumption underlying the weighted probability model (Komorita, 1974) is that because of the logistic problem of communicating offers and counteroffers, players will attempt to form small rather than large coalitions. As the number of potential coalition members increases, the severity of the problem of achieving both reciprocity and unanimous agreement on the terms of the

[1] Other models, including Gamson's (1961) minimum resource and Komorita and Chertkoff's (1973) bargaining theories might also be tested in this study. These two models, however, either explicitly limit themselves to nonveto games or are not particularly amenable to veto games (Murnighan, 1978a) and therefore would not be appropriately tested here.

offer also increases. In addition, the number of potential defectors from the coalition increases with its size; hence, a large coalition is not only less desirable but may also be more difficult to maintain.

The weighted probability model assumes that an individual's share of the prize should be a function of the *number* and *size* of alternative minimum winning coalitions available to him or her. In the 30(21–10–7–5–4) game, Player A has four alternatives (AB, ACD, ACE, and ADE), including one two-person coalition. Player B has only one alternative (AB). Players C, D, and E each have two alternatives, both being three-person coalitions. The present form of the model assigns twice as much of an "advantage" to two-person opportunities as it does to three-person opportunities and predicts that the player's outcomes should be proportional to this "advantage." This "advantage" or weight is determined by the equation

$$P(C_j) = \frac{1/(n_j - 1)}{\Sigma 1/(n_j - 1)},$$

where $P(C_j)$ = the probability of coalition j, n_j = the number of members of coalition j and the summation is over all minimum winning coalitions. Thus, the AB coalition is predicted to form 40% of the time, and each of the three-player coalitions is predicted to form 20% of the time. The probability of inclusion for a particular player depends on the probabilities of the coalitions in which he or she is a member: Player i's probability of inclusion is equal to the sum of the probabilities of each of the coalitions he or she is a member of, $[\Sigma\, P(C_j),\ i \in C_j]$. In the 30(21–10–7–5–4) game, Player A will be included in all the winning coalitions ($P_A = 1.0$); the other players will each be included 40% of the time ($P_B = P_C = P_D = P_E = .40$). Player i's predicted payoff is proportional to his or her probability of inclusion:

$$R_{ij} = \frac{P_i}{\Sigma P_k};\ i, k \in C_j,$$

where R_{ij} = Player i's payoff (reward) in coalition j, P_i = Player i's probability of being included in a minimum winning coalition, and the summation is over all members of C_j. Payoffs in the AB coalition, then, should be 72–28; in the three-player coalitions, payoffs should be 55–22–22.

Although the original presentation of the weighted probability model (Komorita, 1974) made no predictions about the players' overall payoffs in a game (termed "success in bargaining" by Komorita and Moore, 1976), payoff predictions can be derived without additional assumptions. For any player i, the proportion of the total payoffs should equal the sum of the products of the player's predicted payoff in each coalition and the probability of that coalition. In the 30(21–10–7–5–4) game, Player A is included in every coalition and is predicted to receive a payoff of 72 on 40% of the trials and a payoff of 55 on 60% of the trials. Player A's overall mean payoff, then, should be 61.8. Player B, who should receive a payoff of 28 on 40% of the trials, should receive an overall mean payoff of 11.2. Similarly, for Players C, D, and E, the predicted mean payoffs are 8.8 each.

The Roth–Shapley Model

Unlike the weighted probability model, the Roth–Shapley model (Roth, 1977a, b; Shapley, 1953) does not predict the probabilities that different coalitions will form or the payoffs that members of the winning coalitions will receive. Rather, it specifies the proportion of the total payoff a player can expect from the play of a game.

The Roth–Shapley model is purely prescriptive: Its derivation depends only on a set of underlying assumptions concerning the format of the game. In games with constant payoffs for the winning coalition, this expectation is determined by dividing the number of times a player's resources are "pivotal" (in the sense that a losing coalition is converted to a winning coalition by the addition of this player) by the total number of permutations of the players. For instance, consider one permutation, DCEBA, in the 30(21–10–7–5–4) game. Because Player A holds veto power, he or she is the pivotal player in this permutation: The DCEB "coalition" can obtain no payoff without Player A. Indeed, in every permutation that places A third, fourth, or fifth, Player A is pivotal. Player A is also pivotal when Player B is first and A is second.

Player A is pivotal, therefore, in 75 of the 120 permutations that are possible. (In any five-person game, 51 permutations exist). Because Player B can form a winning two-player coalition with A, and the other players cannot, B will be pivotal more often than Players C, D, and E. In this game, B is pivotal 18 times, and C, D, and E are each pivotal 8 times. Dividing by the total number of permutations (120) yields expected proportions of the payoffs of (.65, .15, .07, .07, and .07) for Players A, B, C, D, and E, respectively.

The Core

As with most game-theoretic solution concepts (Murnighan, 1978a), the core (when it exists) identifies a stable payoff for each potential coalition, without predicting the relative frequencies of the different coalitions. The core is based on the concept of domination. One outcome dominates another if the members of a "new" coalition receive greater payoffs than they did before, *and* they have the power to enforce the new outcome. For example, in the 30(21–10–7–5–4) game, a coalition between A and B may form, giving A 80 and B 20. This outcome is dominated by an ADE coalition with a payoff of 90–5–5. In the ADE coalition, Players A, D, and E receive payoffs that are better than the payoffs they received in the AB coalition, and the ADE coalition can obtain the 100-point payoff. But just as the ADE (90–5–5) outcome dominates AB (80–20), every outcome can be dominated by one where Player A receives more and more points by forming a smaller coalition or one with previously excluded players. The only outcome that is *undominated* and therefore in the core of this game gives Player A the entire payoff. Although Players B, C, D, and E may not be disposed to accept this payoff, it remains the only undominated outcome and is the only outcome in the core. Like the Roth–Shapley model, the core is a purely prescriptive model; behavioral assumptions (e.g., about the players' willingness to obtain particular outcomes) are not incorporated into the model. Thus, this extreme prediction might not be expected to occur in an actual game. Nevertheless, the core implies that Player A's payoffs should increase as play in the 30(21–10–7–5–4) game continues. Indeed, whenever a game gives one player veto power, the core is the same, with the veto player receiving the entire payoff.

The core is included in, and for veto games subsumes, other game-theoretic models. In particular, the various forms of the bargaining set (Aumann & Maschler, 1964; Murnighan, 1978a) are all identical with the core for this game, and their predictions are also tested with, although not differentiated from, the core. Other game-theoretic literature (Aumann, 1964; Debreu & Scarf, 1963) also suggests that the core may result from very competitive play. Thus, support for the core would also suggest that the games used in this study were played very competitively.

Table 1 lists the predictions of each of the models for each of the games. Clearly the core makes more extreme predictions than either of the other two models. Differentiations between the weighted probability and the Roth–Shapley models are more difficult, as they have been in previous studies (Murnighan, 1978b; Murnighan & Roth, 1978). Only in the 28(17–8–7–6–5) game are the predictions very different, and then only for Player A. Thus, the data from the 28(17–8–7–6–5) game provide the best test of these two models. An additional test of the weighted probability model's predictions for the frequencies of each coalition can also be conducted. The model predicts equal frequencies for the minimum winning coalitions in the 20(17–6–5–4–3) and 28(17–6–5–4–3) games. For the 38(27–13–12–6–5) and the 30(21–10–7–5–4) games, the two-person coalitions should form twice as often as the three-person coalitions.

For the games in the Michener et al. (1976) study, the core prediction is the same as for this study; the weighted probability model predicted payoffs of 60–40 in AB coalitions, which should have formed twice as often as ACD coalitions (with 60–20–20 payoffs); and the Roth–Shapley value for the games is (58, 25, 8, 8), for overall payoffs. The data in that study show that Player A averaged well over 60 points in each of the games. For the Michener et al. (1976) study, then, both the weighted probability and

Roth–Shapley models underestimated the veto players' outcomes. Because each group played each of the games only once, it is impossible to determine whether the veto players' payoffs increased over trials. Nevertheless, their outcomes were a considerable distance from the core.

The present study affords a finer test of the models and also assesses the effects of "default" conditions, a manipulation of time pressures and aspirations, that are not incorporated into any of the models.

Time Pressure and Aspirations

Research on two-person bargaining has often manipulated the time bargainers are given to reach an agreement (e.g., Yukl, 1974). When both parties lose all or part of a payoff if they do not agree by a certain time, agreements tend to be more likely and the players' aspirations and demands tend to drop (Rubin & Brown, 1975). Time pressure has not, however, been systematically studied in coalition situations. In the present study, if no coalition formed after three attempts, one of three "default" conditions resulted: (a) All five players received a payoff of zero; (b) the veto player received 60 points, whereas each of the others received 10; or (c) bargaining continued until an agreement was reached (i.e., there were no defaults). These conditions might be expected to affect the aspirations of the veto and nonveto players differentially. For the veto player, either of the default conditions should lead to reduced aspirations, particularly as the third attempt to form a coalition approaches. In addition, the 60–10 default condition might be expected to lead to higher aspirations for the veto players than the zero default condition would (cf. Rubin & Brown, 1975). On the other hand, should the veto players' payoffs approach the core, as they did in at least one prior study (Murnighan & Roth, 1978), players may not perceive the two default conditions differently: Both may be regarded as serious setbacks.

For the nonveto players, the zero default condition might reduce each individual's own aspirations, compared to the no default condition. But there is always the possibility that "hurting" the veto player has positive value. Indeed, in previous research where the strongest of the players always received the letter "A" (e.g., Murnighan, 1978b), considerable anti-A feelings surfaced quickly. In addition, invoking the zero default may make Player A so cautious as to increase the nonveto players' payoffs in subsequent coalitions (cf. Murnighan & Roth, 1978). In the 60–10 condition, on the other hand, the nonveto players' aspirations should clearly increase. Only a small minority of the theoretical predictions (see Table 1) indicate that each of the nonveto players will be receiving 10 points on the average. Defaults, therefore, are advantageous to the nonveto players in this condition and should result in "tougher" bargaining (Bartos, 1970) even when defaults do not occur.

Considering the interactions of the default conditions with other aspects of the games, however, reduces the clarity of their potential effects. For instance, in the 60–10 default condition, the mixed-motive nature of coalition games removes restrictions on the potential payoffs that might result. Although pushing for a default as often as possible and refusing to take a payoff that is less than 10 points seems like a desirable strategy, the possibility of exclusion from the winning coalition, with loss of any payoff for that trial, may limit such a strategy's usefulness. Indeed, players who are willing to accept smaller payoffs typically are included more frequently in the winning coalition and also, at least in one study, accumulated larger overall payoffs (Murnighan & Roth, 1977). Thus the nonveto players are faced with a dilemma: Although pushing for a default payoff in either default condition may result in greater joint or long-term gain, a willingness to take a smaller payoff may potentially increase individual, short-term gain (cf. Thibaut & Kelley, 1959). Indeed, using strict definitions for rationality (Simon, 1976), individual nonveto players who reject payoffs that are greater than the default payoff on the third attempt are "irrational," at least in the short run. Add to this the possibility that giving deference to the powerful player may result in repeated inclusion in winning coalitions,

and it is clear that the nonveto players may be torn in several directions.

Although several pressures are inherent in these coalition situations, the effect of aspirations is predicted to lead to higher payoffs for the veto player in the no default condition compared to the zero and 60–10 conditions. A priori predictions contrasting the 60–10 and zero default conditions, however, are difficult to make.

Method

Subjects

Subjects were 120 master's-level students enrolled in a behavioral science course in a commerce department. For the most part, these students had little or no previous background in behavioral science. To complete the project requirements in the course, students had a choice between writing a library research paper or participating in the bargaining exercises and completing a paper analyzing their own strategies in each of the games. The performance in either option accounted for 15% of the grade in the course. Students who participated in the bargaining exercises (presented as "exercises in strategy selection") were informed that their performance in the games would be compared to the performance of other students in the same positions (i.e., with the same vote total) as themselves. Students could obtain an "A" grade for the project in any of three ways, by: (a) scoring better than the average total points for players *in the same position as themselves*, (b) scoring above the mean for similarly positioned players in the other groups in three of the four games, (c) doing an excellent job of analyzing their strategies in their paper. All students, regardless of their game performance, were required to complete the paper. Only with a relatively poor paper and below average performance in the games would a student earn a "B" for the project.[2]

Graduate students were used in this study in an attempt to parallel "real world" coalition bargainers more closely. Although none of the theories tested here speak to the expertise of the populations they address, it seems that most individuals involved in coalition bargaining, such as the examples mentioned earlier (Middle Eastern oil producers, corporation executives, and members of the United Nations Security Council), are both more intelligent and more experienced in bargaining situations than the general population. Including graduate students who have some stake in the outcomes of the bargaining and who are given repeated experience with coalition bargaining should more closely mirror such a population (cf. Murnighan, Note 1).

There were 24 5-person groups in all: 6 composed of 5 males; 7 with 1 female and 4 males; 8 with 2 females and 3 males; 2 with 3 females and 2 males; and 1 with 4 females and 1 male. Most of the 33 female participants were commerce majors.

Design and Analysis

A number of factors were manipulated in this study. Each group played each of the four games for a total of 12 trials in one of the three default conditions. Thus, games and trials were within-group variables, and default conditions was a between-group factor. Each game was played by five players, one in each of the A, B, C, D, and E positions; players were assigned different positions in the different games. Thus, player positions was a within-group variable. Finally, the games were played in one of eight different orders. Two Latin square designs were used to assign groups to the eight game–order combinations. Thus, each game was played an equal number of times in each order. The variable called "order," then, merely indicates whether a game was played 1st, 2nd, 3rd, or 4th for each group. The final design (see Table 2), then, included five crossed factors: games (4), default conditions (3), trials (12), player positions (5), and order (4), with two subjects in each unique condition. The between-group variables, again, were default conditions and order.

Because the players may have discussed the games with each other, their behavior in the games may not have been strictly independent, violating one of the assumptions of analysis of variance. To attempt to correct for this deficiency, the conservative test proposed by Box (1954), discussed by Winer (1962), was used in all analyses for the repeated factors. Although the value of the F ratios remains the same in each case, the degrees of freedom are considerably reduced. For instance, for the main effect of games on the veto player's payoffs, the normal degrees of freedom are 3 and 12; for the conservative test, the degrees of freedom are reduced to 1 and 4.

Procedure

The participants were given instructions about the coalition games in the third class of the semester, prior to the first session. Several examples of the use of the procedure (in games not used later) were discussed. Theories on bargaining were *not* included in this discussion and were only discussed after completion of the exercises. The players were told that there would be 12 trials in each game, and on each trial the winning coalition would divide a prize of 100 points among its members. They were instructed to do as well as they could (i.e., maximize their

[2] Participation by students from an instructor's own class can raise ethical issues concerning potential coercion. A copy of the safeguards taken by the authors to guard the rights of the participants in this study is available on request.

Table 2
The Assignment of the 24 Groups to the Condition–Order–Game Combinations

Condition and order	Game			
	20(17–6–5–4–3)	38(27–13–12–6–5)	30(21–10–7–5–4)	28(17–8–7–6–5)
No default				
1	4, 7	3, 8	2, 6	1, 5
2	3, 6	2, 7	1, 5	4, 8
3	2, 5	1, 6	4, 8	3, 7
4	1, 8	4, 5	3, 7	2, 6
Zero default				
1	12, 15	11, 16	10, 14	9, 13
2	11, 14	10, 15	9, 13	12, 16
3	10, 13	9, 14	12, 16	11, 15
4	9, 16	12, 13	11, 15	10, 14
60–10 default				
1	20, 23	19, 24	18, 22	17, 21
2	19, 22	18, 23	17, 21	20, 24
3	18, 21	17, 22	20, 24	19, 23
4	17, 24	20, 21	19, 23	18, 22

Note. Each of the groups included Players A–E; they played each game for 12 trials.

points) because their performance would (and did) determine part of their course grade.

Students were randomly assigned to groups, with the constraint that members be unacquainted. Each group met for their sessions at the same hour each week for four weeks. They were told which game they would be playing each week, but not which position they would hold throughout each game. Participants were told that they would not be in a "strong" position more than once. They were allowed to discuss the games with other members of their class, excluding members of their own group. In addition, the players were encouraged to formulate strategies for each position prior to each game.

Each group played each of the games for a total of 12 trials, where a trial was defined as the formation of a winning coalition. For each game, player positions were designated A, B, C, D, and E, with Player A having the most resources, Player B having the second most resources, and so forth. Assignment of players to positions was predetermined by randomly assigning members of each group to a particular "slot." The slot was unknown to the players and indicated what position each player would play in each of the four games. For instance, the first slot might mean that an individual would be assigned Player Position A in the 30(21–10–7–5–4) game, B in the 20(17–6–5–4–3) game, and so forth. The assignments within each slot were determined randomly, with the constraint that no player would be assigned the A position more than once and that the player who was not assigned to the A position in any game was given the B position in the 30(21–10–7–5–4) game. These constraints were rough attempts to establish equity among the assignments.

During the games, the players were seated around a set of opaque partitions that shielded them from view of each other and the experimenter. At each group's first session, the experimenter read specific instructions about the procedures for the games. The players made offers on each trial by means of written offer slips, which required indicating to whom one wished to send offers and also a proposal regarding the division of rewards for the prospective coalition members. For example, if Player X wished to form an XY coalition, he or she addressed an offer to Player Y and specified a division of the rewards (e.g., 60 for X and 40 for Y) on the offer slip. A player was required to send an offer slip to each player included in the proposed coalition. Thus, if a player proposed a three-person coalition, two identical offer slips were sent, one to each of the proposed coalition partners. Players were also told that they could not send an offer to one person to form one coalition and a second different offer to another person to form another coalition. This procedure, originally used by Komorita and Meek (Note 2), allowed two-, three-, and four-person coalitions to form in a single step. Thus, although large coalitions may be more difficult to form, the only difficulty inherent in this procedure is the need to fill out additional offer slips. After the players had completed the offers, the experimenter collected, examined, and distributed them to the proper persons.

After receiving an offer, each person could accept or reject it by marking "Accept" or "Reject" at the bottom of each offer slip. A person receiving more than one offer could accept at most one, unless the offers proposed the identical payoff division for the same coalition. Hence each person could only accept offers to form a single coalition on each trial. Furthermore, in determining a winning coalition, any

player's proposal, if accepted, had priority over any offer he or she might accept, thus committing the player to his or her own offer. After the offers had been accepted or rejected, the experimenter collected the offer slips and announced the winning coalition, if one had formed, and the payoff division. A coalition was declared winning if all the proposed coalition partners accepted the offer. If no coalition formed because at least one person rejected each of the proposed coalitions, the procedure was repeated. This procedure allowed for acceptance within the group of two proposals on the same trial. For instance, if A sent an offer to C, D sent an offer to A, and both offers were accepted, AC would be declared the winning coalition because A was committed to his or her offer (invalidating his or her acceptance). While the offer D sent to A did not result in a coalition, it indicated the exact nature (i.e., how much D was willing to offer) of D's interest. When two coalitions were accepted in this manner, such that each was invalidated by the other (e.g., ACD and ACE), the players were informed of the situation and the trial was rerun.

A practice trial was conducted before the start of the first session. Immediately after the practice trial, the players were assigned to their positions for that game. Lists of the resources (i.e., votes) for each position and the winning coalitions were also provided. No verbal communication was permitted thereafter; hence, the players could not identify each others' positions.

The instructions were summarized for the players at the start of their second, third, and fourth sessions. Practice trials were not run, but assignment to one of the five positions and information about the resources for each of the positions and the winning coalitions were distributed *after* the players were seated behind the partitions. Thus, for each game, the players were not informed of the identity of the players in the other positions.

Lengthy discussions of power, coalition, and bargaining theories were conducted during several classes at the end of the semester. The central focus in each of these discussions was the students' behaviors during the exercises and their analyses of strategies in their papers. Individuals were not identified in these discussions, and all other questions were answered.

Results

Sex Effects

Vinacke (1959) and Komorita and Moore (1976) have reported that males tend to be, in Vinacke's terms, "exploitative" and that females tend to be "accommodative" in coalition situations. There appeared to be no effects due to sex in this study: Males tended to obtain overall payoffs as veto players only marginally higher than females did (89.8 vs. 87.2); similarly, females obtained only marginally higher payoffs as nonveto players (2.6 vs. 2.3). However, the random assignment of players to slots resulted in a disproportionate number of male veto players in the 20(17–6–5–4–3) game. Thus it is difficult to make firm conclusions.[3]

Coalition Frequencies

Because each game resulted in a different set of minimum winning coalitions, the frequencies of the different coalition types were analyzed in four analyses of variance, one for each game. Thus, four Default (3) × Order (4) × Trial Block (3 blocks of 4 trials each) × Coalition Type (which differed over games, see Table 1) analyses of variance were run, with the frequencies of each coalition type converted to arcsin values[4] as the dependent variable in each case. Trials where nonminimum winning coalitions formed (not predicted by any of the models, and occurring only 2% of the time) or where defaults resulted were not included in these analyses. The data are shown in Table 3.

In the 28(17–8–7–6–5) game, significant effects were found for default conditions, $F(2, 12) = 3.92$, $p < .05$, and the default by coalition type interaction, $F(10, 60) = 2.44$, $p < .05$. The default main effect merely reflects the fact that due to a total of 15 defaults in this game, there were more coalitions forming in the no default condition than in the two default conditions. The Default × Coalition Type interaction appears to be primarily a result of an unusually high frequency of ABD coalitions in the no default condition and relatively equal frequencies for the other coalitions in each default condition.

In the 30(21–10–7–5–4) game, the only significant effect was for coalition type, $F(3,$

[3] Indeed, because there were no female veto players in the no default 20(17–6–5–4–3) condition, reliable statistical analyses were not possible.

[4] The transformation used was

$$x' = 2 \arcsin \sqrt{x},$$

where x is expressed as a proportion of the maximum possible frequency for any coalition within a trial block. Also, for $x = 0$, $1/(4n)$ (where $n = 4$) was used; for $x = 1$, $1 - 1/(4n)$ was used (Bartlett, 1947).

Table 3
Coalition Frequencies and the Predictions of the Weighted Probability Model (in Parentheses)

Game	Coalition										
	AB	AC	AD	AE	ABC	ABD	ABE	ACD	ACE	ADE	Default
20(17–6–5–4–3)	80 (72)	62 (72)	50 (72)	80 (72)	1	0	0	0	1	2	0
38(27–13–12–6–5)	88 (112)	111 (112)	—	—	1	0	0	0	1	81 (56)	6
30(21–10–7–5–4)	134 (112)	—	—	—	1	1	0	40 (56)	45 (56)	59 (56)	5
28(17–8–7–6–5)	—	—	—	—	41 (48)	48 (48)	45 (48)	29 (48)	51 (48)	57 (48)	15

Note. Totals do not always add to 288 because of the exclusion of infrequent coalitions such as ABDE, ACDE, etc. Coalitions that were not predicted to form have no parenthetical entry.

36) = 13.91, $p < .0001$. Post hoc tests indicated that the AB coalition ($n_{AB} = 134$) formed significantly more than either ACD, ACE, or ADE (ns = 40, 45, 59, respectively). In the 38(27–13–12–6–5) and 20(17–6–5–4–3) games, no significant effects were found.

These data show some support for the weighted probability model's frequency predictions (see Table 3). In the 28(17–8–7–6–5) and the 20(17–6–5–4–3) games, only one effect was significant for coalition types (a Default × Coalition Type interaction), primarily resulting from a high frequency for one coalition type in one particular condition. No main effects were found for coalition types; overall, then, the weighted probability model's predictions in these two games were supported. In the 30(21–10–7–5–4) game, the AB coalition was significantly more frequent than the three-person coalitions, as predicted. However, it was considerably more frequent than even the model predicted. Finally, in the 38(27–13–12–6–5) game, there were no significant differences among the coalitions, although the model predicts that AB and AC will each form twice as frequently as ADE. Thus the model receives only marginal support, in the two games where equal likelihood predictions are made.

The Veto Players' Payoffs

The veto players' payoffs were analyzed in two Default (3) × Game (4) × Trial (12) × Order (4) analyses of variance, one including the default payoffs, the other excluding them. Because the results of the two analyses were almost identical, the data excluding the default payoffs will only be presented when the outcomes of the analyses differed. In addition, because the distributions of player A's payoffs were negatively skewed, both in the entire distribution and within almost every cell of the design, medians and means for significant effects are reported. Although the analysis of variance is robust with respect to skewed distributions, reporting the medians may be a better representation of the central tendency within each cell.[5]

Significant main effects were found for games, $F(1, 4) = 19.83$, $p < .05$, for trials, $F(1, 12) = 9.93$, $p < .01$, and for order, $F(1, 12) = 6.85$, $p < .05$, and were marginally significant for the default conditions, $F(2, 12) = 3.12$, $p < .10$. Post hoc tests using the Neuman–Keuls procedure indicated that the veto players received significantly more in the 20(17–6–5–4–3) game than in the 38(27–13–12–6–5) or 30(21–10–7–5–4) games and received the lowest payoffs in the 28(17–8–7–6–5) game (see Table 4). The veto players' payoffs consistently increased over trials. Player A received significantly less in the first game played ($M = 84.4$, $Mdn = 90$), and no differences, again using the Newman–Keuls technique, were noted among the other orders ($91.1 \leq M$s ≤ 92.5, $95 \leq Mdn$s ≤ 97). The marginal default effect suggests that Player A received the highest payoffs in the no default condition. This effect disappeared, however, when the default trials were removed, $F(2, 12) = 2.17$, *ns*.

[5] Tests of homogeneity, using Huynh and Feldt's (1970) chi-square test, were also run for each of the significant effects. (The small sample size precluded an overall test.) None of the chi-squares reached significance at the .05 alpha level.

Table 4
Means and Medians of the Veto Players' Payoffs in Each Game Over the 12 Trials

Game	Trial												*M*	*Mdn*
	1	2	3	4	5	6	7	8	9	10	11	12		
20(17–6–5–4–3)	90.4	90.4	93.6	95.1	95.7	94.5	95.9	95.7	96.5	96.0	96.8	95.9	94.7	98
38(27–13–12–6–5)	81.4	87.4	87.6	88.4	91.1	92.1	92.8	94.0	93.6	94.4	94.6	92.2	90.8	94
30(21–10–7–5–4)	81.8	85.9	87.5	89.5	91.8	91.0	91.0	90.9	90.6	93.4	93.8	92.9	90.0	95
28(17–8–7–6–5)	78.7	79.7	84.2	77.0	77.5	86.9	84.4	92.6	95.5	99.4	90.5	91.4	83.9	90
M	83.1	85.9	88.2	87.5	89.0	91.1	91.0	90.8	91.5	93.1	93.9	93.1	—	—
Mdn	88	90	90	92	94	94	95	96	96	97	97	97	—	—

Nonveto Players' Payoffs

The analysis of the nonveto players' payoffs was similar to those for Player A but included an additional factor, the players' positions (i.e., B, C, D, or E). The effects that do not include the players factor reproduce the results for the veto players' payoffs and are not reported. The significant effects that did include the players factor were the players main effect, $F(1, 12) = 15.10$, $p < .01$, and the Games × Players interaction, $F(3, 9) = 3.98$, $p < .05$. Table 5 displays the means for the Games × Player interaction and the associated main effects. Post hoc tests indicate that Player B received significantly higher payoffs than the other players in the 30(21–10–7–5–4) game did and that Players B and C received significantly higher payoffs than Players D and E in the 38(27–13–12–6–5) game. In both cases, then, nonveto players who held positions that allowed for the formation of a two-person coalition with Player A received significantly higher payoffs than nonveto players did in the same game who were required to form three-person agreements.

In addition, combining the mean payoffs for players who were required to form three-person coalitions compares very favorably to the mean payoffs for players who could form two-person coalitions. For instance, in the 30(21–10–7–5–4) game, combinations of Players C and D's payoffs, C and E's payoffs, or D and E's payoffs range from 3.88 to 4.10; Player B's mean payoffs of 4.00 are almost identical. Thus, it appears that, across the coalition type, the payoffs of the veto player's partnership sets were fairly stable within a game. This result is predicted by several game theoretic models (e.g., Aumann & Maschler's, 1964, bargaining set).

One way to explore the determinants of the veto players' payoffs is to investigate the correlations between their payoffs and the offers they received, along with their own demands. It has been suggested that players who are

Table 5
Mean Payoffs for the Nonveto Players in Each Game, Including Default Payoffs

Game	Player				*M*
	B	C	D	E	
20(17–6–5–4–3)	1.76_{cd}	$.98_{d}$	1.07_{d}	1.49_{cd}	1.32_{c}
38(17–13–12–6–5)	2.87_{abc}	2.88_{abc}	1.42_{d}	1.33_{d}	2.13_{b}
30(21–10–7–5–4)	4.00_{a}	1.88_{cd}	2.00_{cd}	2.10_{bcd}	2.50_{b}
28(17–8–7–6–5)	3.82_{a}	3.14_{ab}	3.15_{ab}	3.58_{a}	3.42_{a}
M	3.11_{a}	2.22_{b}	1.91_{b}	2.12_{b}	

Note. Means with common subscripts within a main effect or the interaction are *not* significantly different from one another at the .05 level using the Newman–Keuls procedure.

given veto power can obtain almost any payoff they desire, at least when there are no default conditions (Murnighan & Roth, 1978). In this study, Player A's demands (his or her offers to self) were more highly correlated (significantly) with A's payoffs than were either the mean of the nonveto players' offers or the highest of the nonveto players' offers. Player A's payoffs and demands showed a correlation of .94 ($n = 1126$) when default payoffs were excluded and correlations ranging from .88 to .98 in the four games. The highest of the nonveto players' offers revealed an overall correlation with A's payoffs of .76, with correlations ranging from .66 to .83 in each of the games. Needless to say, each of these correlations is significant. More important, in each game the correlations between A's payoffs and A's demands were significantly higher (using two-tailed t tests, Edwards, 1967, pp. 250–251) than the corresponding correlations between A's payoffs and the nonveto players' offers. It appears, then, that veto players controlled their payoffs more than the nonveto players, replicating previous findings (Murnighan & Roth, 1978), even when defaults were possible.

Defaults

The frequency of defaults in the game/default conditions is shown in Table 6. Given the total of only 26 defaults, reliable statistical analyses cannot be used on these data. In addition to the data in the table, which suggest that the 60–10 default condition and the 28(17–8–7–6–5) games led to more frequent defaults, defaults were more frequent early in each session, in the first set of four trials ($n = 15$), rather than later, in the second (6) or third set (5).

The frequency of trials that reached at least three attempts is also shown in Table 6. The data suggest that reaching the third attempt was most likely to lead to a default in the 28(17–8–7–6–5) and 38(27–13–12–6–5) games in the 60–10 default condition.

The uneven distribution of trials that reached three attempts necessitated a least squares analysis of variance of the offers to the veto player over the three attempt trials. Using default conditions (3), games (4), players (5), and attempts (first, second, and third) as independent variables, the results yielded a suggestive interaction between default conditions and attempts, $F(1, 50) = 3.84$, $p < .07$. The means in the interaction suggest relatively stable, slightly increasing offers to A over the three attempts in the no default and zero default condition and decreasing offers to A in the 60–10 condition. Even in this condition, however, the mean offers to A dropped only slightly, from 83.3 to 80.6.

Discussion

The results clearly do not support the weighted probability and Roth–Shapley models. Just as they did in the Michener et al. (1976) study, the veto players' payoffs far exceeded the predictions of either model. Although the weighted probability model received some support from the data on coalition frequencies, it consistently underestimated the veto players' payoffs. The Roth–Shapley model is similarly inadequate, even though its predictions are somewhat closer to

Table 6
Frequency of Defaults in Each Game/Default Condition

Default	Game 28	30	38	20	Total
None	– (21)	– (10)	– (2)	– (4)	– (37)
Zero	7 (13)	0 (13)	2 (6)	0 (2)	9 (34)
60–10	8 (15)	5 (14)	4 (8)	0 (4)	17 (41)
Total	15 (49)	5 (37)	6 (16)	0 (10)	26 (112)

Note. The frequency of trials that reached three attempts is shown in parentheses.

the data. The models received some very marginal support from the players' payoffs for the significant differences that were found for nonveto players who could form two-player coalitions in the 38(27–13–12–6–5) and the 30(21–10–7–5–4) games: Even though these players' payoffs were smaller than predicted, they were significantly larger than those of players in the same games who were required to form at least three-person coalitions with Player A, as predicted. Also, in rank ordering the players' payoffs, both the models did fairly well (rank correlation coefficients between observed and predicted payoffs of .93), suggesting a correct depiction of the games' structures, if not the specific outcomes.

The core, on the other hand, was clearly not reached. However, the payoffs of the veto players consistently moved toward the core in each of the four games studied. The core is also not supported by the significant difference found among the four games. In addition, the payoffs for the veto players in the 28(17–8–7–6–5) game were considerably below those in the core. Even in this game, however, the payoffs consistently increased over trials. In addition, the consistently strong correlations, across games and the other conditions, between the veto players' demands and payoffs strongly suggests that they had control of their payoffs. Just as Murnighan and Roth (1978) found that when the veto players demanded 99% or more of the payoff, over 90% of the time they were receiving it, so too the present data suggest near-complete control for the veto player. Whether veto players could push their payoffs closer to the entire payoff if they were tougher in their demands remains to be tested in future research. This study, along with other recent research (e.g., Fiorina & Plott, 1978; Murnighan & Roth, 1977, 1978), shows promise for the predictive capabilities of the core.

The findings, then, clearly do not offer strong support for the weighted probability model, the Roth–Shapley value, or the core. Each received marginal support, however: the core by the increasing payoffs received by the veto players and the other models by their ordering of the magnitudes of the players' payoffs across games. The effects for games and for trials stand out and suggest that a combination of the properties of the two kinds of models may be necessary to predict the outcomes of coalition bargaining games that include a veto player. Komorita and Chertkoff's (1973) bargaining theory, although not testable in veto games, suggest that payoffs will change over trials, increasing for the most powerful player, as they have here. The difficulty in directly adapting such an insight is in determining where a veto player's payoffs will begin and where they will end. In the present study, the final median payoffs of the veto player were adequately, but not perfectly, represented by the core. The starting points were quite close to values midway between the core and either the weighted probability or Roth–Shapley models (i.e., 90, 85, 81–82, and 75–80 in the 20, 38, 30, and 20 games, respectively). A model that incorporates trials as a factor, with a veto player's payoffs increasing from such a core-weighted probability – Roth-Shapley midpoint on the early trials to the core on later trials, dependent on the game being played, would explain this study's data fairly well. Difficulties arise, however, in conceptually justifying such predictions in a logical way. In addition, the different conditions that surround the bargaining situation, including, for example, the physical separation of individuals, as was the case in this study, may also determine the magnitude of the payoffs obtained by the bargainers. Given the fact that few studies have been conducted on the dynamics of games that include veto players, further research, even of an exploratory nature, might contribute sufficient information for the construction of theoretical propositions that are viable both conceptually and empirically. Given the current lack of clear-cut results, it appears that not enough is understood about veto games to present such a theory.

The effects for the other variables studied here, order and the default conditions, were somewhat unexpected. The players were given considerable preparation before the game; nevertheless, most expressed the belief that the first session was needed to establish the firsthand understanding necessary to bargain effectively in the games. Indeed, players in

the A position tended to increase their payoffs, especially in the 28(17–8–7–6–5) game, after their first game. These results suggest that experience may be a key factor in coalition research, if not in "real world" situations. Whereas most research on coalitions asks participants to play in only a single session, this study highlights the effects of experience, suggesting that stability is only achieved after a rather involved first experience.

The lack of effects found for the default conditions was particularly surprising. It would seem, at least, that the default conditions should have constrained Player A's ability to obtain such high payoffs. The fact that one player has veto power in each of the games appears to have overwhelmed any effects of this manipulation. Combined with the results from Murnighan and Roth (1978), the conclusion that veto players can and will ruthlessly pursue their own rewards is difficult to refute. Indeed, these results support the widely held assumption that strict controls are necessary to restrain monopolistic organizations.

There are occasions, however, when more socially beneficial behaviors result. In this regard, it is important to describe one "exceptional" group that played the 28(17–8–7–6–5) game first, in the no default condition. At the start of the game, the veto player made a proposal of 60–20–20 (60 points for himself, and 20 for two of the other players). After a few rejections, this agreement was finally achieved. Even though Player A continuously received offers exceeding 60 points, he *enforced* 60–20–20 payoffs on every trial, rotating among the nonveto players to insure equal outcomes for each of them. By the end of the session, all of the players were exchanging 60–20–20 offers. This group's first session was interesting—their second was remarkable! The new veto player, now playing the 38(27–13–12–6–5) game, demanded 98 of the 100 points on his first offer. He received four offers of 60–20–20! Inflexibility of both A and the nonveto players resulted in 32 unsuccessful attempts prior to the formation of the first coalition. And the payoff finally agreed upon in that coalition was 60–20–20. It appears that a norm had developed: Payoffs of 60–20–20 were defined as appropriate; others were not. Although this norm gradually eroded (due mainly to the impatience of players who were excluded from the winning coalition), the veto players in this group received payoffs that were considerably below the mean over all veto players, in each of the four games. The nonveto players, on the other hand, were above the mean for their position for each of the games. According to the grading system announced prior to play, scoring above the mean for all players in the same position in three out of four games (or above the overall mean) would insure an "A" for the project. Thus, all five players in this group received an "A"!

Clearly this was an unusual group. However, one explanation for the group's success is that the first veto player's strategy provided this individual with a chance to obtain a high grade for himself *and* contribute to a high grade for the other players in the group. Unlike prisoners' dilemma games, then, where individual and group gains are generally in conflict, coalition games like these offer players at least one strategy that can be implemented to maximize both individual and group outcomes.

Conclusion

In summary, this study investigated four coalition games that included a veto player. The games varied the number of two-person coalitions that the veto player could form to obtain a fixed payoff in three default conditions. The games had a significant effect on the veto players' payoffs, with payoffs diminishing as fewer two-person coalitions were possible. The other noteworthy effects included support for the core, particularly because of the consistently increasing payoffs for the veto player. The default conditions had little impact on the results. Indeed, the lack of significant effects for this variable suggests that players with veto power can control the game they are playing almost at will. These results further illustrate the need for controls of some kind for monopolistic organizations. Future research might also investigate the effects noted in the one "exceptional" group: When

joint gains can be tied to higher individual gains (or at least potentially higher individual gains), then (and possibly only then) cooperative, group-oriented behaviors may emerge (cf. Argyris, 1957). Indeed, the incidence of cooperative behavior in highly competitive situations such as those in this study has rarely been investigated (an exception is Dawes, McTavish, & Shaklee, 1977) and would help to move research on coalitions into the determination of structural mechanisms to increase cooperation.

Reference Notes

1. Murnighan, J. K. *Experience and coalition behavior: The question of generalizability*. Unpublished manuscript, University of Illinois, 1979.
2. Komorita, S. S., and Meek, D. *Some preliminary tests of a bargaining theory of coalition formation*. Unpublished manuscript, Indiana University, 1972.

References

Argyris, C. *Personality and organization*. New York: Harper, 1957.

Aumann, R. J. Markets with a continuum of traders, *Econometrica*, 1964, *32*, 39–50.

Aumann, R. J., & Maschler, M. The bargaining set for cooperative games. In M. Drescher, L. S. Shapley, & A. W. Tucker (Eds.), *Advances in game theory*. Princeton, N.J.: Princeton University Press, 1964.

Bartlett, A. The use of transformations. *Biometrics*, 1947, *3*, 39–52.

Bartos, O. Determinants and consequences of toughness. In P. Swingle (Ed.), *The structure of conflict*. New York: Academic Press, 1970.

Box, G. E. P. Some theorems on quadratic forms applied in the study of analysis of variance problems. *Annals of Mathematical Statistics*, 1954, *25*, 290–302, 484–498.

Dawes, R. M., McTavish, J., & Shaklee, H. Behavior, communication, and assumptions about other people's behavior in a commons dilemma situation. *Journal of Personality and Social Psychology*, 1977, *35*, 1–11.

Debreu, G., & Scarf, H. A limit theorem on the core of an economy. *International Economic Review*, 1963, *4*, 235–246.

Edwards, A. L. *Statistical methods*. New York: Holt, Rinehart & Winston, 1967.

Fiorina, M. P., & Plott, C. R. Committee decisions under majority rule: An experimental study. *American Political Science Review*, 1978, *72*, 575–598.

Gamson, W. A. A theory of coalition formation. *American Sociological Review*, 1961, *26*, 373–382.

Huynh, H., & Feldt, L. S. Conditions under which mean square ratios in repeated measurements designs have exact *F*-distributions. *Journal of the American Statistical Association*, 1970, *65*, 1582–1589.

Komorita, S. S. A weighted probability model of coalition formation. *Psychological Review*, 1974, *81*, 242–256.

Komorita, S. S., and Chertkoff, J. M. A bargaining theory of coalition formation. *Psychological Review*, 1973, *80*, 149–162.

Komorita, S. S., & Moore, D. Theories and processes of coalition formation. *Journal of Personality and Social Psychology*, 1976, *33*, 371–381.

Luce, R. D., & Raiffa, H. *Games and decisions*. New York: Wiley, 1957.

Michener, H. A., Fleishman, J. A., & Vaske, J. J. A test of the bargaining theory of coalition formation in four-person groups. *Journal of Personality and Social Psychology*, 1976, *34*, 1114–1126.

Murnighan, J. K. Models of coalition formation: Game theoretic, social psychological, and political perspectives. *Psychological Bulletin*, 1978, *85*, 1130–1153. (a)

Murnighan, J. K. Strength and weakness in four coalition situations. *Behavioral Science*, 1978, *23*, 195–208. (b)

Murnighan, J. K., & Roth, A. E. The effects of communication and information availability in an experimental study of a three-person game. *Management Science*, 1977, *23*, 1336–1348.

Murnighan, J. K., & Roth, A. E. Large group bargaining in a characteristic function game. *Journal of Conflict Resolution*, 1978, *22*, 299–317.

Roth, A. E. The Shapley value as a von Neumann-Morgenstern utility. *Econometrica*, 1977, *45*, 657–664. (a)

Roth, A. E. Bargaining ability, the utility of playing a game, and models of coalition formation. *Journal of Mathematical Psychology*, 1977, *16*, 153–160. (b)

Rubin, J. Z., & Brown, B. R. *The social psychology of bargaining and negotiation*. New York: Academic Press, 1975.

Shapley, L. S. A value for *n*-person games. *Annals of Mathematical Studies*, 1953, *24*, 307–317.

Simon, H. A. *Administrative behavior* (4th ed.). New York: Free Press. 1976.

Thibaut, J. W., & Kelley, H. H. *The social psychology of small groups*. New York: Wiley, 1959.

Vinacke, W. E. Sex roles in three-person games. *Sociometry*, 1959, *22*, 343–360.

Winer, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1962.

Yukl, G. A. The effects of situational variables and opponent concessions on a bargainer's perceptions, aspirations, and concessions. *Journal of Personality and Social Psychology*, 1974, *29*, 227–236.

Received July 14, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 11, 1947–1956

Weather, Mood, and Helping Behavior: Quasi Experiments With the Sunshine Samaritan

Michael R. Cunningham
Elmhurst College

To establish the relationship of weather variables with helping behavior, two quasi-experimental field studies were conducted. In the first study, executed in the spring and summer and subsequently replicated in the winter, the amount of sunshine reaching the earth was found to be a strong predictor of a participant's willingness to assist an interviewer. Smaller relationships were also found between helping and temperature, humidity, wind velocity, and lunar phase. A second study was conducted indoors to control for comfort factors, and sunshine, lunar phase, and participant's age and sex were found to predict the generosity of the tip left for a restaurant waitress. Sunshine and temperature temperature were also significantly related to self-reports of mood.

So many things I would have done, but clouds got in my way. . . . (Joni Mitchell, "Both Sides Now")

One of the most pervasive background environmental variables in human life, a factor that shapes agricultural economies, makes or breaks recreational plans, and serves as a perennial topic for superficial conversation, is the weather. Yet although the weather appears to affect both emotion and social behavior (Campbell & Beets, 1977; for reviews, see Huntington, 1945; Larson, 1965; Moos, 1976; Tromp, 1974; Winslow & Herrington, 1949), the specific weather variables most closely associated with behavior changes remain obscure.

A brief scan of some of the research literature will illustrate the wide variety of meteorological indices purportedly linked to various human behaviors.

Dexter (1904), for example, found very low barometric pressure, excessive humidity, and abnormal winds to be related to poor student deportment and increased teacher use of corporal punishment. Deteriorating or stormy weather has also been related to reduced task performance by workers (Muecher & Ungeheuer, 1961). Weather has also been implicated in psychological disturbances; Mills (1934) and Digon and Bock (1966) found a relationship between low barometric pressure, which usually characterizes stormy days, and suicide, whereas Lester (1970) obtained a relationship between the amount of snow during the winter months and the frequency of suicide.

The sky conditions may also affect mood or psychological well-being. Winslow and Herrington (1936) found a correlation of .78 between the amount of sunshine apparent on a given day and judgments of atmospheric "pleasantness" (associations were also found with increased barometric pressure and decreased humidity). Finally, Persinger (1975) observed that low moods in his 10 subjects were associated with fewer hours of sunshine and high humidity.

Although the foregoing studies suggested that weather may influence emotion and be-

Thanks to Thomas Brothen, Charles Carver, Carolyn Cunningham, William Graziano, Alice Isen, Michael Louguee, Paul Rosenblatt, Catherine Thayer, Sandra Titus, and Hans Wendt for their insightful comments on an earlier draft of this manuscript. Grateful appreciation also to Paul Cronstrom, Mike Fleming, Ruth Nachman, Linda Rudin, and Cindy Weber for their help as experimenters, and to the staff of the Minneapolis–St. Paul station of the National Weather Service and the manager and waitresses of Henrici's Restaurant for their invaluable assistance.

Requests for reprints should be sent to the author, Department of Psychology, Elmhurst College, Elmhurst, Illinois 60126.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3711-1947$00.75

havior, straightforward conclusions are difficult. In no study have all of the implicated weather variables been considered simultaneously to determine the relative strength of their relationships with behavior. In addition, a number of the studies suffered from the methodological weakness of using indices based on the average level of a given weather variable, and the frequency of a certain behavior within relatively wide time blocks, such as a day, week, or month, to establish a relationship (e.g., Banziger & Owens, 1978). But since a day or week in which temperatures varied from 30 °F to 70 °F (−1 °C to 21 °C) produces a different experience from a stable 50 °F (9 °C) day, such summary statistics may be highly misleading. Further, if weather and behavior are not sampled concurrently, it is difficult to eliminate the argument that purported weather effects are due to some third variable or by-product of the weather (e.g., mobility restriction).

The use of experimentally controlled environments would eliminate these objections. Griffitt and Veitch (1971) found that under conditions of manipulated high temperature, participants' evaluation of both other people and the experimental environment was significantly more negative than under conditions of comfortable temperature. Corresponding changes in the participant self-ratings of affect suggest that the temperature effects probably were mediated by alterations in the individual's mood. Subsequent work has indicated that there may be an upper limit to the temperature–aggression relationship; too much heat may actually inhibit aggression (Baron & Bell, 1976).

Yet although temperature can be easily manipulated, controlling humidity, barometric pressure, wind velocity, and amount of sunshine in a factorial design seems unfeasible. A compromise would be to adopt a quasi-experimental design (Campbell & Stanley, 1963) in which uncontrolled variations in the weather are systematically observed and related to changes in a carefully measured dependent variable using randomly selected subjects.

Helping or altruism is a social behavior that has been found to vary as a function of the mood or emotional state of the subject (Cunningham, Steinberg, & Grev, in press). If the weather affects individuals' emotional state or sense of well-being, then it might be expected to affect helping, and two studies tentatively suggested such a relationship. Lockard, McDonald, Clifford, and Martinez (1976) reported that panhandlers were more successful in spring than in autumn, and Cialdini, Vincent, Lewis, Catalan, Wheeler, and Darby (1975) footnoted a similar observation. Neither study, however, connected a specific weather variable to an increase in helping.

The present series of studies was designed to assess the impact of a variety of weather, climatic, and air quality indicators on helping, using two different indices. The first study, concerned with helping an interviewer, was initially conducted outdoors during the late spring and summer and was subsequently replicated during the winter. The second study involved restaurant tipping, and data were collected indoors during the early spring.

Predictions for the effects of weather on both forms of helping were drawn from the results of the earlier research. Since behavior seemed to be adversely affected by low barometric pressure and high humidity, negative correlations were predicted for those variables with interpersonal helping. Although previous research has found only that mood was adversely affected by high temperature, it is reasonable to imagine that mood may be depressed by the cold as well. An inverted U-shaped relationship was thus predicted for temperature and helping, with the most helping expected at moderate temperatures. Finally, since mood seemed to be affected by sunlight, a positive correlation was predicted between helping and the amount of sunshine reaching the earth due to the absence of cloud cover.

Experiment 1

Method

Participants. Data were gathered outdoors from 540 participants at two off-campus locations in the city of Minneapolis and two locations on the main campus of the University of Minnesota. Every 3rd individual over the apparent age of 16 passing the experimenter was approached, and the estimated age

and sex of the subject were recorded. Fifteen participants were approached each day during the afternoon on 36 weekdays randomly selected during the spring and summer of 1974 and the winter and early spring of 1974–1975. An average of 4 people each day refused to stop for the study and were not included in the primary analyses. Temperatures ranged from −18° C to 38° C. Data were not collected during any periods of precipitation.

Procedure. One male and one female were employed as experimenters each season. One female was used in winter and another in summer. One experimenter was blind to the purposes of the experiment each season. The measure of helpfulness was obtained by approaching passersby with the statement:

> Hi. I'm from the sociology department of the University of Minnesota and we're conducting a survey of social opinions. Although the survey is 80 questions long, you don't have to answer all the questions. How many questions would you be willing to answer for me?

The number of questions the participant was willing to answer was employed as an interval measure of helping. Participants were then debriefed following their response.

Weather readings taken at the beginning of the specific hour in which the participants were approached were employed as predictor variables. Separate estimates were obtained of the amount of sunlight reaching the earth, atmospheric temperature, barometric pressure, relative humidity, wind velocity, and lunar phase. To determine if changing weather had an effect, the increase or decrease in temperature and barometric pressure over the last hour and last 3 hours prior to the behavioral measurement was calculated. Three measures of air pollution were also included: carbon monoxide, sulfur dioxide levels, and Apex, a weighted composite of CO, SO_2, and particulate levels. Weather readings were obtained from the National Weather Service, Minneapolis–St. Paul Forecast Office, with the exceptions of sunshine level, which was obtained by incident light reading using a Gossen Luna-Pro light meter, and lunar phase, obtained from an almanac.

Results

Preliminary analysis revealed no association between the age and sex of the subjects or location and the weather variables. This independence, and the well-trafficked public places used for sampling, supported the assumption of random assignment of subject to weather conditions and the employment of each subject as an independent case. There was also no association between experimenters and helping.

A general regression equation was computed for the summer, the winter, and the combined data, using the 13 weather variables as predictors of helping. Each of the three regression equations—for the summer, $F(13, 256) = 7.41$, $p < .001$; winter, $F(13, 256) = 7.47$, $p < .001$; and combined sample, $F(13, 526) = 8.99$, $p < .001$—was highly significant and, analogous to a significant overall F test in analysis of variance (ANOVA), justified examination of the individual correlations. The Pearson correlations between the weather variables and helping are presented in Table 1.

The most significant variable in this study, across both summer and winter, was sunshine. People were likely to be more helpful when the sky was clear and a large amount of sunshine was striking the earth, compared to when the sky was cloudy and less sunlight reached ground level. The positive association of sunshine with helping can apparently not be attributed to temperature, since the correlation between the two variables was moderate, $r(538) = .29$, $p < .001$, and in the partial correlation analysis described below, both variables contributed independently to the prediction of helping.

Temperature was also a significant predictor of helping, although the nature of the relationship varied from summer to winter. Consistent with predictions, temperature was negatively associated with helping in the summer and positively related to helping in the winter. To determine if the relationship of temperature with helping was nonlinear, a curvilinear equation was constructed. Temperature scores were divided into seven equal intervals and were weighted to produce a curve with an apex centering about 19 °C (65 °F), consistent with the experimentally obtained optimal outdoor temperature for clothed active individuals (Yaglow & Miller, 1925). The correlation of the curvilinear equation with helping was positive, $r(538) = .30$, $p < .001$, and was significantly better than the linear relationship of temperature with helping, $z = 2.11$, $p < .02$.

Further, both sunshine and the curvilinear temperature index were employed in a general regression equation to predict helping. Both sunshine and curvilinear temperature were significant, $F(2, 537) = 55.82$ and 31.11, re-

Table 1
Correlations of Weather and Subject Variables With Amount of Help Offered an Interviewer

Item	Summer	Winter	Combined
Sunshine	.32***	.40***	.36***
Temperature	−.16*	.37***	.11***
Temperature increase (1 hr.)	−.19**	−.09	−.13***
Temperature increase (3 hrs.)	−.05	−.01	−.03
Barometric pressure	.03	.05	.04
Barometric pressure increase (1 hr.)	.21***	−.11*	.07
Barometric pressure increase (3 hrs.)	.23***	−.02	.09
Relative humidity	−.19**	−.22***	−.20***
Wind velocity	.20**	−.15*	.01
Air pollution index (Apex)	.11*	−.12*	−.02
Sulfur dioxide	.02	−.26***	−.15***
Carbon monoxide	.12*	−.16*	−.03
Lunar phase[a]	−.15*	−.16*	−.15***
Age	−.10*	.01	.04
Sex[b]	.03	.16*	.10

Note. Summer and winter *n*s = 270.
[a] Large value = full moon. [b] Large value = female.
* $p < .05$. ** $p < .01$. *** $p < .005$.

spectively, $p < .001$, with sunshine accounting for 13% and curvilinear temperature accounting for 5% of the total variance in helping.

Another predicted significant association found in this study was a negative correlation between relative humidity and helping. People were more likely to be helpful when the humidity was low than when the humidity was high.

As a converse to temperature, wind velocity was positively correlated with helping in the summer but negatively correlated with helping in the winter. This is reasonable, since a cooling breeze provides a desirable relief from summer heat but is an undesirable contributor to wind chill in the winter months.

Unexpectedly, a small negative relationship was found between lunar phase and helping. In both summer and winter, people seemed to be less helpful when the moon was full than when the moon was less full. Further analyses reported below, however, raise questions concerning this relationship.

A number of other correlations of weather variables with helping appear in Table 1 for one season but not for the other. Each of these variables correlated more highly with another weather variable than they did with helping, and they must be regarded as unreliable until they are replicated.[1]

Partial correlation analyses were conducted to determine the relative independence and strength of the significant weather variables including sunshine, temperature, relative humidity, wind velocity, and lunar phase. The results of these analyses for the summer, winter, and combined sample are presented in Table 2.

The partial correlation analysis revealed that sunshine, temperature, relative humidity, and wind velocity were each significantly related to helping in the summer, in the same directions as in the zero-order correlations. A regression equation using these four predic-

[1] A number of derived statistics involving combinations of various weather variables were tried as predictors of helping. A discomfort index proposed by the U.S. National Weather Service (Mather, 1974, p. 245), which consisted of a weighted multiplicative equation involving temperature and humidity, proved to be negatively related to helping but with a correlation smaller than curvilinear temperature or humidity separately, $r(538) = -.13$, $p < .05$. A wind chill statistic (Mather, 1974, p. 246) that attempted to combine both temperature and wind velocity proved unrelated to helping, $r(538) = .03$, *ns*, perhaps because of the curvilinear relation of temperature to helping. Finally, a statistic proposed by Dirnagl (1977) that combined temperature, humidity, and barometric pressure showed only a modest association with helping, $r(538) = .14$, $p < .01$.

tors was highly significant, multiple $R = .44$, $F(4, 265) = 16.07$, $p < .001$, and accounted for 20% of the variance in helping, with sunshine accounting for 10% of the total variance. In the winter data, sunshine, temperature, and wind velocity were significantly related to helping, whereas humidity, although in the right direction, was not significant. A regression equation with these three significant weather variables was again significant, multiple $R = .49$, $F(3, 266) = 27.44$, $p < .001$, and accounted for 24% of the variance, with sunshine accounting for 16% of the total variance. Lunar phase was significant in neither the summer nor the winter. Further, because of the curvilinear relationship of temperature and interactive relationship of wind velocity with temperature and helping, the partial correlation analysis on the combined sample revealed only sunshine to be significantly associated with helping.

Table 2
Partial Correlations of Selected Weather Variables With Helping

Item	Summer	Winter	Combined
Sunshine	.20***	.28***	.29***
Temperature	−.26***	.26***	−.002
Relative humidity	−.19***	−.07	−.05
Wind velocity	.14**	−.12*	.02
Lunar phase[a]	−.02	.03	−.05

Note. Summer and winter *n*s = 270.
[a] Large value = full moon.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Discussion

This study found significant associations between helping behavior and a number of weather variables. In both zero-order and partial correlation analysis, significant associations were found between sunshine and helping in both seasons, such that helping was greater on bright sunny days compared to cloudy days. In both sets of analyses, helping was greater during periods of cooler temperature and higher wind velocity in the summer, and warmer temperature and lower wind velocity in the winter. Helping was also greater when the relative humidity was lower, although in the partial correlation analyses this association was significant only for the summer data.

The association of lunar phase with helping was ambiguous. While lunar phase was negatively correlated with helping in the zero-order correlations, no association of lunar phase and helping was found in the partial correlations. Since lunar phase was negatively associated with other predictors of helping in this sample—e.g., sunshine and lunar phase, $r(538) = -.29$, $p < .001$—the original correlations may have been spurious.

The association with helping of some of the weather variables, such as wind velocity, temperature, and humidity, might be attributable to the comfort or discomfort that the individuals anticipate experiencing while helping. Being asked to answer some questions when the temperature is low, for example, entails spending additional time in the cold, and that added cost factor could reduce helping.

The comfort interpretation seems less effective as an explanation of the sunshine relationship, however. During Minnesota winters the temperature is often cold while the sky is bright and clear. Considering just those cases where the temperature was below freezing (0 °C) and physical comfort was presumably low, sunshine was still associated with a greater likelihood of helping, $r(177) = .23$, $p < .01$.

The role of comfort in the sunshine–helping relationship might also be examined using the data on the rate at which individuals refused to stop and listen to the interviewer solicit participation. Although this is an ambiguous test of helping because people did not hear all of the request, refusals at least were not based on the length of time outdoors required for helping. Nevertheless, fewer people refused to stop, and thus more were helpful, on sunny days than on cloudy days, $r(122) = -.32$, $p < .01$.

A more direct method for investigating the role of comfort in the association of sunshine and helping would be to use a setting in which comfort factors such as temperature, humidity, and wind velocity could be controlled, whereas sunlight would be free to vary. A dependent variable that does not require a time

Table 3
Correlations of Age and Sex With Tipping and Weather

Item	Age *r*	Sex *r*[a]
Tip	.12*	.29**
Sunshine	.25***	.24***
Temperature	−.54***	−.50***
Barometric pressure	.24***	.12*
Relative humidity	.20**	.29***
Wind velocity	.42***	.40***
Lunar phase[b]	.02	.04

Note. $N = 130$.
[a] Larger value = female. [b] Larger value = full moon.
* $p < .09$. ** $p < .01$. *** $p < .002$.

commitment would avoid the problem of the linkage between helping and remaining in a positive or negative environment. A follow-up study was thus conducted that examined the effects of weather on the tips left by patrons of an indoor eating establishment.

Experiment 2

Method

Data were gathered from 130 parties dining during the afternoon at a moderately expensive climate-controlled medium-sized restaurant at a shopping center in a western suburb of Chicago. The study was conducted during the afternoon on 13 randomly selected days during the work week in April, May, and June of 1978, and 10 parties were observed each day. The restaurant had windows on two sides, and the temperature ranged from 4° C (40° F) to 27° C (81° F) outdoors and was at 21° C (70° F) indoors.

Six waitresses served to record data, and all were blind to the purposes of the study. On each of the designated days, one waitress was asked to record information about the first 10 parties she waited on after 1:00 p.m. The waitress recorded the number of people in the group, the total amount of the check, the amount of the gratuity, the approximate age and sex of the person or persons leaving the tip, and whether liquor was served. Prior to recording information on the customers, the waitress was asked to record her own mood on a 5-point scale ranging from *good mood* to *bad mood*. Waitresses were carefully instructed in the use of the self-report scale and were asked to be honest, since in a previous study, randomly selected participants were found to be less than adept in using introspective self-report scales (Cunningham, Steinberg, & Grev, in press). For this reason and to avoid becoming intrusive, the moods of the tippers themselves were not assessed.

Weather information was obtained from the National Weather Service, Chicago Forecast Office, and included temperature, barometric pressure, and wind velocity. Sunshine level and lunar phase were obtained as in Study 1.

Table 4
Correlations of Weather and Group Characteristics With Percentage of Check Left As Tip

Item	Zero order correlation *r*	Age & sex partialed out *r*
Sunshine	.20***	.15**
Temperature	−.14**	−.02
Barometric pressure	−.05	−.08
Relative humidity	.19**	.10*
Wind velocity	.03	−.08
Lunar phase[a]	.14**	.15**
Group size	−.06	—
Liquor served	−.003	—

Note. $N = 130$.
[a] Large value = full moon.
* $p < .10$. ** $p < .05$. *** $p < .001$.

Results

Since it is customary to leave as a gratuity an amount of money proportional to the size of the check, the principal dependent variable analyzed in this study was the ratio of the amount of money left as a tip to the total bill for the meal. Prior to examining the effect of weather on tipping, the relationship of weather to the status variables age and sex was examined to insure independent sampling.[2] It was found, however, that there was an association between weather and the age and sex of people in the restaurant on a given day. As Table 3 indicates, older people and women were more likely to dine out when the sunshine was bright, the weather was cool, and the humidity, barometric pressure, and wind velocity were high. Although these associations are interesting in their own right, since older people and women also left a more substantial tip than the others, the analysis of the impact of weather on tipping neces-

[2] To insure that variations in tipping were not due to the behavior of individual waitresses, an analysis of variance was conducted, using the six waitresses as independent variables. No effect due to waitresses was found, $F(5, 124) = .025$, *ns*.

arily must take into account this self-selec-ion factor.

Although a regression equation using just he seven weather variables was significant, $F(6, 123) = 4.24$, $p < .01$, two sets of corre-ations are presented in Table 4, the zero-rder correlation between each weather varia-le and tipping and second-order partial orrelations, with the association of age and ex removed from the correlations of weather nd tipping. As Table 4 reveals, despite the act that sunshine was related to an increase n the number of older people and females lining, sunshine was nevertheless significantly elated to helping. This correlation was maller than in Experiment 1, which might e expected, since participants were indoors nd were thus partially screened from effects f the sun.

Although inspection of the zero-order corre-ations suggested some reversals of the find-ngs of Experiment 1 with respect to tempera-ure and humidity, the partial correlations llayed such concerns. As might be expected n a climate-controlled restaurant, outdoor emperature, barometric pressure, and wind elocity had no significant direct effect on ipping. Group size and whether liquor was erved were unrelated to tipping, whereas hu-nidity was marginally correlated. Apparently ontrary to Experiment 1, there was a positive ssociation between the fullness of the moon nd the size of the tip.

Additional partial correlation analyses were erformed to determine the relationship of ach weather variable to tipping, with the ffects of the other weather variables and age nd sex removed. These analyses are pre-ented in Table 5.

Once again, sunshine was significantly cor-elated with tipping. Relative humidity was lso positively correlated with tipping, whereas lunar phase showed a marginal posi-ive association. A regression analysis using unshine, temperature, relative humidity, and unar phase as predictors of tipping produced ignificant associations of sunshine and rela-ive humidity with tipping, $F(4, 125) = 4.73$, $p < .01$, with sunshine accounting for 4% of he total variance.

Table 5
Partial Correlations of Selected Weather Variables With Tipping

Item	Partial *r*
Sunshine	.23***
Temperature	.03
Relative humidity	.17**
Lunar phase[a]	.11*

Note. $N = 130$. The effects of age and sex have been partialed out as well as the other weather variables.
[a] Large value = full moon.
* $p < .10$. ** $p < .03$. *** $p < .005$.

To test for a curvilinear relation of temperature with tipping, a weighted equation centered on 19 °C (65 °F) was constructed. Curvilinear temperature was positively related to tipping, $r(128) = .19$, $p < .02$, as a zero-order correlation but was insignificant when the effects of age and sex were removed, $r(128) = .09$.

Examination of the relationship of the weather variables to the self-reported mood of the waitresses, presented in Table 6, provides further insight into the nature of the weather effects. As Table 6 reveals, both sunshine and temperature were significantly related to a more positive mood of the waitress. Waitress mood (obtained prior to the receipt of tips) was not itself a significant predictor of tipping, $r(128) = .004$, suggesting that the effect of weather on tipping may have been mediated by the mental or emotional state of the customers.

Table 6
Correlations of Weather With Waitress's Mood

Item	*r*
Sunshine	.60*
Temperature	.75**
Barometric pressure	.32
Relative humidity	−.36
Wind velocity	−.20
Lunar phase[a]	−.35

Note. Larger number is more positive mood. $N = 13$.
[a] Large value = full moon.
* $p < .05$. ** $p < .01$.

Discussion

The second study effectively controlled for the effect on helping of comfort factors such as temperature and wind velocity by investigating the effect of weather on a helping action that did not require the participants to remain outdoors for a period of time in order to be helpful. Nonetheless, this study replicated Experiment 1 by finding that the outdoor sunshine level was significantly related both to the gratuity left for the waitress in a restaurant and to the waitress's self-reported mood.

Of course, the fact that sunshine was associated with both customers' helping and waitresses' mood does not necessarily mean that changes in customers' mood produced the variations in helping. The observed variations in both helping and mood as a function of sunshine could have involved separate mediational mechanisms. Yet because of the strong experimental evidence for an effect of mood on helping (Isen & Levin, 1972; Cunningham, Steinberg, & Grev, in press) and the lack of a plausible alternate mediator for the effect of sunshine on helping, mood seems a prime candidate. The present study did not, however, provide information on how sunshine might affect mood. Explanations based on symbolic associations, aesthetics, and biological processes all seem reasonable.

Sunshine level could influence mood through its symbolic connection with pleasant or disappointing events. Thus sunshine could increase mood by stimulating thoughts of swimming, picnics, and other outings, whereas cloudy days could be associated with the disappointment of canceled plans and the annoyance of rain and snow and could in that way alter the individual's mood.

Alternatively, sunshine could produce a more positive mood by illuminating the environment in a more stimulating and pleasing manner. Clear sunlight has spectral characteristics different from the light on cloudy days and with its stronger intensity may enhance colors and sharpen detail. The scenery on a cloudy day may appear more dull and monochromatic, by contrast. Maslow and Mintz (1956) have demonstrated that the aesthetic quality of indoor settings affected mood and person perception, and sunshine might similarly influence mood primarily through aesthetic responses to the environment.

Sunshine could also influence mood through its effect on physiological processes. There are indications both from experimental studies on light deprivation and clinical studies of blind and cataract patients that the level of light detected may affect adrenal corticosteroid production and other endocrine functions, hemoglobin formation, thyroid activity, the detoxification capacity of the liver, and the overall regulation of circadian and cirannual biorhythms (Luce, 1970). Solar radiation can also increase the atmospheric concentration of negative ions, and increased negative ion concentration has been linked to increased oxidation of serotonin and increased relaxation in humans (Krueger & Reed, 1976; Randall, 1970). Alternately, since cloud cover filters out ultraviolet rays, decreased ultraviolet light may retard physiological processes such as Vitamin D production (Ott, 1973). Yet since no direct measures of blood chemistry were obtained in the present studies, such interpretations of the effect of sunlight, like those based on associations and aesthetics, remain speculative.

The series of studies reported here employed a quasi-experimental correlation design, and the lack of total control over the independent variables produced some ambiguous results. First, in Experiment 2 older people and women were more likely to eat in the restaurant, and thus participate in the study, on sunnier, cooler, windier, more humid, higher pressure days. Although it is believed that statistical control for such self-selection factors was adequate, complete random assignment to conditions would have been more desirable.

Nowhere are the weaknesses of the correlational approach more apparent than in the case of lunar phase effects. In the zero-correlation analysis in Experiment 1, a negative association with helping was found in both seasons, but this relationship disappeared in the partial correlation analysis. Experiment 2 found a significant positive relationship with tipping, although no association was found

with waitresses' mood. Since a recent review of the lunar phase literature pertaining to homicide, suicide, and psychiatric admission frequency (Campbell & Beets, 1978) indicated substantial inconsistency in findings across studies and argued that published significant associations were due to Type I errors, the present small but significant correlations are regarded as somewhat of a nuisance. As with other variables, such as sunshine, employed in this study, a full explanation of proposed effects will require measurement of physical, physiological, and psychological processes influenced by a factor such as the moon. Given the overabundance of speculation concerning the nature of lunar influences, no interpretation of the present findings will be offered.

Two other apparent inconsistencies should be noted. In Experiment 1 humidity was negatively associated with helping, whereas in Experiment 2 the association was positive. This discrepancy is reasonable, however, if one assumes that the higher the outdoor humidity, the more relief the restaurant patron experienced when coming into a climate-controlled environment. Appreciation or more positive mood following such relief could have increased tipping.

Another inconsistency concerns the relationship of temperature with the various dependent measures. Temperature showed an inverted U-shaped relationship with helping in Experiment 1 but was positively related to waitresses' mood in Experiment 2. Yet the positive relation with waitresses' mood is compatible with an overall curvilinear relationship, given the more restricted temperature range of Experiment 2. Temperature was not, however, clearly linearly or curvilinearly related to tipping in Experiment 2. This is not surprising, however, given the climate-controlled environment of the restaurant. Further, although waitresses' mood was recorded when the waitresses first came in to begin their shifts, tipping did not occur until the patrons had been away from the outdoor temperature for some time.

The lack of a substantial relationship of barometric pressure with helping in either study was somewhat unexpected in light of previous research. Yet it is well to bear in mind Piccardi's (1962) observation that simply ascending to the top of a skyscraper involves a pressure change equivalent to that stemming from a tornado. Barometric pressures in Experiment 1 ranged from 29.63 to 30.37 and in Experiment 2 from 29.60 to 30.27, which is an average range, but since data were not gathered during rainy or stormy weather, it is possible that some pressure effects were missed.

Also unexpected was the finding that the consumption of alcohol was not related to tipping, nor was the size of the party, contrary to the results of previous research (Freeman, Walker, Borden, & Latane, 1975). But since the precise amount of alcohol consumed was not recorded, and since it was frequently difficult to determine if one person or the entire group contributed to the tip, few conclusions can be drawn.

Although this series of studies has stressed the role of weather variables in helping, the conclusion should not be drawn that the weather is the only or even the major factor contributing to altruism. The strongest predictor, sunshine level, accounted for only 13% of the helping variance when participants were outdoors and 4% of the variance when people were indoors, leaving a great deal to be accounted for by social and individual difference factors. Yet given its relative neglect in the literature, future investigations might examine the relationship between sunshine and other affect-linked behaviors. Alternative indicators of helpfulness or friendliness, such as people's willingness to start a conversation with a stranger or give to charity, might be examined, as well as negative affect-related behaviors such as reports of crime, marital distress, suicide, initial contact with a psychiatric clinic, and constricted nonverbal behavior (Cunningham, 1977). Given the reported friendliness of Californians, it would also be of interest to examine the friendliness of various cities around the world as a function of the amount of sunlight they receive (cf. Feldman, 1968; Robbins, DeWalt, & Pelto, 1972). And the next time experimental subjects are not

performing as predicted, the laboratory psychologist might look outdoors to see what kind of a day it is.

References

Banziger, G., & Owens, K. Geophysical variables and behavior: II. Weather factors as predictors of local social indicators of maladaption in two non-urban areas. *Psychological Reports*, 1978, *43*, 427–434.

Baron, R. A., & Bell, P. A. Aggression and heat: The influence of ambient temperature, negative affect, and a cooling drink on physical aggression. *Journal of Personality and Social Psychology*, 1976, *33*, 245–255.

Campbell, D. E., & Beets, J. L. Meteorological variables and behavior: An annotated bibliography. JSAS *Catalog of Selected Documents in Psychology*, 1977, *7*, 1. (Ms. No. 1403)

Campbell, D. E., & Beets, J. L. Lunacy and the moon. *Psychological Bulletin*, 1978, *85*, 1123–1129.

Campbell, D. T., & Stanley, T. C. *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally, 1963.

Cialdini, R. B., Vincent, J. E., Lewis, S. K., Catalan, T., Wheeler, D., & Darby, B. L. Reciprocal concession procedure for inducing compliance: The door-in-the-face technique. *Journal of Personality and Social Psychology* 1975, *31*, 206–215.

Cunningham, M. R. Personality and the structure of the non-verbal communication of emotion. *Journal of Personality*, 1977, *45*, 564–584.

Cunningham, M. R., Steinberg, J. M., & Grev, R. Wanting to and having to help: Separate motivations for positive mood and guilt-induced helping. *Journal of Personality and Social Psychology*, in press.

Dexter, E. School deportment and the weather. *Educational Review*, 1904, *19*, 160–168.

Digon, E., & Bock, H. Suicides and climatology. *Archives of Environmental Health*, 1966, *12*, 279–286.

Dirnagl, K. Neuere Untersuchungsergebnisse zur Beeinflussung des menschlichen Befindens durch das Wetter. *Therapiewoche*, 1977, *27*, 858–870.

Feldman, R. L. Response to compatriots and foreigners who seek assistance. *Journal of Personality and Social Psychology*, 1968, *10*, 202–214.

Freeman, S., Walker, M. R., Borden, R., & Latane, B. Diffusion of responsibility of restaurant tipping: Cheaper by the bunch. *Personality and Social Psychology Bulletin*, 1975, *1*, 584–587.

Griffitt, W., & Veitch, R. Hot and crowded: Influences of population density and temperature on interpersonal affective behavior. *Journal of Personality and Social Psychology*, 1971, *17*, 92–98.

Huntington, E. *Mainsprings of civilization*. New York: Wiley, 1945.

Isen, A. M., & Levin, P. F. Effects of feeling good on helping: Cookies and kindness. *Journal of Personality and Social Psychology*, 1972, *15*, 294–301.

Krueger, A. P., & Reed, E. J. Biological impact of small air ions. *Science*, 1976, *193*, 1209–1213.

Larson, C. T. *Ser. 1: Environmental abstracts*. Ann Arbor: University of Michigan, 1965.

Lester, D. The snow-fall and suicide in Buffalo. *The Bulletin of the Suicide Prevention and Crisis Service*, 1970, *2* (Supplement), 46–47.

Lockard, J. S., McDonald, L. L., Clifford, D. A., & Martinez, R. Panhandling: Sharing of resources. *Science*, 1976, *191*, 406–408.

Luce, G. G. *Biological rhythms in psychiatry and medicine*. Washington, D.C.: National Institute of Mental Health, 1970.

Maslow, A. H., & Mintz, N. L. Effects of esthetic surroundings: 1. Initial effects of three esthetic conditions upon perceived "energy" and "well-being" in faces. *Journal of Psychology*, 1956, *41*, 247–254.

Mather, T. R. *Climatology: Fundamentals and applications*. New York: McGraw-Hill, 1974.

Mills, C. Suicides and homicides and their relation to weather changes. *American Journal of Psychiatry*, 1934, *91*, 669–677.

Moos, R. H. *The human context*. New York: Wiley, 1976.

Muecher, H., & Ungeheuer, H. Meteorological influence on reaction time, flicker fusion frequency, job accidents and use of medical treatment. *Perceptual and Motor Skills*, 1961, *12*, 163–168.

Ott, T. N. *Health & light*. Old Greenwich, Conn.: Devin-Adair Company, 1973.

Persinger, M. S. Lag response in mood reports to changes in the weather matrix. *International Journal of Biometeorology*, 1975, *19*, 108–114.

Piccardi, G. *The chemical basis of medical climatology*. Springfield, Ill.: Charles C Thomas, 1962.

Randall, W. The behavior of clouds correlates with thyroid activity in cats. *Psychonomic Science*, 1970, *21*, 270–271.

Robbins, M. C., DeWalt, E. R., & Pelto, P. J. Climate and behavior: A biocultural study. *Journal of Cross-Cultural Psychology*, 1972, *3*, 331–334.

Tromp, S. W. *Progress in Biometeorology Division A: Progress in human biometeorology* (Vol. 1, Part 1A). Amsterdam: Swets and Zeitlinger BV, 1974.

Winslow, C. E., & Herrington, L. P. Subjective reaction of human beings to certain outdoor atmospheric conditions. *American Society of Heating and Ventilating Engineers Transcript*, 1936, *42*, 119–132.

Winslow, C. E., & Herrington, L. P. *Temperature and human life*. Princeton, N.J.: Princeton University Press, 1949.

Yaglow, C. P., & Miller, W. E. Effective temperature with clothing. *American Society of Heating and Ventilating Engineers*, 1925, *31*, 89–100.

Received July 31, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 11, 1957-1965

Divergence of Private Evaluations of Alternatives Prior to a Choice

Paul J. Brounstein
University of Maryland, College Park

Nancy Ostrove
Lake Erie College

Judson Mills
University of Maryland, College Park

In an attempt to reconcile past research based on reactance theory that found convergence of evaluations of alternatives prior to a choice, and research based on choice certainty theory that found divergence, it was proposed that convergence occurs in overtly expressed evaluations but that divergence occurs in private evaluations. According to this interpretation, the freedom to choose a nonpreferred alternative is threatened by the overt expression of a preference that involves commitment but is not threatened by the private holding of a preference. As expected from this interpretation, using the procedure in which Linder and Crane and Linder, Wortman, and Brehm found convergence of overtly expressed evaluations, divergence was found when evaluations were made privately, that is, supposedly without the experimenter being aware of the nature of the evaluations.

Brehm's (1966) theory of psychological reactance has been applied to evaluations of alternatives prior to a choice by Wicklund (1968, 1970), who suggested that a preference for one alternative over another may threaten the freedom to choose the nonpreferred alternative and to reject the preferred alternative. Reactance theory assumes that a threat to a freedom leads to attempts to reassert that freedom. The freedom to choose a nonpreferred alternative can be reasserted by increasing the attractiveness of the nonpreferred alternative relative to the preferred alternative. This change would result in convergence of evaluations of the alternatives.

Elaborating this analysis, Linder and Crane (1970) hypothesized that ratings of the attractiveness of alternatives will converge as a choice approaches. Linder and Crane had female undergraduates rate the desirability of two interviewers after reading brief descriptions of them. Some of the subjects had been led to believe that they would choose which of the two would interview them about highly personal topics following a brief "get-acquainted" period with both potential interviewers. The length of the get-acquainted period was varied. Linder and Crane found that the difference in the ratings, given the two interviewers, was smaller when the choice was expected after a 3-minute get-acquainted period than when the get-acquainted period was expected to last 15 minutes or when no choice was expected.

Linder, Wortman, and Brehm (1971) used a procedure similar to that of Linder and Crane but varied the time to the choice while holding the amount of time in the get-acquainted period constant. Replicating the Linder and Crane effect, they found that the difference in the ratings given the two interviewers was smaller when the time to the

This study was supported by a grant from the National Science Foundation. Computing facilities were provided by the University of Maryland Computer Science Center. The authors are grateful to Darwin Linder for providing copies of the instructions and materials used in the study by Linder and Crane (1970).

Paul Brounstein is now at the Granville Corporation, Washington, D.C. Nancy Ostrove is now at the Uniformed Services University of Health Sciences, Bethesda, Maryland.

Requests for reprints should be sent to Judson Mills, Department of Psychology, University of Maryland, College Park, Maryland 20742.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3711-1957$00.75

choice was 3 minutes than when it was 10 minutes or when no choice was expected.

The studies by Linder et al. provide evidence of convergence of alternatives prior to a choice. However, it is possible that the convergence occurred not because the freedom to choose the nonpreferred alternative was threatened by the private holding of a preference but rather because it was threatened by the overt expression of a preference that involved commitment. In the studies by Linder et al., the ratings of the alternatives were not completely private but were made with the knowledge that they would be seen by the experimenter. This could have made the subjects feel that expressing a preference for one alternative would commit them to choosing it, for to do otherwise would leave the experimenter with the impression that they were inconsistent or flighty. Such an impression should be expected to be stronger, the shorter the time from the preference to the choice.

The assumption that the overt expression of a preference that involves commitment threatens the person's freedom to choose the nonpreferred alternative, whereas the private holding of a preference does not, provides a means of reconciling the findings of Linder et al. with research by O'Neal and Mills (O'Neal & Mills, 1969; Mills & O'Neal, 1971; O'Neal, 1971) that has provided evidence of divergence in evaluations of alternatives prior to a choice. The research by O'Neal and Mills was based on a theory of choice certainty (Mills, 1968), which assumes that people want to be certain when taking an action that it is better than the other alternatives. Certainty about a prospective choice can be increased by increasing the attractiveness of the preferred alternative relative to the nonpreferred alternative, which would result in divergence of evaluations of the alternatives.

O'Neal and Mills (1969) tested the hypothesis that individuals faced with a choice about other persons will demonstrate a greater halo effect in their impressions of those persons. They found that the anticipation of making a choice about which women in a set of photographs were promiscuous increased the intercorrelation of rankings of the women on desirable traits such as *artistic* and *sincere*. Giving one person consistently high rankings on desirable traits and another consistently low rankings indicates a greater difference in attractiveness than do rankings that are not as consistently high or low. In a follow-up study, Mills and O'Neal (1971) replicated the findings of O'Neal and Mills and showed that they could not be attributed to greater attention paid to the choice photographs.

Using a choice about the subject's partners for a group task, O'Neal (1971) showed that the greater the importance of the anticipated choice, the greater the magnitude of the halo effect. When the prospective choice would definitely determine the subject's partners, it increased the intercorrelation of rankings of the potential partners on desirable traits; but when there was only a small chance that the choice would determine the subject's partners, it did not. O'Neal also found evidence that the influence of an anticipated choice on the halo effect depends on the presence of unexplained arousal. Subjects who were aroused by means of caffeine and were not informed about the source of their arousal showed the effect of anticipated choice on the halo effect, whereas those who were given caffeine and told that it was a stimulant, or those who were given a placebo, did not. The reasoning behind the prediction of this finding was that a lack of certainty about an anticipated choice is typically accompanied by arousal, and when a person anticipating a choice is aroused and does not know the source of his arousal, he may interpret it as resulting from a lack of certainty about the choice, which may have the effect of increasing his desire for certainty.

In the studies by O'Neal and Mills that found divergence in the attractiveness of alternatives prior to a choice, the subjects were not asked to make direct evaluations of the alternatives. Although the rankings of the choice alternatives in those studies were given to the experimenter, they were made on a number of different positive traits and did not, as Wicklund (1974) has noted, involve a bold, blatant discrimination, as in the studies by Linder et al. The assumption that convergence occurs in overtly expressed evaluations of alternatives prior to a choice but that divergence occurs in private evaluations is consistent with the previous research.

The purpose of the present research was to test the interpretation that when evaluations of alternatives involve public commitment, they will converge prior to a choice, but when evaluations do not involve commitment, they will diverge prior to a choice. It was assumed that reactance theory, which predicts convergence in the attractiveness of alternatives prior to a choice, and choice certainty theory, which predicts divergence, are both valid but that predictions of the two theories apply to different circumstances. The prediction of convergence from reactance theory was assumed to apply when there is an overt expression of a preference that involves commitment, and the prediction of divergence from choice certainty theory was assumed to apply when evaluations of the alternatives are private.

Experiment 1

The first study was designed to vary the anticipation of a choice and whether evaluations of the alternatives were public or private, while paralleling the procedure of Linder et al. as closely as possible. A predecision condition, similar to their 3-minute predecision condition, and a no decision condition, the same as their no decision condition, were included. The procedure of Linder et al. was modified slightly so that a private assessment condition, in which the subjects would think that their evaluations would not be known by the experimenter, could be included as well as a public assessment condition, similar to the assessment procedure Linder et al. used. In addition to the same direct assessment of the desirability of the two interviewers used by Linder et al., the subjects were asked to indicate how applicable a number of positive traits were to the interviewers.

In the only other change from the procedure used by Linder et al., the instructions concerning the get-acquainted period with both interviewers were omitted. It was thought that the expectation of meeting the two interviewers would have the effect of minimizing changes in the attractiveness of the interviewers. Research by Walster, Berscheid, and Barclay (1967) has shown that changes in the attractiveness of an alternative are attenuated by the expectation that objective information about that alternative will be forthcoming. In addition, if information that might increase certainty about the choice were expected, there would be less need for the subjects to increase their certainty by changing their evaluations of the alternatives. The instructions concerning the get-acquainted period were not discussed by Linder et al. in relation to the arousal of reactance, and there seemed no reason why the amount of reactance should be affected by the omission of the get-acquainted period.

Method

The subjects were 80 female students in introductory psychology who volunteered, knowing that they would earn extra credit toward their course grades. An equal number was randomly assigned to each of the four experimental conditions: predecision–public assessment, no decision–public assessment, predecision–private assessment, no decision–private assessment. Within each of the four conditions, half of the subjects were run by a female experimenter and half by a male experimenter.

When the subject arrived, she was shown to the experimental room and was given the same instructions concerning the National Foundation for Opinion Research as in the studies by Linder et al. Subjects in the predecision conditions were told that they would be interviewed concerning personal matters ranging from dating patterns and social regulations to their views on sex and that they would be able to choose the interviewer from two graduate students in clinical psychology who had both been judged satisfactory in terms of technique but who differed in their personality characteristics. Subjects in the no decision conditions were told that they would not be interviewed but were being asked to read written descriptions of two interviewers in order to determine the effects of the descriptions. All subjects read the same personality sketches about the two interviewers, Carter and Williams, used by Linder et al.

Afterwards, in a departure from the procedure used by Linder et al., the subject was seated in front of an impressive-looking apparatus previously employed in studies using the Bogus Pipeline (Jones & Sigall, 1971), with the explanation that it was a part of a new measurement technique about which they would like her opinions. The subject was told that when the machine was on, it was hooked up to a central information bank that automatically recorded responses made by turning a dial. The experimenter showed the subject how the dial controlled the movement of a needle on a scale marked from 0 to 30.

Mentioning that he/she would turn on the machine, the experimenter flicked the "on" switch and a loud clattering noise emanated from the apparatus.

Table 1
Means of the Absolute Difference in the Ratings for Experiment I

Assessment	Decision condition	
	No decision	Predecision
Public	10.7	7.9
Private	8.1	15.4

Note. $n = 20$ per cell.

Quickly turning the switch to "off," the experimenter remarked with exasperation that there seemed to be a short circuit somewhere and said that the subject could still practice on the machine "to get the feel of it" while the problem was being fixed. As if on the spur of the moment, the experimenter suggested that the subject use her impressions of the two interviewers during the practice session. For subjects in the predecision conditions, the experimenter added that the practice rarely took more than 3 minutes, after which the subject would choose her interviewer.

The subjects in the public assessment conditions were asked to indicate their responses to items read aloud by turning the dial to the appropriate needle reading and reading the number aloud to the experimenter before returning the dial to 0. In clear view of the subject, the experimenter wrote down her responses.

The subjects in the private assessment conditions were asked to indicate their responses to items read aloud just by turning the dial to the appropriate reading and returning it to 0. The experimenter sat behind a partition that blocked the subject's and experimenter's views of one another, supposedly so that the experimenter's presence would not distract the subject. From a dial concealed behind the partition and connected to the subject's dial, the experimenter recorded the subject's responses.

As in the procedure of Linder et al., the subject indicated how much she agreed with the statement, "I would like to have my interview with Mr. Carter" on a scale from 0 (not at all) to 30 (very much) and rated Williams on the same scale. The subject also indicated how well different traits applied to each of the two men on a scale from 0 (extremely uncharacteristic) to 30 (extremely characteristic). The traits were: capable, kind, patient, responsible, sincere, and understanding.

Following the evaluations of the interviewers, the subjects in the predecision conditions were asked if they had already made a choice about their interviewer. Two of the subjects in the predecision–public assessment condition and one in the predecision–private assessment condition said that they had decided previously. The experimenter casually mentioned to all subjects that there was something more to the study and asked if they had any idea what it might be. The response of one person run under the predecision–public assessment condition indicated suspicion about the procedure, and she was not included as a subject. Finally, the true purpose of the experiment was explained, and each subject promised not to discuss it with others.

Results

The absolute value of the difference between the ratings given Carter and Williams was calculated in the same manner as in the studies by Linder et al. Means of the absolute difference in the ratings for the experimental conditions are presented in Table 1. As can be seen from Table 1, when the assessment of the alternatives was public, the results were similar to those obtained in the studies by Linder et al.; the absolute difference in the ratings of Carter and Williams was smaller in the predecision–public assessment condition than in the no decision–public assessment condition. When the assessment of the alternatives was private, the results were reversed; the absolute difference in the ratings was greater in the predecision–private assessment condition than in the no decision–private assessment condition.

An analysis of variance of the absolute difference in the ratings revealed a significant Decision × Assessment interaction, $F(1, 72) = 6.96$, $p < .01$. The main effects for decision and assessment were not significant, nor were any of the effects involving sex of the experimenter. A planned comparison revealed that the difference between the predecision–public assessment condition and the no decision–public assessment condition was not significant. A second planned comparison revealed that the difference between the predecision–private assessment condition and the no decision–private assessment condition was significant, $F(1, 72) = 7.80$, $p < .01$.

An indirect measure of the evaluation of the two interviewers was computed by taking the absolute difference between the sum of the scores given Carter on the six traits and the sum of the scores given Williams on the traits. The results for the indirect measure were in the same direction as for the direct measure; that is, the absolute difference was less in the predecision–public assessment condition than in the no decision–public assessment condition and was greater in the predecision–private

ssessment condition than in the no decision–rivate assessment condition. However, the ifferences between the conditions were not gnificant for the indirect measure.

Discussion

The results of the first study lend support to the proposed interpretation of the research based on reactance theory (Linder & Crane, 1970; Linder, Wortman, & Brehm, 1971) that found evidence of convergence of evaluations of alternatives prior to a choice and the research based on choice certainty theory (Mills & O'Neal, 1971; O'Neal, 1971; O'Neal & Mills, 1969) that found evidence of divergence. The interpretation reconciles the apparently conflicting results of the past research by making the assumption that the reactance theory prediction of convergence applies to the overt expression of a preference that involves commitment, whereas the choice certainty theory prediction of divergence applies to private evaluations of alternatives. As expected from the proposed interpretation, using the procedure in which Linder et al. found convergence of overtly expressed evaluations, divergence was found when evaluations were made privately.

Although the results for the public assessment conditions paralleled the results in the studies by Linder et al., they were not strong enough to be statistically significant. The only substantial change from the procedure of Linder et al. was the dropping of the instructions concerning the get-acquainted period with both interviewers. As mentioned in the introduction, it was thought that the expectation of meeting both interviewers in the get-acquainted period would make it less likely that the subjects would change their evaluations of the interviewers. It was assumed that the instructions about the get-acquainted period did not influence the arousal of reactance.

It is possible that the anticipation of getting more information concerning the alternatives does affect the amount of reactance created by the overt expression of a preference prior to a choice. According to reactance theory, the importance of the freedom that is threatened determines the amount of reactance created by the threat. The freedom to choose the non-preferred alternative and reject the preferred alternative may be more important when additional information about both of the alternatives is expected. If so, then the convergence effect should be stronger when the instructions concerning the get-acquainted period with both interviewers are included.

Experiment 2

The second study was designed to replicate and extend the findings of the first study. The variables manipulated were time to the choice and whether or not additional information was expected, as well as whether evaluations of the alternatives were public or private. In addition, an attempt was made to replicate exactly the procedures of the 10- and 3-minute predecision conditions of Linder, Wortman, and Brehm (1971).

One major prediction was that the reactance effect—that is, convergence in the evaluations of alternatives as time to choice decreases—would occur when the evaluations are public and additional information is expected. From the assumption that when additional information is not anticipated, the importance of the freedom to choose the non-preferred alternative and reject the preferred alternative is reduced, it was predicted that the convergence effect expected from reactance would be reduced when additional information is not expected. It follows that when evaluations of the alternatives are public and the choice close, the absolute difference in the ratings of the alternatives should be smaller when additional information is anticipated than when it is not anticipated.

The other major prediction was that the choice certainty effect—that is, divergence in the evaluations of alternatives as time to choice decreases—would occur when the evaluations are private and additional information is not expected. It was assumed that the expectation of receiving additional information that might increase certainty about the choice would decrease the need to increase certainty through the reevaluation of alternatives. From this assumption it was predicted that the divergence effect would be reduced when additional information was expected. It follows that when evaluations are private and the

choice close, the absolute difference in the ratings of the alternatives should be greater when additional information is not anticipated than when it is anticipated.

Method

The subjects were 100 female undergraduates in introductory psychology who volunteered, knowing that they would receive extra credit toward their course grades. Ten subjects were randomly assigned to each of the eight experimental conditions of a $2 \times 2 \times 2$ factorial design varying time to choice (10 minutes, 3 minutes), information (expected, not expected), and assessment (public, private). Ten subjects were later randomly assigned to each of the two replication conditions.

The procedure was the same as in Experiment 1, with the following exceptions. All subjects were run by a male experimenter, and all were told that they would be interviewed and would be able to choose their interviewer. In the descriptions of the interviewers, the name Carter was changed to Caster.

Half of the subjects in the factorial design were told that they would be able to meet both interviewers in a brief get-acquainted session prior to the choice (information expected), whereas the other half were not so informed (information not expected). Just prior to the ratings of the interviewers, subjects in the 10-minute conditions were told that one of the interviewers was tied up for the next few minutes, whereas subjects in the 3-minute conditions were told that the interviewers should be ready momentarily.

In the 10 minute – information expected conditions the instructions were as follows:

> After the practice trials, when both men are finally available, you'll be able to have your chat with them. This session will last a couple of minutes, so considering the practice, get-acquainted period, and all, you'll be making your final choice of interviewer in approximately 10 minutes.

In the 10 minute – information not expected conditions the instructions were:

> After the practice trials, I'll get everything ready for the interview; that will also take a couple of minutes, so considering the practice and all, you'll be making your choice of interviewer in approximately 10 minutes.

In the 3 minute – information expected conditions the instructions were:

> After the practice trials, you'll be able to have your chat with the two interviewers, which will last a couple of minutes, so considering the practice and the get-acquainted period, you'll be making your choice of interviewer in approximately 3 minutes.

In the 3 minute – information not expected conditions the instructions were:

> After the practice trials, I'll get everything ready for the interview; that will also take a couple of minutes, so considering the practice and all, you'll be making your choice of interviewer in approximately 3 minutes.

After suggesting that the subject use her first impressions of the two interviewers for the practice session, the experimenter went behind the partition and determined from a random assignment list whether the subject's preference ratings were to be assessed under public or private conditions. The instructions for the public and private assessment conditions were the same as in Experiment 1, but in both conditions the experimenter was behind the partition, and the instructions were given by means of a tape recording to eliminate any possibility that the experimenter could influence the subject's responses by the way he administered the dependent measures. The order of the questions concerning Caster and Williams was counterbalanced.

Subjects in the 10-minute and 3-minute replication conditions were given a booklet that was identical to the one used by Linder et al. (except that Carter was changed to Caster). They expected to receive additional information about the choice alternatives, and the segment of the procedure dealing with the measurement device was omitted. The experimenter manipulated the time-to-choice variable in exactly the same manner as in the factorial design. Ratings of the two interviewers were made on questionnaires that were filled out in the presence of the experimenter and were returned directly to him.

After the dependent measures had been collected, subjects in both the factorial design and the replication conditions were asked if they had already made a choice about their interviewer. One subject in each of the four information not expected conditions said that she had decided previously. The experimenter casually mentioned to all subjects that there was something more to the study and asked if they had any idea what it might be. The responses of four persons indicated accurate suspicion about the procedure and they were not included as subjects. One was run under the 3 minute – information not expected – public assessment condition, one under the 10 minute – information expected – private assessment condition, one under the 3 minute – information not expected – private assessment condition, and one under the 10-minute replication condition.

Finally, the true purpose of the experiment was explained, and each subject promised not to discuss it with others.

Results

The means of the absolute difference in the ratings given Caster and Williams for the eight experimental conditions of the factorial

design are presented in Table 2. As can be seen from Table 2, when the assessment of the alternatives was public, the results were similar to those of Linder et al. and the public assessment conditions of Experiment 1; the absolute difference in the ratings was smaller in the 3-minute condition than in the 10-minute condition when additional information was expected and also when it was not expected. Similar results were obtained in the replication conditions; the mean for the 3-minute replication condition was 7.2 and the mean for the 10-minute replication condition was 8.4.

When the assessment of the alternatives was private and additional information was not expected, the results paralleled those of the private assessment conditions of Experiment 1; the absolute difference in the ratings was greater in the 3 minute – information not expected condition than in the 10 minute – information not expected condition. When additional information was expected, there was virtually no difference between the 3-minute and 10-minute conditions.

A $2 \times 2 \times 2$ analysis of variance of the absolute difference in the ratings revealed that the following effects were significant: the main effect of assessment, $F(1, 92) = 5.11$, $p < .05$; the main effect of information, $F(1, 92) = 9.78$, $p < .01$; the interaction between time to choice and assessment, $F(1, 92) = 4.35$, $p < .05$.[1] The major predictions were tested with the use of planned comparisons.

The major prediction from reactance theory was that the absolute difference in the ratings of the alternatives would be less in the 3 minute – information expected – public assessment condition than in the 10 minute – information expected – public assessment condition. This comparison was not significant. The comparison between the 3 minute – information not expected – public assessment condition and the 10 minute – information not expected – public assessment condition was also not significant. The two 3 minute – public assessment conditions were contrasted with the two 10 minute – public assessment conditions, and the contrast was not significant. The 3-minute replication condition and the 10-minute replication condition were not significantly different. The 3-minute replication condition and the two 3-minute public assessment conditions were contrasted with the 10-minute replication condition and the two 10-minute public assessment conditions. This contrast was also not significant. In addition, the contrast between the 3 minute – information expected – public assessment condition and the 3 minute – information not expected – public assessment condition was not significant.

Table 2
Means of the Absolute Difference in the Ratings for Experiment 2

Assessment	Information	Time to choice: 10 min.	Time to choice: 3 min.
Public	Expected	7.1	5.4
	Not expected	9.5	6.6
Private	Expected	7.0	7.1
	Not expected	10.0	14.9

Note. $n = 10$ per cell.

The major prediction from choice certainty theory was that the absolute difference in the ratings would be greater in the 3 minute – information not expected – private assessment condition than in the 10 minute – information not expected – private assessment condition. This comparison was significant, $F(1, 92) = 4.53$, $p < .05$. As expected, the 3 minute – information not expected – private assessment condition was significantly greater than the 3 minute – information expected – private assessment condition, $F(1, 92) = 11.47$, $p < .01$.

Discussion

The results of the second study fortify the proposed interpretation, which reconciles the

[1] The error term for the analysis of variance and the planned comparisons was derived by pooling the estimate of error from the factorial and replication cells. Exclusion of the replication cells from the error term would not affect the significance levels of any of the comparisons. There would not be any change in the significance levels with exclusion of the four subjects who said they had already made a choice.

apparently conflicting results of the past research (Linder & Crane, 1970; Linder, Wortman, & Brehm, 1971; O'Neal & Mills, 1969; Mills & O'Neal, 1971; O'Neal, 1971) by making the assumption that the reactance theory prediction of convergence of alternatives prior to a choice applies to the overt expression of a preference that involves commitment, whereas the choice certainty theory prediction of divergence applies when evaluations are private. As expected from the proposed interpretation, when preferences were privately assessed and additional information was not expected, the absolute difference in the ratings of the alternatives was greater, the closer the choice. The replication of the divergence effect found in the private assessment conditions of Experiment 1 lends substantial support to part of the proposed interpretation based on choice certainty theory.

That divergence of private evaluations did not occur when additional information was expected is consistent with the choice certainty theory analysis. If additional information is anticipated, there should be little reason to reevaluate the choice alternatives, since the forthcoming information may be sufficient to achieve certainty about which alternative is best. As expected from this reasoning, when evaluations were private and the choice close, the absolute difference in the ratings was greater when additional information was not expected than when it was expected.

The divergence effect that was found in the present research is reminiscent of the divergence that occurs in evaluations of alternatives after a decision (Brehm, 1956). However, there is no basis for regarding the divergence in the present research as postdecisional. It occurred in the private assessment conditions but not in the public assessment conditions, where there was just as much reason, if not more, for subjects already to have made a choice. Also, very few of the subjects in any of the conditions of either Experiment 1 or Experiment 2 said that they had already made a choice.

As in Experiment 1, the results for the public assessment conditions were in the direction of the convergence effect predicted from reactance theory but were not strong enough to be statistically significant. The same was true for the results for the replication conditions, which duplicated as closely as possible the 3-minute and 10-minute conditions of Linder, Wortman, and Brehm (1971).

It is not clear why the convergence effect was not stronger in the present research. The possibility that the lack of a stronger convergence effect in Experiment 1 was due to the omission of the get-acquainted session with both interviewers—which might have reduced the importance of the freedom to choose the nonpreferred alternative—was not borne out by the results of Experiment 2. The tendency for the alternatives to converge in the public assessment conditions was not stronger when the get-acquainted session was included (information expected conditions) than when it was omitted (information not expected conditions).

The fact that the data from the public assessment conditions and the replication conditions were consistently in the predicted direction, together with the significant differences found previously by Linder et al., makes it reasonable to conclude that the convergence effect predicted from reactance theory is reliable. Since it occurs when preferences are public but not when they are private, it can also be concluded that the freedom to choose the nonpreferred alternative and to reject the preferred alternative is threatened by the overt expression of a preference that involves commitment but is not threatened by the private holding of a preference.

References

Brehm, J. W. Post decision changes in the desirability of alternatives. *Journal of Abnormal and Social Psychology*, 1956, *52*, 384–389.

Brehm, J. W. *A theory of psychological reactance*. New York: Academic Press, 1966.

Jones, E. E., & Sigall, H. The Bogus Pipeline: A new paradigm for measuring affect and attitude. *Psychological Bulletin*, 1971, *76*, 349–364.

Linder, D. E., & Crane, K. A. Reactance theory analysis of predecisional cognitive processes. *Journal of Personality and Social Psychology*, 1970, *15*, 258–264.

Linder, D. E., Wortman, C. B., & Brehm, J. W. Temporal changes in predecision preferences among choice alternatives. *Journal of Personality and Social Psychology*, 1971, *19*, 282–284.

Mills, J. Interest in supporting and discrepant information. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook*. Chicago: Rand McNally, 1968.

Mills, J., & O'Neal, E. Anticipated choice, attention and the halo effect. *Psychonomic Science*, 1971, *22*, 231–233.

O'Neal, E. Influence of future choice, importance, and arousal on the halo effect. *Journal of Personality and Social Psychology*, 1971, *19*, 334–340.

O'Neal, E., & Mills, J. The influence of anticipated choice on the halo effect. *Journal of Experimental Social Psychology*, 1969, *5*, 347–351.

Walster, E., Berscheid, E., & Barclay, A. M. A determinant of preference among modes of dissonance reduction. *Journal of Personality and Social Psychology*, 1967, *7*, 211–215.

Wicklund, R. A. *Regret as a result of threat to decision freedom*. Unpublished doctoral dissertation, Duke University, 1968.

Wicklund, R. A. Prechoice preference reversal as a result of a threat to decision freedom. *Journal of Personality and Social Psychology*, 1970, *14*, 8–17.

Wicklund, R. A. *Freedom and reactance*. Potomac, Md.: Erlbaum, 1974.

Received August 11, 1978 ■

Corrections to DePaulo, Rosenthal, Eisenstat, Rogers, and Finkelstein

In the article "Decoding Discrepant Nonverbal Cues" by Bella M. DePaulo, Robert Rosenthal, Russell A. Eisenstat, Peter L. Rogers, and Susan Finkelstein (*Journal of Personality and Social Psychology*, 1978, Vol. 36, No. 3, pp. 313–323), the second, third, and fourth full paragraphs in the left-hand column of p. 320 (starting with "A two-way ANOVA . . .") should be changed to read as follows:

"For the three samples, t tests were computed comparing video primacy for the scenes differing on one affective dimension (slightly discrepant) versus the scenes differing on both affective dimensions (very discrepant). As predicted, there was (slightly) less video primacy (more audio primacy) for the very discrepant scenes than for the slightly discrepant scenes, combined $p = .085$, median $d = .19$."

In addition, the next-to-last sentence in the second full paragraph in the right-hand column of p. 320 should be changed to read:

"However, they were characterized by varying degrees of discrepancy, and subjects tended to attend more to audio cues on the more discrepant scenes than on the less discrepant ones."

Thus, the predicted effect for degree of discrepancy (i.e., the tendency to show relatively greater audio primacy for more discrepant scenes) was much smaller than originally reported. This effect was not stronger for scenes involving the face than for scenes involving the body. This correction is necessitated by a keypunch error that was discovered after publication.

Journal of Personality and Social Psychology
1979, Vol. 37, No. 11, 1966-1981

Consensus Information, Prediction, and Causal Attribution: A Review of the Literature and Issues

Saul M. Kassin
Purdue University

The present review of the literature suggests that it is useful to distinguish between two types of consensus information—normative expectancies (e.g., Jones & McGillis' prior probability concept) and explicit base rates (e.g., Kelley's conception of observed covariation across actors). Normative expectancies, which may be derived from a knowledge of one's own behavior (i.e., the false-consensus effect) or the behavior of others, provide one basis for prediction and causal inference. Explicit, sample-based consensus may also be employed, but under somewhat restrictive conditions: (a) when prior expectations are neutralized and/or (b) when the consensus manipulation is particularly strong, salient, easily translatable, representative of the criterial population, and causally relevant. A number of additional issues are reviewed (e.g., the cognitive strategies by which observers reject base rates), and recommendations for the direction of future research are made.

The attribution literature has recently been inundated with studies aimed at determining whether or not and under what conditions naive observers employ consensus information for predicting and explaining behavioral events. To date some experiments have shown that consensus is a valuable source of information, whereas others have indicated that it is not. This paper will attempt to review and organize the diverse findings and discuss some of the important theoretical issues that have been raised.

Theory

Attribution theories portray the layperson as a perceiver who, in order to understand what caused some form of behavior to occur, actively seeks and utilizes various kinds of information. One of the more intuitively appealing of the proposed relationships is the link between consensus and causal attribution. When an individual behaves differently from others (low consensus), observers are likely to view properties of that individual as having had a causal impact; when an actor behaves in a manner consistent with the performance of others (high consensus), observers are apt to infer that some invariant feature or object in the environment elicited the common reaction.

Heider (1958) illustrated this hypothesis when he suggested that consensus influences attributions about the nature of a percept, the origin of motivation and pleasure, and the causal locus of success and failure. Later attribution theories readily adopted the proposed role of consensus in the inference process. For Kelley (1967), consensus—defined as *observed* covariation across actors (i.e., do other actors in the situation behave similarly) —is one of three criteria for assessing the validity of an environmental attribution, the others being distinctiveness (does the actor's behavior occur only in the presence of a particular entity) and consistency (does the actor behave similarly toward the entity at different times). For temporally consistent behaviors, Kelley (1967, 1973) predicted that

I would like to thank the following people for their helpful comments on earlier drafts of the manuscript: Gary L. Wells, Reuben M. Baron, Charles A. Lowe, Miron Zuckerman, Jack W. Brehm, Lawrence S. Wrightsman, Frederick X. Gibbons, and an anonymous reviewer.

Reprint requests should be sent to Saul M. Kassin, Department of Psychological Sciences, Purdue University, West Lafayette, Indiana 47907.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3711-1966$00.75

those high in consensus (and/or distinctiveness) are causally attributed to the entity, whereas those low in consensus (and/or distinctiveness) are attributed to the actor. Jones and Davis (1965) conceived of consensus as a prior probability variable: Based on a tacit knowledge of norms, social constraints, and values, observers compare an actor's behavior to the *expected* behavior of others. As such, Jones and Davis predicted that behaviors that are low in their "assumed desirability" are more informative of an actor's attributes than those known to be high in their attractiveness or desirability.

In short, theorists have distinguished between two kinds of consensus information—observers' beliefs about what other people *would do* if they were present and observers' firsthand knowledge of what others in the situation *actually do*. The former case exemplifies Jones and Davis's (1965) prior-probability concept and will be referred to as the perceiver's normative expectancy or "implicit consensus." The latter case corresponds to Kelley's (1967) definition of observed covariation and will be referred to as sample-based or explicit-consensus information.

Observers have thus been portrayed as information processors who are "characteristically normative and nomothetic" (Jones et al., 1972, p. 85) and are acutely sensitive to variations in both assumed and observed consensus. Other subsequent formulations (Weiner, 1974) and theoretical integrations (Ajzen & Fishbein, 1975; Anderson, 1974; Jones & McGillis, 1976) have accepted the role of consensus in the attribution process. Moreover, consensus manipulations have been successfully applied by researchers to affect perceptions of task difficulty (Weiner, Frieze, Kukla, Reed, Rest, & Rosenbaum, 1971) and attributions for emotional disorders (Valins & Nisbett, 1971).

Research: An Overview

The Normative Expectancy

Prior to observing the target actor, perceivers—as members of a social structure with often clearly defined behavioral norms—expect to see certain behaviors performed more frequently than others. These crude (i.e., generalized across classes of actors and situations) expectations serve as an initial basis for predicting the frequency of a given act in the population and the behavior potential of a particular individual. For drawing causal inferences, an unexpected act is conceptually equivalent to a low consensus behavior, whereas an expected event is equivalent to one of high consensus.

Generalized behavioral expectancies. Normative expectancies may be derived first from a knowledge of one's own behavior. Heider (1958) often referred to actors' egocentric tendency to use their own behavior as a normative standard. Specifically, he (and recently Ross, 1977) delineated a two-stage process. First, individuals tend to see their own behaviors as environmentally caused and hence consensual. After all, I choose Brand X over Brand Y because Brand X is better; since Brand X is better, most people (would) also choose it. In the second stage, observers who know their own behavior interpret another actor's discrepant performance as low in consensus and attributable to personal causes.

The effects of self-generated consensus have been amply demonstrated. Hansen and Donoghue (1977) had subjects sample a beverage and then watch a confederate who drank either a similar amount or more. In addition to inferring population performance from their own behavior, subjects attributed the other actor's similar performance to the beverage and dissimilar behavior to the actor. Ross et al. (1977) conducted an analogous questionnaire study in which subjects read brief stories of events culminating in a behavioral choice. As it turned out, subjects' own stated choices were related to their estimates of how others would behave. In fact, subjects made more confident and extreme trait ratings to characterize the typical person who would elect the alternative option than to describe the typical person making the same behavioral choice.[1]

Certainly an individual need not actually observe his or her own behavior in order to draw normative inferences. Implicit consensus

[1] Note that these demonstrations of the false-consensus effect are confounded with subject self-selection.

may be generated through a simple knowledge of the population base rates for certain behaviors. In some studies, behaviors that were low in their attractiveness or assumed desirability induced dispositional inferences about the target actors (Ajzen, 1971; Jones, Davis, & Gergen, 1961). In view of the strong correspondence between assumed desirability and perceived probability of occurrence (Ajzen, 1971; Eisen, Note 1), these studies indirectly support the normative expectancy–causal attribution hypothesis. In other studies, Lay, Burron, and Jackson (1973) and Kassin and Lowe (in press) actually pretested the base rates and perceived base rates for stimulus behaviors, respectively. Lay et al. (1973), for example, found that subjects made more confident inferences about the traits of a person who supposedly endorsed a low base-rate item (e.g., "I joke and talk rather than work whenever possible") than about one who endorsed a high base-rate item (e.g., "I love to tell and listen to jokes and funny stories").

Actor-based expectancies. Of course, behavioral expectancies are often refined by an observer's knowledge about characteristics of the target actor (e.g., sex, age, occupation). Jones and McGillis (1976) described actor-based expectancies as follows: "Category membership suggests a modal behavior expectancy or the presence of one categorizing feature (obesity) suggests other correlated features (jolliness)" (p. 413). Kahneman and Tversky (1973), for example, capitalized on popular stereotypes about occupations. In the well-known engineer–lawyer study, they told one group of subjects that a sample contained 70% lawyers and 30% engineers, whereas another group was presented with a sample of 30% lawyers and 70% engineers. Subjects then read descriptions of five target individuals taken from that group and predicted the likelihood that each person was a lawyer (engineer). One description read:

> Jack is a 45-year-old man. He is married and has four children. He is generally conservative, careful, and ambitious. He shows no interest in political and social issues and spends most of his free time on his many hobbies which include home carpentry, sailing, and mathematical puzzles. (p. 241)

Results indicated that subjects underutilized the 70–30 base rates and relied largely on the personality-stereotypic cues (e.g., classifying the individual as an engineer if he shows no interest in political and social issues, enjoys mathematical puzzles, and so on). Since this study, observers' reliance on brief personality sketches for making *category* predictions has been repeatedly demonstrated (e.g., Zuckerman, 1978a).

Predictions about *behavior* are similarly influenced by actor-based expectancies. For example, Miller, Gillen, Schenker, and Radlove (1974) had subjects read about Milgram's (1963) obedience study and then guess how selected stimulus persons responded. They found that the target person's sex and attractiveness affected this measure—subjects predicted that males and unattractive individuals would administer greater levels of shock than females and physically attractive participants would.

Situation-based expectancies. Normative expectancies may also be qualified by situational demands (Herzberger, in press; Jones et al., 1961; Jones & McGillis, 1976). Lowe and Kassin (1977) manipulated the implicit consensus for helping behavior by altering several factors of the described situation (e.g., the personality and appearance of the requester). For subjects who were not presented with the base rates for helping, population estimates were as anticipated: 67% (high implicit consensus), 44% (moderate implicit consensus), and 37% (low implicit consensus). More importantly, causal attribution and trait inference followed from these estimates—subjects evaluated a nonhelping actor more negatively when helping was expected. Others have noted similar effects. Miller et al. (1974) found that Milgram's obedience study elicited clear nonobedience expectations. Nisbett and Borgida (1975) reported that their subjects expected helping behavior in the Darley and Latané (1968) seizure study.

Summary. In sum, it appears that normative expectancies that are based on previous self- and other-observation provide an initial basis for prediction and attribution. To a large extent, these expectations are actor specific (e.g., "males are generally more aggressive than females") and situation specific

(e.g., "people are generally not aggressive in school"). When information about the actor and/or situation is unavailable or uninformative, then observers may fall back on normative expectancies that are more generalized (e.g., "People are generally [not] aggressive").

Explicit Consensus

Explicit consensus refers to the *actual* behavior of individuals in a sample. Early tests of Kelley's (1967) model supported the consensus–attribution hypothesis (Frieze & Weiner, 1971; McArthur, 1972). In the prototypical study, McArthur (1972) presented subjects with behavioral descriptions (e.g., "John laughs at the comedian") to which high or low consensus, distinctiveness, and consistency were appended. Consensus information took the form "almost everyone (hardly anyone) else laughed at the comedian" and was utilized as predicted toward person attribution and especially toward stimulus attribution. In subsequent experimentation, Orvis, Cunningham, and Kelley (1975) found that knowledge of consensus was sufficient for subjects to "figure out" the distinctiveness and consistency of an event—high and low consensus induced inferences of high consistency as well as high and low distinctiveness, respectively.

Finally, two somewhat established empirical patterns are noteworthy. First, explicit consensus appears to be particularly informative about the *stimulus* in a situation (Orvis et al., 1975; McArthur, 1976; Zuckerman, 1978b). Subjects given a stimulus attribution for fictitious events readily infer high consensus (Zuckerman & Mann, in press), solicit consensus more than any other kind of information (Garland, Hardy, & Stephenson, 1975), and rank it as the most important variable governing these decisions (Vestewig, Note 2). Developmentally, the high consensus – stimulus inference relationship precedes the link between low consensus and person inferences (DiVitto & McArthur, 1978). Second, a growing body of evidence suggests that perhaps *observers* make better use of consensus information than do actors, who tend to rely more heavily on distinctiveness (Hansen, 1976; Hansen & Lowe, 1976) or on a knowledge of their own previous behavior (Hansen & Donoghue, 1977; Hansen & Stonner, 1978).

Despite the findings cited above, certain recent developments in the literature have cast some doubt as to whether observers are as sensitive to consensus as was initially believed. Within the attribution area, McArthur (1972) noted that although main effects for consensus on attributions were obtained, it was the least effective of Kelley's three variables. This conclusion about the *relative* ineffectiveness of consensus was extended to issues such as developmental trends in the use of Kelley's variables (DiVitto & McArthur, 1978; Dix, Herzberger, & Erlebacher, Note 3; Karlovac, Feldman, Higgins, & Ruble, Note 4; Swann & Collins, Note 5) and has been widely cited as evidence for the absolute ineffectiveness of consensus.

In the social judgment literature, analogous deficiencies in observers' use of base rates toward prediction have been reported (e.g., the Kahneman & Tversky, 1973, studies). On the one hand, the apparent comparability of results between the attribution (e.g., Nisbett & Borgida, 1975) and judgment (e.g., Kahneman & Tversky, 1973) studies is impressive in view of their contrasting research paradigms (for an excellent discussion, see Fischhoff, 1976). On the other hand, one must be cognizant of the vastly different conclusions that follow from the fact that researchers in the two disciplines pose somewhat different questions. Attribution researchers make the traditional comparison between subjects' responses and the null hypothesis (i.e., does consensus influence inferences, or is it ignored?), whereas the social judgment investigators compare subjects' predictions with those implied by a formal probability model (i.e., do subjects make optimal or suboptimal use of the base rate? see Borgida, 1978; Wells & Harvey, 1978).

Citing the McArthur (1972) and Kahneman and Tversky (1973) experiments, Nisbett and his colleagues performed a number of experiments in which some subjects but not others received base-rate information. No differences in their dependent measures emerged, leading the authors to conclude

hat perceivers ignore consensus. Because of he importance of this research, two studies in particular will be described.

Nisbett and Borgida (1975) presented subjects with descriptions of two experimental situations: the high fear condition of a shock tolerance study (Nisbett & Schachter, 1966) and the emergency condition of a bystander intervention study (Darley & Latané, 1968). Consensus–information subjects read frequency distribution tables indicating that most participants in the shock study tolerated a large degree of shock and that most participants in the helping study either took a long time to help or did not help at all. No consensus–information subjects did not read about these results. In one study, subjects read brief descriptions of participants and guessed how these target actors had behaved. In a second study, subjects were told that the described actors behaved in the extreme, high base-rate manner (i.e., tolerated maximum shock, did not help). They then rated that participant on a number of relevant traits and indicated whether that behavior was personally or situationally caused. Overall, knowledge of base rates had no effect on prediction, attribution, or trait inference.

In another study, Nisbett et al. (1976) failed in three attempts to mitigate depression with the use of high-consensus information that was designed to externalize the perceived cause of the affect, thereby reducing worry. One study that dealt with the "Sunday blues" illustrates their general approach. Male undergraduates filled out a number of mood scales, a questionnaire reporting on their academic and social activities for the day, and rating scales for cartoon funniness. After these initial Sunday premeasures, one group received statistics showing that many students experience the Sunday blues (92% occasionally, 65% often), a second group was provided with an explanation of the phenomenon to supplement this consensus, and a third group was not given any base-rate information. Postmanipulation measures revealed that no changes in mood took place in any of the groups, including those presented with the consensus.

Research since these influential experiments has focused on determining those conditions that mediate the efficacy of explicit consensus. One outcome of the more recent work in this area is the widespread recognition that the interaction between expectancies and sample-based consensus is important and that the consensus–attribution relationship is a complex one whose boundary conditions are easily violated. What has thus far emerged is a collection of empirical findings and issues that suggests the following modest proposition: Any conclusions about the effects of consensus information must be qualified—by stating what kind of consensus manipulation (e.g., implicit or explicit, concretely or abstractly presented), what kind of effect (i.e., individual or population prediction, causal attribution, trait inference), and compared to what response criterion (i.e., the null hypothesis vs. a normative model).

Normative Expectancies Versus Explicit Consensus

The Consistency Problem

Self-generated consensus appears to provide a powerful frame of reference, one that may be based upon a lifetime of self- and other-observation. Some investigators have thus suggested that experimental consensus manipulations might be ineffective because of interference from prior expectations. After all, subjects are typically presented with base rates about a target behavior that in itself arouses normative expectations. Additionally, subjects may observe or read about the surrounding situation, the target stimulus, and/or the target actor, all of which contribute to this prior belief. In other words, sample base rates often comprise only one source of consensus information.

In cases where an observer is confronted with base rates for a behavior that is high or low in implicit consensus, the degree of consistency–inconsistency between the two could partly determine the efficacy of the manipulated (i.e., explicit) consensus. At one extreme, the impact of base rates (i.e., relative to a no-base-rate control) may be minimal if they are *redundant* with prior expectations. In this regard, Wells and Harvey (1977) noted that Nisbett et al. (1976) probably

presented subjects with a redundant and hence uninformative consensus manipulation. Recall that these latter investigators sought to mitigate depressive moods with the use of high-consensus information. Nisbett et al. (1976) noted that the Sunday blues are commonly experienced by students; yet perhaps subjects were also aware that this negative affect was normative. Although the no consensus subjects' prior beliefs were not assessed, their estimates might not have differed from those of the base-rate subjects. In other words, perhaps the manipulation was uninformative, having little effect on already existing beliefs.

A more common problem arises when base rates are rejected for being highly *discrepant* with implicit consensus. Instances of extreme inconsistency are fairly well documented. Nisbett and Borgida (1975) noted this kind of discrepancy in their study when they compared the "naive estimates" made by the no base-rate subjects (i.e., extreme shock tolerance/not helping a victim were unexpected behaviors) with the base rates actually presented (i.e., shock tolerance/nonhelping were high-consensus behaviors). Miller et al. (1974) presented their subjects with the surprising results of Milgram's (1963) obedience study and also found that subjects did not employ the consensus for prediction or explanation. Finally, Kahneman and Tversky (1973) reported that base rates had little effect on category prediction when they conflicted with the expectancies aroused by their target-case descriptions.

In sum, it appears that observers' use of sample-based consensus is minimal when this information is highly redundant or discrepant with baseline expectancies. We turn now to discussions about the cognitive strategies by which individuals reject discrepant base rates and the ways in which the effectiveness of explicit consensus has been increased.

Cognitive Strategies for the Rejection of Explicit Consensus

If observers do indeed reject base rates that are inconsistent with their beliefs, what are the cognitive strategies by which they reject the new information? One possibility is that the perceiver may simply disbelieve the base rate. Results of a survey sponsored by a political party as well as the manipulations of a psychology experimenter may thus be dismissed as false and deceptive. More important, however, are those situations where the observed consensus is *believed* but is rejected as uninformative.

The first mechanism that comes to mind is the possibility for distorted recall. Perhaps subjects who read that very few participants helped a seizure victim actually recalled a higher base rate immediately prior to their dependent variable responses. This possibility is ruled out by consensus recall results (Nisbett & Borgida, 1975; Wells & Harvey, 1977), which held up even when manipulation and assessment were separated in time (Nisbett & Borgida, 1975). Wells and Harvey (1977) did identify one strategy typically employed by observers who are confronted with discrepant base rates. Subjects—either informed or uninformed of sample randomness—read the bystander intervention study and base rates indicating that helping was either high or low in consensus. They then indicated the extent to which the participants comprising the base rate were "representative" and "different" from students in general. In the no-knowledge-of-randomness condition, the lower the base rates for helping were, the more different from people in general that sample of actors was rated. Having been presented with highly unexpected information, subjects readily assumed a *bias* in the observed sample as a means of rejecting the base rate as uninformative of the "true" consensus.

What about observers' inferences about the *size* of a sample? Kahneman and Tversky (1972) asserted that "people often take seriously a result stated in percentages, with no concern for the number of observations, which may be ridiculously small" (p. 44). Wells and Harvey's (1977) findings, however, suggest a second mechanism that perceivers might invoke when confronted with a discrepant *percentage*—to infer that it was based on a small sample. Accordingly, Kassin (1979) had subjects read one of two situations in which helping behavior was either expected or unexpected and results indicating either a high

or low base-rate percentage. As predicted, expectancies and base rates interacted—subjects for whom the consensus percentages were discrepant with normative expectations (i.e., high expectancy, low base rate and low expectancy, high base rate) assumed that the consensus had been derived from a small sample. In sum, observers have at their disposal certain mechanisms with which they justify their dismissal of sample-based consensus information.

Effective Manipulations of Explicit Consensus

In view of the problems associated with highly discrepant or redundant consensus information, it is not surprising that significant effects for sample-based consensus have generally been obtained under somewhat restrictive conditions—either when the manipulation is strengthened (e.g., when it is particularly salient) or when normative expectancies have been neutralized (i.e., in the absence of information about the target actor and stimulus). What follows now is a review of those conditions that have successfully increased observers' use of base rates.

Neutralization of Expectancies

A number of investigators have obtained strong effects for sample-based consensus by denying subjects access to expectancy-arousing information. For example, Kassin and Lowe (Note 6) conducted two experiments in which consensus and distinctiveness information were illustrated through the movement of objects (after Heider & Simmel, 1944). Specifically, a series of animated films was shown depicting interactions between three triangles (Target Persons A, B, C) and three squares (Target Stimuli X, Y, Z). In one experiment, Target Actor B bumped Target Stimulus Y. For high consensus, Persons A and C also bumped Y; for low consensus, A and C approached but did not bump Y. Distinctiveness was similarly operationalized by varying the behavior of Actor B toward the other stimuli, X and Z. In contrast to some previous findings, consensus had a large effect on causal attribution and was often cited by subjects in their free response explanations; it also had a particularly strong effect on predictions about the target stimulus.

Hansen and Lowe (1976), who manipulated consensus by having observers view actors' bogus reactions to musical stimuli, also insured control over expectancies—by pretesting the stimuli and selecting those that had been rated as neutral and equivalent. In this way, observers' tendencies to infer high consensus from their own preferences (Hansen & Donoghue, 1977; Ross et al., 1977) were minimized. Finally, the study by Feldman, Higgins, Karlovac, and Ruble (1976) is particularly informative of this boundary condition for an explicit-consensus effect. Consensus was presented in videotaped sequences of an actor making a choice among items and others either agreeing or disagreeing with that choice. As it turned out, causal attributions were influenced by the manipulation, but only for subjects who did not actually see the stimuli from which the actor's choice was made. Of course, instances where base rates influence target-case prediction only when individuating information is unavailable (Carroll & Siegler, 1977; Kahneman & Tversky, 1973; Zuckerman, 1978a) reflect the same pattern. In fact, both Kahneman and Tversky (1973) and Zuckerman (1978a) found that the impact of base rates on target prediction was consistently greater for subjects who did not read a description of the stimulus person than for those who did.

Factors That Strengthen Consensus

Clearly, most of the recent research effort has been directed toward isolating those conditions that strengthen manipulations of explicit consensus. These factors are reviewed below.

Magnitude of Consensus

How high is high consensus? Because consensus refers to a *proportion*, one obvious prediction is that the stronger the consensus manipulation (e.g., 90%–10% rather than 60%–40%), the greater will be its effect on predictions and causal inferences. In McArthur's (1972) and other similar question-

naire experiments, the actual proportions were not specified. For those presentations of consensus in which subjects directly viewed the behavior of others, the three (Hansen & Lowe, 1976) or four (Feldman et al., 1976) others behaved unanimously, thereby creating small-sample base rates of 100%–25% and 100%–20%, respectively. Only since the Nisbett and Borgida experiments (1975) have attribution researchers begun to present their subjects with more complex base rates composed of larger samples of "others." Wells and Harvey (1977) suggested that Nisbett and Borgida's (1975) base rate manipulations were relatively weak. Experiments were thus conducted in which the same two situations were described but with a lower low consensus and a higher high consensus added. As predicted, the stronger version of consensus had a somewhat greater effect than did the original manipulation (see also Smith, Note 7).

Salience or Availability of the Base Rate

Consensus is not necessarily an abstract form of information. Rather, manipulations of sample-based consensus may vary along a continuum of salience and concreteness. At one extreme, it may be operationalized as a "remote, pallid, and abstract" (Nisbett et al., 1976) summary base rate. At the other extreme, consensus information may be conveyed more concretely—through the live or videotaped actions of others (e.g., Feldman et al., 1976) or through the animated movements of objects (Kassin & Lowe, Note 6). Borgida and Nisbett (1977) have argued convincingly that concrete information is more compelling than similar information presented in abstract form. To the extent that perceivers successively consider possible causes in order of their salience (Smith, Note 7), consensus-based inference may be facilitated by factors that increase the variable's attention value and perhaps its availability for recall (Taylor & Fiske, 1978).

Sequential versus simultaneous presentation. As with other variables in social perception, consensus often unfolds over time. Recall that Feldman et al. (1976) manipulated the consensus of an actor's choice. In a simultaneous-presentation condition, the four other actors then agreed or disagreed with this choice in unison; in a serial-presentation condition, four other actors then agreed or disagreed individually. A Consensus × Mode of Presentation interaction indicated that consensus affected causal attributions more when it was presented sequentially than when it was presented simultaneously. In fact, a comparable study (reported by Manis, 1977) has revealed that subjects' use of base rates for making categorical *predictions* is also increased by a sequential mode of presentation. Feldman et al. (1976) explained this effect as "a tendency for subjects to process simultaneous information as a single bit of information, but process successive information as independent bits of information. Thus, there would be 'more' consensus information processed by subjects in the successive presentation condition" (p. 697). The importance of these findings is underscored by their ecological validity. Since consensus information in "real-world," nonlaboratory settings is often acquired through multiple observations that are separated by time, it is probably a powerful determinant of prediction and attribution.

Order of information presentation. The recognized importance of temporal factors has stimulated a number of interesting hypotheses concerning the *order* of information presentation at three levels: Between consensus and other cues, between the target actor versus others' behavior, and within the others' behavior. At the first level, Ruble and Feldman (1976) varied the order in which consensus, distinctiveness, and consistency were presented in a questionnaire study and found a recency effect—high versus low consensus had the least impact when presented first and the greatest impact when presented last (see also Zuckerman, 1978b). The relative strengths of consensus and distinctiveness may thus be vulnerable to the order of information presentation.

The order in which the target actor versus the others' behavior is viewed is also an important characteristic of consensus when its perception is embedded in time. In the videotapes of Feldman et al. (1976), the target actor always made a choice before rather than after the others did. On the assumption that

observers are attuned to conformity and compliance cues, Kassin (1977) varied the consensus and order of behavior in a contrived aggression experiment and found that an actor's high-consensus aggression was personally attributed less often when he delivered shock after the others did than when he did so before the others. Thus, the order of the target and others' behavior may mediate the consensus effect, at least for behaviors that entail social pressure and when the order is not fixed by social constraints (e.g., "respond when you decide" rather than "respond in order of seating").

Finally at the third level, there is ample evidence from research in impression formation (Jones & Goethals, 1971) to suggest that the order in which the others' behaviors appear might actually elicit differential *estimates* of consensus (e.g., One's estimate that "most actors choose X over 0" could vary if the choices unfold as xxxxx00, 00xxxxx, or xx0xx0x).

Translatability of the base rate. Carroll and Siegler (1977) suggested that perhaps Kahneman and Tversky's (1973) subjects viewed the category base rates as relevant but had difficulty drawing a direct implication from them. Specifically, they noted, "if they attempt to probability match (cf. Weir, 1964), there would be no direct correspondence between a 70%/30% division in the population and how a sample of five people can be divided. Seeing no direct solution to the problem of how to use the base rate information, subjects may then decide to ignore it" (p. 393). Accordingly, Carroll and Siegler (1977) examined the effect of translatability on base-rate utilization.

In that study (reported Experiment 2), half the subjects predicted the occupations of 10 individuals sampled from a population of 20 with a 70%–30% base rate (these figures afforded direct translation into a 7–3 partition for the 10-member sample). The other subjects were presented with base rates that were 75%–25% and hence did not allow for whole number translation to a sample of 10. Results indicated that despite the increased strength of the consensus manipulation in the latter condition, predictions were influenced only by the translatable base rates. A similar effect was reported by Tversky and Kahneman (1974), who labeled this availability heuristic "imaginability." Thus, the ease with which a base rate may be applied to a prediction task affects subjects' use of the information.

Representativeness of the Sample

Of perhaps even greater importance than the salience or availability of explicit consensus is its perceived validity. One severe limitation of explicit consensus is that it is based on the observation of a limited sample. Thus, although the sample base rates should always influence guesses about the behavior of persons from that group, utilization of these data toward prediction of other samples or the population at large is necessarily constrained by the perceived representativeness of the original sample. This limitation was demonstrated in studies showing that base-rate information may be undermined by inferences that the sample is biased (Wells & Harvey, 1977) or small (Kassin, 1979).

Knowledge of randomness. One variable that enhances the meaningfulness of a base rate is knowledge of random selection. Theoretically, consensus utilization requires a belief that the observed sample of actors is representative of the population (see Wells & Harvey, 1978). In line with Kelley's (1967) model, a consensual behavior should be situationally attributed. However, this inference may be undermined by a perceiver's dismissal of the base rate as being attributable to some idiosyncratic characteristics of the observed sample. For example, if all the actors in a group solve a puzzle, an observer may not infer task ease if he or she believes that the individuals are geniuses. The perceiver must know or be able to assume that the actors are representative of the population, and hence of average intellect, in order to make the appropriate inferences.

Wells and Harvey (1977) dealt extensively with this issue. Noting that Nisbett and Borgida's (1975) subjects might have assumed that the base rate sample was biased, Wells and Harvey told half of their subjects that the samples from which the base rates were derived had been randomly selected. Sure

enough, percentage estimates of population performance and causal attributions for the behavior of another were influenced by consensus only for subjects in the knowledge-of-random-sampling condition, a finding corroborated by Hansen and Donoghue (1977). Guesses about the behavior of persons who were in the original sample, however, were affected by consensus with and without the randomness variation. Consistent, then, with the notion that knowledge of randomness increases the generalizability of a base rate to *new* samples, such instructions affected extra-sample prediction (i.e., the prediction of individuals in the general population) but not intrasample guesses (i.e., guessing the behavior of target actors from the original sample).

An intriguing and related side issue concerns the "others" comprising consensus. Kahneman and Tversky (1973) insisted that people are more confident when making inferences from correlated than from independent cues. However, Goethals (1972) found that an actor's confidence is increased more by agreement with a dissimilar other than with a similar other, tentatively suggesting that high consensus is more informative about a stimulus when there is variability in the personal characteristics of the actors in a sample. Finally, note that Borgida (1978) proposed an additional, more concrete method by which to convey representativeness information—to show subjects a few "typical" people from the sample who behaved consensually.

Sample size. Ajzen and Fishbein (1975) suggested that when a proportion is held constant, information gain and hence causal attribution should increase with sample size, that is, with the number of observed events. The fact that "30 out of 40 students failed the exam" should thus provide more information than would "3 out of 4 students failed the exam." Accordingly, Kassin (1979) presented subjects with a description of a helping experiment and two conflicting sets of results—one derived from a sample of 10 actors and the other from a sample of 50. For some subjects, helping was high in base rates for the large sample, (60%, 70%, 80%, *or* 90%) and low in base rate for the small sample (40%, 30%, 20%, *or* 10%, respectively). For others, the base rates were conversely low for the large sample and high in the small sample. Results for both individual and population predictions indicated that although subjects did not make "optimal" use of the information (i.e., a weighted average), they did demonstrate an overall reliance on the larger-sample base rate.

Causal Relevance of Consensus

Base rates with an explanation. On the premise that people possess and often invoke their intuitive theories about the causes of events, Ajzen (1977) proposed that "causality heuristics" mediate the consensus–prediction relationship:

> When asked to make a prediction, people look for factors that would cause the behavior or event under consideration. Information that provides evidence concerning the presence or absence of such causal factors is therefore likely to influence predictions. Other items of information, even though important by the normative principles of statistical prediction, will tend to be neglected if they have no apparent causal significance. Statistical information is used mainly when no causal information is available. (p. 304)

To test this proposition, Ajzen (1977) had subjects predict fictitious students' academic success (i.e., grade point average) from two cues—one that was intuitively causal (IQ or study time) and one that was not (income or distance from campus). In addition, he told subjects that each cue had demonstrated either a strong or weak relationship to the criterion. Results indicated that subjects' predictions were influenced more by causal and empirically strong cues than by noncausal and weak cues, respectively. More important, the causal nature of the information had a greater impact than did its empirical validity. In fact, the causal–weak cue was given greater weight than the noncausal–strong cue (e.g., even when IQ was a weak predictor and distance was strong, IQ weighed more heavily in subjects' decisions). Within the framework of these results, Ajzen explained subjects' underutilization of base rates in Kahneman and Tversky's (1973) lawyer–engineer study by noting that the proportion of lawyers (engi-

neers) did not *cause* any member of the sample to become an engineer or lawyer, whereas the target–case description did provide such causal information.

By way of implication, one way to increase observers' use of base rates is to augment them with a reasonable explanation or "causal theory." Smith (Note 7) presented subjects with the bystander intervention and shock tolerance descriptions employed by Nisbett and Borgida (1975). In addition to the standard high- and no-base-rate groups, others were provided with supplementary explanations (e.g., "74% went all the way to the last shock because they were low in anxiety" or "because they wanted the experimenter to approve of them"). Sure enough, this manipulation increased subjects' use of the consensus toward individual prediction, leading the author to conclude that "the reason becomes salient as a possible cause of behavior" (p. 24). Finally, note that Tversky and Kahneman (1978) have also conceptualized their base-rate results within the context of a causality heuristic.

Actions versus occurrences. After Kruglanski (1975), Zuckerman (1978b) distinguished between behaviors that are under the actor's voluntary control (actions) and those that are not completely voluntary (occurrences). Although occurrences may be attributed to either internal or external factors, actions are by definition caused by internal factors and should not logically be affected by consensus. The reasoning is as follows: Consensus provides information about the provocation or eliciting power of a stimulus (see Anderson, 1974; McArthur, 1976; also see McArthur's, 1972, distinction between manifest and subjective verb categories). Yet if the stimulus is ruled out as a plausible (external) cause, consensus information becomes irrelevant.

To test this hypothesis, Zuckerman (1978b) had subjects read about one of two kinds of events—actions (e.g., "Jerry attended the Sunday meeting") and occurrences (e.g., "Mary passed the exam in history"). These statements were accompanied by high or low consensus, distinctiveness, and consistency information. Consistent with previous research (e.g., McArthur, 1972; Orvis et al., 1975), consensus accounted for the highest percentage of variance in stimulus attribution. More germane to the present discussion, consensus had a substantially larger effect on the attribution of occurrences than on actions. Apparently, the consensus–attribution link is strongest for behaviors that are not completely voluntary.

Summary

In sum, explicit consensus varies along a number of important dimensions that influence its efficacy. The following characteristics of the variable were reviewed: the strength or magnitude of the proportion, the salience of the information and the ease with which it may be applied, the perceived representativeness (and hence generalizability of the base-rate sample), and the causal relevance of the base rate. We turn now to a consideration of the dependent variables.

The Dependent Variables

Prediction

Subjects have been asked to predict the behavior of target members of the original sample, frequencies or individuals in a new population, and/or their own behavior. An important issue here concerns the response criterion by which subjects' performance is evaluated (for a comprehensive discussion of the response criterion controversy, see Borgida, 1978, and Wells & Harvey, 1978). It was noted earlier that the conclusions drawn about base rates and prediction have been based on one of two criteria. The first is the null hypothesis—do the predictions of base rate versus no base rate or high- versus low-base-rate subjects statistically differ at a conventionally acceptable level? If they do differ, one would conclude that consensus affected predictions; if no difference emerged, one would conclude that subjects "ignored" consensus. An alternative strategy is to compare subjects' predictions with those implied by a normative model—do the predictions of base-rate subjects perfectly match those implied by normative standards? If they do, one may

conclude that subjects made perfect utilization of base rates; if not, one would conclude that observers "underutilized" consensus. Kahneman and Tversky's (1973) lawyer–engineer study illustrates the distinction. Recall that their subjects predicted the occupations of individuals taken from a population of 70% or 30% engineers (lawyers). Overall, subjects' mean estimates were 55% in the high engineer condition and 50% in the low engineer condition, a result indicating that although subjects did not ignore the base rate (i.e., the difference between 50% and 55% was statistically significant), they did underutilize it (i.e., 55% compared to 70%).[2] With this distinction in mind, the available research indicates that although observers generally do not ignore consensus information (Nisbett & Borgida, 1975, is an exception), they do not make full utilization of it either (see Wells & Harvey, 1977; 1978).

In general, population predictions are affected by both implicit and explicit consensus, the latter being especially effective when observers are assured that the base-rate sample is representative (Hansen & Donoghue, 1977; Wells & Harvey, 1977) and large (Kassin, 1979), when the base rates are directly translatable into a prediction (Carroll & Siegler, 1977), and when they unfold sequentially (Manis, 1977). The prediction of individual cases is also affected, though not optimally, by base rates (Lowe & Kassin, 1977; Wells & Harvey, 1977; Zuckerman, 1978a), except when relevant personal information (i.e., actor-based expectancies) is also available (Carroll & Siegler, 1977; Kahneman & Tversky, 1973; Nisbett & Borgida, 1975; Zuckerman, 1978a). Finally, note that base rates appear to have a greater impact on observers' discrete predictions (e.g., Category A vs. Category B) than on subjective probabilities associated with these outcomes (Manis, Dovalina, Avis, & Cardoze, in press).

Causal Attribution

A number of important differences between predictions and attributions deserve mention. Whereas the prediction task requires direct translation from base rates, causal attribution demands that more complex inferences be made (Nisbett & Borgida, 1975). Also, in contrast to prediction data that may be compared to those implied by a formal model, "there is no generally agreed upon criterion for explanatory adequacy" (Fischhoff, 1976, p. 439). The null hypothesis (e.g., consensus vs. no consensus) rather than optimality thus serves as the unambiguous criterion by which to judge causal attributions. Finally, in contrast to the assessment of predictions, which is straightforward and operationally consistent across studies, there is considerable disagreement over how to operationalize attribution measures that are far less sensitive and reliable.

Nisbett and Bellows (1977), among others, have argued that people infer causality from their normative expectations. This relationship has been established, even in cases where base-rate information is also available (e.g., Hansen & Donoghue, 1977; Lowe & Kassin, 1977). On the other hand, observers use explicit consensus toward causal attribution only under certain well-defined conditions—when prior expectancies are neutralized and/or when the base rate is strong, salient, representative of the population, and causally relevant. Note also that sample base rates do not affect personal and situational attribution equally. Instead, they generally provide more information about the stimulus than they do about the person (DiVitto & McArthur, 1978; Garland et al., 1975; McArthur, 1976; Orvis et al., 1975; Zuckerman, 1978b; Zuckerman & Mann, in press; Vestewig, Note 2). The actual behavior of actors thus provides information about the power and characteristics of the focal stimulus or general situation (see

---

[2] Actually, Wells and Harvey (1978) are quick to point out that "proper utilization" in Kahneman and Tversky's (1973) study does not necessarily imply target case predictions of 70% and 30%. Their subjects had two sources of information—the personality sketches and the category base rates. From a Bayesian perspective, if an individual case description is both accurate and diagnostic, it can serve to overcome the prior odds and reduce subjects' reliance on the 70–30 probabilities. Consequently, a number of possible predictions could be acceptable within the normative model.

also the discussion by DiVitto & McArthur, 1978).

One final limitation within the domain of causal attribution is notable. That is, although consensus is informative for validating the causal categories of person versus situation, it is not necessarily of interest to observers who are evaluating alternative categories (e.g., ability versus effort, intentional versus unintentional). In short, the utility of consensus is restricted to a narrow range of deductively relevant causal categories (Kruglanski, Hamel, Maides, & Schwartz, 1978).

Trait Inference

A number of investigators have assessed subjects' inferences about the characteristics of the target actor and stimulus. Trait inference should follow from causal attribution such that actor-caused behaviors should yield more extreme and more confident trait ratings than situationally caused events would. As with causal attribution, dispositional inferences about the actor are often based on a comparison of his or her behavior with the *expected* norm (Ajzen, 1971; Lay et al., 1973; Lowe & Kassin, 1977; Ross et al., 1977). In other words, the actor's characteristics are inferred from his or her unexpected (i.e., low implicit consensus) behavior.

The effect of sample-based consensus on trait inferences has only recently received empirical attention. Nisbett and Borgida (1975) reported that inferences about actors' traits were unaffected by consensus. Wells and Harvey (1977), however, demonstrated that base rates do affect these inferences when the traits being assessed are relevant to the target behavior. Having read the bystander intervention study, their base-rate subjects rated nonhelping actors on trait dimensions, whereas their no-base-rate subjects evaluated the relatedness between the target behavior and each dimension. Of nine traits seen as unrelated to the target behavior, none was affected by the base rates. Yet, three out of four "relevant" traits were influenced by the consensus manipulation. Similarly, Hansen and Stonner (1978) found that consensus affects inferences about stimulus characteristics (e.g., task difficulty), but only when the base rates and attributes are perceived to be causally related.

Summary and Conclusions

The present review suggests that any conclusions concerning the effects of consensus must be qualified—by delineating what type of consensus, what type of effect, and compared to what response criterion. Normative expectancies that are based on self-observation (i.e., the false-consensus effect) or previous other-observation (i.e., Jones & McGillis's, 1976, consensus variable) provide an initial basis for prediction and a frame of reference within which causes are inferred. These expectancies may be of a general nature or they may be refined by a perceiver's knowledge about the (class of) target actor and surrounding behavioral situation. Explicit, sample-based consensus (i.e., Kelley's, 1967, conception of a covariation across actors) is similarly employed, but under somewhat restrictive circumstances. That is, base rates appear to be relatively ineffective when they are either too inconsistent or are redundant with normative expectations. In some studies, the power of consensus has thus been demonstrated only when actor- or situation-based expectancies have been neutralized. In other studies, consensus utilization has been increased by varying characteristics of the experimental manipulation such as the magnitude of the consensus proportion, the salience of the base rate (sequential vs. simultaneous presentation, order of information presentation) and the ease with which it may be applied (translatability), the perceived representativeness of the base-rate sample (knowledge of random selection, sample size), and the causal relevance of the base rate (base rates with an explanation, actions vs. occurrences).

On the dependent variable side, consensus information is employed differently toward the various tasks of prediction, causal attribution, and trait inference. Generally stated, observers appear to underutilize but not to ignore base rates when making categorical and behavioral predictions. This conclusion is tempered, of course, by the host of factors that strengthen the presentation of base rates.

For making causal attribution and trait inferences, consensus seems to have a greater impact on stimulus than on person inferences. Finally, only traits and characteristics perceived as relevant to the target behavior are influenced by consensus information.

To date, a wide variety of issues surrounding the consensus controversy have been addressed. These include actor–observer differences in base-rate utilization, developmental trends in the use of consensus, the relative efficacy of consensus and distinctiveness, concrete versus abstract information, response criteria in social judgment research, and so on. What has emerged from this recent surge in research is an empirical, though somewhat atheoretical, redefinition of the boundary conditions for a consensus effect. Yet two important issues remain largely unresolved. First, how do normative expectancies develop and on what informational characteristics are they based? The literature suggests that implicit consensus is a viable determinant of observers' causal beliefs. Nevertheless, a theoretical account of these expectancies, one that would enable us to predict their direction and magnitude, is currently lacking. A second important issue demanding theoretical integration concerns the relationship between implicit and explicit consensus. In most life situations, people observe base rates for events that are high or low in their subjective probabilities of occurrence. Although the available data indicate that base rates are more effective when information about the target actor or situation is absent, nobody has tested the logical hypothesis that the degree of discrepancy between implicit and explicit consensus mediates utilization of the latter. At this point, attention might fruitfully be directed toward examining this relationship.

Reference Notes

1. Eisen, S. V. *Actor-observer differences in causal attribution and consensus, distinctiveness, and consistency information.* Paper presented at the meeting of the Eastern Psychological Association, Boston, April 1977.
2. Vestewig, R. *Self-reported information processing differences and causal inference.* Paper presented at the meeting of the Eastern Psychological Association, Boston, April 1977.
3. Dix, T. H., Herzberger, S., & Erlebacher, A. *Using consensus information: Developmental study of Kelley's attribution theory.* Paper presented at the meeting of the American Psychological Association, Toronto, Canada, September 1978.
4. Karlovac, M., Feldman, N. S., Higgins, E. T., & Ruble, D. N. *The effect of consensus information on causal attributions.* Paper presented at the meeting of the American Psychological Association, Washington, D.C., September 1976.
5. Swann, W. B., & Collins, W. A. *When persons become causes: Age and consensus as moderators of causal attribution.* Unpublished manuscript, 1978.
6. Kassin, S. M., & Lowe, C. A. *Kelley's attribution theory: A perceptual approach.* Paper presented at the meeting of the Eastern Psychological Association, Philadelphia, April 1979.
7. Smith, E. R. *Consensus, prediction, and explanation: Toward a model of the attribution process.* Unpublished manuscript, 1977.

References

Ajzen, I. Attribution of dispositions to an actor: Effects of perceived decision freedom and behavioral utilities. *Journal of Personality and Social Psychology,* 1971, *18,* 144–156.

Ajzen, I. Intuitive theories of events and the effects of base-rate information on prediction. *Journal of Personality and Social Psychology,* 1977, *35,* 303–314.

Ajzen, I., & Fishbein, M. A Bayesian analysis of attribution processes. *Psychological Bulletin,* 1975, *82,* 261–277.

Anderson, N. H. Cognitive algebra: Integration theory applied to social attribution. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 7). New York: Academic Press, 1974.

Borgida, E. Scientific deduction—Evidence is not necessarily informative: A reply to Wells and Harvey. *Journal of Personality and Social Psychology,* 1978, *36,* 477–482.

Borgida, E., & Nisbett, R. E. The differential impact of abstract vs. concrete information on decisions. *Journal of Applied Social Psychology,* 1977, *7,* 258–271.

Carroll, J. S., & Siegler, R. S. Strategies for the use of base rate information. *Organizational Behavior and Human Performance,* 1977, *19,* 392–402.

Darley, J. M., & Latané, B. Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology,* 1968, *8,* 377–383.

DiVitto, B., & McArthur, L. Z. Developmental differences in the use of distinctiveness, consensus and consistency information for making causal attributions. *Developmental Psychology,* 1978, *14,* 474–482.

Feldman, N. S., Higgins, E. T., Karlovac, M., & Ruble, D. N. Use of consensus information in causal attributions as a function of temporal presenta-

tion and availability of direct information. *Journal of Personality and Social Psychology,* 1976, *34,* 694–698.

Fischhoff, B. Attribution theory and judgment under uncertainty. In J. H. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research* (Vol. 1). Hillsdale, N.J.: Erlbaum, 1976.

Frieze, I., & Weiner, B. Cue utilization and attributional judgments for success and failure. *Journal of Personality,* 1971, *39,* 591–605.

Garland, H., Hardy, A., & Stephenson, L. Information search as affected by attribution type and response category. *Personality and Social Psychology Bulletin,* 1975, *1,* 612–615.

Goethals, G. R. Consensus and modality in the attribution process: The role of similarity and information. *Journal of Personality and Social Psychology,* 1972, *21,* 84–92.

Hansen, R. D. Actors rely on distinctiveness and observers rely on consensus: Two causal schemata. *Personality and Social Psychology Bulletin,* 1976, *2,* 242–247.

Hansen, R. D., & Donoghue, J. M. The power of consensus: Information derived from one's own behavior. *Journal of Personality and Social Psychology,* 1977, *35,* 294–302.

Hansen, R. D., & Lowe, C. A. Distinctiveness and consensus: The influence of behavioral information on actors' and observers' attributions. *Journal of Personality and Social Psychology,* 1976, *34,* 425–433.

Hansen, R. D., & Stonner, D. M. Attributes and attributions: Inferring stimulus properties, actor's dispositions, and causes. *Journal of Personality and Social Psychology,* 1978, *36,* 657–667.

Heider, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

Heider, F., & Simmel, M. An experimental study of apparent behavior. *American Journal of Psychology,* 1944, *57,* 243–259.

Herzberger, S. D. The attributor's unprompted use of strength and direction of consensus information. *Social Behavior and Personality,* in press.

Jones, E. E., & Davis, K. E. From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2). New York: Academic Press, 1965.

Jones, E. E., Davis, K. E., & Gergen, K. J. Role playing variations and their informational value for person perception. *Journal of Abnormal and Social Psychology,* 1961, *63,* 302–310.

Jones, E. E., & Goethals, G. R. *Order effects in impression formation: Attribution context and the nature of the entity.* Morristown, N.J.: General Learning Press, 1971.

Jones, E. E., Kanouse, D. E., Kelley, H. H., Nisbett, R. E., Valines, S., & Weiner, B. (Eds.). *Attribution: Perceiving the causes of behavior.* Morristown, N.J.: General Learning Press, 1972.

Jones, E. E., & McGillis, D. Correspondent inferences and the attribution cube: A comparative reappraisal. In J. H. Harvey, W. J. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research* (Vol. 1). Hillsdale, N.J.: Erlbaum, 1976.

Kahneman, D., & Tversky, A. Subjective probability: A judgment of representativeness. *Cognitive Psychology,* 1972, *3,* 430–454.

Kahneman, D., & Tversky, A. On the psychology of prediction. *Psychological Review,* 1973, *80,* 237–251.

Kassin, S. M. Order effects in consensus information and causal attributions. *Psychological Reports,* 1977, *41,* 927–930.

Kassin, S. M. Base rates and prediction: The role of sample size. *Personality and Social Psychology Bulletin,* 1979, *5,* 210–213.

Kassin, S. M., & Lowe, C. A. On the use of single-sentence descriptions of behavior in attribution research. *Social Behavior and Personality,* in press.

Kelley, H. H. Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 15). Lincoln: University of Nebraska Press, 1967.

Kelley, H. H. The processes of causal attribution. *American Psychologist,* 1973, *28,* 107–128.

Kruglanski, A. W. The endogenous–exogenous partition in attribution theory. *Psychological Review,* 1975, *82,* 387–406.

Kruglanski, A. W., Hamel, I. Z., Maides, S. A., & Schwartz, J. M. Attribution theory as a special case of lay epistemology. In J. H. Harvey, W. Ickes, & R. F. Kidd (Eds.), *New directions in attribution research* (Vol. 2). Hillsdale, N.J.: Erlbaum, 1978.

Lay, C. H., Burron, B. F., & Jackson, D. N. Base rates and informational value in impression formation. *Journal of Personality and Social Psychology,* 1973, *28,* 390–395.

Lowe, C. A., & Kassin, S. M. On the use of consensus: Prediction, attribution, and evaluation. *Personality and Social Psychology Bulletin,* 1977, *3,* 616–619.

Manis, M. Cognitive social psychology. *Personality and Social Psychology Bulletin,* 1977, *3,* 550–566.

Manis, M., Dovalina, I., Avis, N. E., & Cardoze, S. Base rates can affect individual predictions. *Journal of Personality and Social Psychology,* in press.

McArthur, L. A. The how and what of why: Some determinants and consequences of causal attribution. *Journal of Personality and Social Psychology,* 1972, *22,* 171–193.

McArthur, L. Z. The lesser influence of consensus than distinctiveness information on causal attributions: A test of the person–thing hypothesis. *Journal of Personality and Social Psychology,* 1976, *33,* 733–742.

Milgram, S. Behavioral study of obedience. *Journal of Abnormal and Social Psychology,* 1963, *67,* 371–378.

Miller, A. G., Gillen, B., Schenker, C., & Radlove, S. The prediction and perception of obedience to authority. *Journal of Personality,* 1974, *42,* 23–42.

Nisbett, R. E., & Bellows, N. Verbal reports about

causal influences on social judgments: Private access versus public theories. *Journal of Personality and Social Psychology*, 1977, *35*, 613–624.

Nisbett, R. E., & Borgida, E. Attribution and the psychology of prediction. *Journal of Personality and Social Psychology*, 1975, *32*, 932–943.

Nisbett, R. E., Borgida, E., Crandall, R., & Reed, H. Popular induction: Information is not necessarily informative. In J. Carroll and J. Payne (Eds.), *Cognition and social behavior*. Hillsdale, N.J.: Erlbaum, 1976.

Nisbett, R. E., & Schachter, S. Cognitive manipulation of pain. *Journal of Experimental Social Psychology*, 1966, *2*, 227–236.

Orvis, B. R., Cunningham, J. D., & Kelley, H. H. A closer examination of causal inference: The roles of consensus, distinctiveness, and consistency information. *Journal of Personality and Social Psychology*, 1975, *32*, 605–616.

Ross, L. The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10). New York: Academic Press, 1977.

Ross, L., Greene, D., & House, P. The "false consensus effect": An egocentric bias in social perception and attribution processes. *Journal of Experimental Social Psychology*, 1977, *13*, 279–301.

Ruble, D. N., & Feldman, N. S. Order of consensus, distinctiveness, and consistency information, and causal attributions. *Journal of Personality and Social Psychology*, 1976, *34*, 930–937.

Taylor, S. E., & Fiske, S. T. Salience, attention, and attribution: Top of the head phenomena. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. New York: Academic Press, 1978.

Tversky, A., & Kahneman, D. Judgment under uncertainty: Heuristics and biases. *Science*, 1974, *185*, 1124–1131.

Tversky, A., & Kahneman, D. Causal thinking in judgment under certainty. In B. Butts & J. Hintikka (Eds.), *Logic, methodology, and philosophy of science*. Dordrecht, Holland: D. Reidel, 1978.

Valins, S., & Nisbett, R. E. *Attribution processes in the development and treatment of emotional disorders*. Morristown, N.J.: General Learning Press, 1971.

Weiner, B. *Achievement motivation and attribution theory*. Morristown, N.J.: General Learning Press, 1974.

Weiner, B., Frieze, I., Kukla, A., Reed, L., Rest, S., & Rosenbaum, R. M. *Perceiving the causes of success and failure*. Morristown, N.J.: General Learning Press, 1971.

Wells, G. L., & Harvey, J. H. Do people use consensus information in making causal attributions? *Journal of Personality and Social Psychology*, 1977, *35*, 279–293.

Wells, G. L., & Harvey, J. H. Naive attributors' attributions and predictions: What is informative and when is an effect an effect? *Journal of Personality and Social Psychology*, 1978, *36*, 483–490.

Zuckerman, M. Use of consensus information in prediction of behavior. *Journal of Experimental Social Psychology*, 1978, *14*, 163–171. (a)

Zuckerman, M. Actions and occurrences in Kelley's cube. *Journal of Personality and Social Psychology*, 1978, *36*, 647–656. (b)

Zuckerman, M., & Mann, R. W. The other way around: Effects of causal attributions on estimates of consensus, distinctiveness, and consistency. *Journal of Experimental Social Psychology*, in press.

Received August 30, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 11, 1982-1992

Learned Helplessness and Learned Effectiveness: Effects of Explicit Response Cues on Individuals Differing in Personal Control Expectancies

W. Larry Gregory, George M. Chartier, and Mark H. Wright
Arizona State University

Female undergraduates ($n = 62$) who scored as extreme internals or externals on the Mirels Personal Fate Control Scale participated in a partial replication of Hiroto's learned helplessness experiment. Lights were added to the treatment apparatus, which made explicit to subjects the contingency or noncontingency between their responses and the termination of an aversive tone. As predicted, the performance of internals was significantly impaired by uncontrollability (learned helplessness), while that of externals was facilitated by controllability (learned effectiveness). Externals performed as well as internals in the "escapable" condition, but their performance was inferior to that of internals in the control condition. Following "inescapable" treatment, internals performed worse than externals. These results are supportive of Lefcourt's theory of cue explication. Implications for locus of control and learned helplessness research are discussed.

The behavioral consequences of both the experimental induction of learned helplessness (Abramson, Seligman, & Teasdale, 1978; Maier & Seligman, 1976) and pre-extant, generalized perceptions of externality (Lefcourt, 1976; Phares, 1976) have been examined in a study by Hiroto (1974). Hiroto used an instrumental learned helplessness induction procedure that has since become common: Subjects, told there is something they can do to terminate an aversive tone, are seated before a small box on which there is a button. In the "escapable" treatment condition, pressing the button terminates the tone. In the "inescapable" condition, nothing the subject does terminates the tone. Rather, it is terminated automatically. Hiroto concluded that the effects of inescapable treatment parallel those of pre-extant externality: Subjects treated with inescapable noise performed significantly more poorly on a subsequent soluble task than did subjects not treated with inescapable noise; and the performance of external locus of control subjects was inferior to that of internal locus of control subjects across all experimental conditions.

Although it is typically found that internals outperform externals at tasks when instructions emphasizing skill are used (Lefcourt, Lewis, & Silverman, 1968; Rotter & Mulry, 1965; Watson & Baumal, 1967), Hiroto's (1974) results indicating that the performance of internals was superior to that of externals in the inescapable condition are puzzling when juxtaposed with other locus of control studies. These studies have shown that internals adjust their aspirations downward following failure, whereas externals are likely to display atypical aspiration shifts (Feather, 1968; Lefcourt, 1967; Lefcourt & Ladwig,

The research reported in this article was supported by National Institute of Mental Health Grant 1 RO3 MH30431-01 to the second author. The authors gratefully acknowledge the assistance of Debbie Doller and Linda Schmeider, who served as research assistants. Special appreciation is extended to Sanford Braver and Dennis Young for their continued statistical guidance.

Requests for reprints should be sent to W. Larry Gregory, Department of Psychology, Arizona State University, Tempe, Arizona 85281 until January 7, 1980. After that date, requests for reprints should be sent to him at the Department of Psychology, New Mexico State University, Las Cruces, New Mexico 88003.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3711-1982$00.75

1965), that internals indicate that they have no control over a situation after failing to avoid an aversive outcome (Gregory, 1978), that internals perform worse than externals on a digits-backwards task when they expect to receive uncontrollable shocks (Houston, 1972), and that internals perform worse than externals on an anagram task following failure (Garrett & Willoughby, 1972). It would seem, then, that internals are often more adversely affected by negative experience than are externals, although Hiroto's findings suggest the opposite.

One possible explanation for these apparently discrepant results might lie in a seemingly minor procedural element: Hiroto (1974) did not utilize lights or other adjunctive feedback stimuli on the treatment apparatus to make explicit to the subjects the contingency or noncontingency between the subjects' button-pushing responses and the termination of the aversive tone. The use of these lights has since become standard during the treatment phase of learned helplessness experiments when instrumental learning is involved (Hiroto & Seligman, 1975; Miller & Seligman, 1975; Price, Tryon, & Raps, 1978; Sacco & Hokanson, 1978). Hence, during the inescapable treatment, it may be argued that some of the subjects pushed the response button just prior to or simultaneously with the automatic termination of the tone. Since there were no response contingency lights to inform a subject as to whether he or she had terminated the tone or whether it had gone off automatically, some subjects might have perceived a relationship between their response and the termination of the tone. Having experienced partial reinforcement, these subjects would be more likely to persist during the second phase of the experiment. Since Hiroto found that internals made significantly more responses than externals did during the inescapable treatment (*Mdn*s = 21 and 5.5, respectively), this scenario is more likely to have occurred for the internal subject than for the external subject. A test of this explanation would require only a partial replication of the Hiroto procedure with the addition of response contingency lights to the treatment apparatus. If this explanation is valid, then Hiroto's finding should be reversed: For subjects in the inescapable condition, the performance of internals should be inferior to that of externals.

But what effect should the addition of the response contingency lights have on the behavior of internals and externals in the control and escapable conditions? The performance of control condition subjects should be unaffected, since there is no treatment and the response contingency lights are inapplicable. Thus internals should again outperform externals, as in Hiroto's (1974) study.

Different predictions can be made for the escapable condition. A number of studies have shown that explicit task cues have a positive effect on the behavior of externals (Dollinger & Taub, 1977; Gregory & Nelson, 1978; Lefcourt, 1967; Lefcourt & Wine, 1969; Taub & Dollinger, 1975). These studies have demonstrated that providing reward or purpose for a task raises the motivational level of externals to equal that of internals or results in externals making aspiration shifts more typical of internals following a success or a failure. Lefcourt (1967) has labeled this process "cue explication" and has restricted its meaning to making the external subject "aware of the availability of reinforcements, and perhaps, of the methods for maximizing his chances of succeeding in given tasks" (p. 378).[1] In a learned helplessness experiment, a response contingency light would make ex-

[1] One study (Wolk & DuCette, 1974, Study 2) involving cue explication has failed to find that it alters the performance of externals to equal that of internals. Under high cue explication conditions, Wolk and DuCette found that externals failed to display intentional or incidental learning at a level equal to that of internals. But it is questionable whether the high cue explication manipulation met the conditions prescribed by Lefcourt, since it consisted of informing subjects that they would be tested on some aspect of the task in addition to taking a test already specified. This does not make explicit the reinforcements available for a good task performance. This procedure could have produced debilitating anxiety or test anxiety, which are associated with externality (Feather, 1967; Ray & Katahn, 1968; Watson & Baumal, 1967). Supporting this argument, Wolk and DuCette found that high cue explication externals performed worse than low cue explication externals in their high difficulty condition.

plicit to the external subject that he or she had just been reinforced for a correct task performance. Whereas Hiroto (1974) found internals to exhibit superior performance in the escapable condition, the cue explication literature would suggest that the addition of the response contingency lights in the escapable treatment condition would enhance the performance of externals to equal that of internals.

Predictions

It was predicted that the addition of response contingency lights to the treatment apparatus for an instrumental learned helplessness induction procedure would result in (a) no differences in performance between internals and externals in the escapable condition, (b) internals outperforming externals in the control condition, and (c) externals outperforming internals in the inescapable condition. For external subjects only, the theory of cue explication predicts that (d) the performance of subjects in the escapable condition should be enhanced above that of subjects in the control condition. Comparative performance among internals would be unaffected by explicit response cues, and they should display the typical learned helplessness pattern: (e) Inescapable treatment should produce performance decrements when compared to the control condition, whereas (f) the control condition should not differ from the escapable treatment condition.

Method

At Arizona State University, students enrolled in introductory psychology must write three brief research reports or participate in 3 hours of laboratory research. Students participating in this experiment had chosen the latter option.

Subjects were 62 females who scored at the extremes (0–2 or 7–9) of the Mirels (1970) personal fate control scale. This was a subsample of 183 females to whom the Rotter (1966) Internal–External Locus of Control Scale was administered immediately prior to the experiment. Subjects were assigned randomly to either the escapable treatment, control, or inescapable treatment conditions. The relatively greater frequency of internals and the method of immediate assignment to condition resulted in unequal numbers of subjects per cell of the 2 × 3 factorial design. Although Hiroto (1974) used subjects scoring at the extremes of the DeKalb Survey Tests (Student Opinion Survey, Form I-E, 1; see James, Note 1), the Mirels scale was employed here, since it is assumed to be more predictive of control over personal events and it is derived from the more commonly used Rotter I-E scale. The James and Rotter scales are highly similar in content. Only female subjects were used, due to their availability in the subject pool.

Apparatus

The treatment apparatus, adapted from Hiroto and Seligman (1975), consisted of a spring-loaded button housed in the center of a circular disc, 30.5 cm in diameter and 7.6 cm high, with two circular lights, one red and one green, each 1.3 cm in diameter, located on either side of the button. In the escapable condition two button presses terminated the aversive tone and stopped a remote timer. In the inescapable condition the button was electronically rendered inoperative.

The test phase apparatus was a modification of Turner and Solomon's (1962) human shuttle box. By sliding a knob to either extreme end contact could be made with a hidden microswitch that controlled tone termination. Specifications of the manipulandum are included in Hiroto (1974).

The aversive stimulus was a 3000 Hz tone generated by a Hewlett Packard audio oscillator, presented to the subject through Realistic Nova 10 stereo headphones at 100 db. Formal instructions were presented to the subject on typewritten index cards. All response latencies were measured by a standard (1/100 sec) automatic timer, and intertrial intervals (ITI) were timed manually on a stopwatch. The oscillator, timer, and controlling circuitry were located in an adjacent room separated by a one-way mirror.

Procedure

Treatment. Following administration of the I-E scale, each subject was escorted individually to the experimental room and was informed that the experiment concerned the effects of noise pollution on the performance of simple tasks and that the subject's continued participation in the study beyond a brief sample of the unpleasant but harmless tone was optional. If, after a 3 sec sample of the 3000 Hz tone, the subject agreed to continue, the appropriate treatment phase was begun. Three subjects refused to participate (one internal and two moderates).

Subjects in the escapable and inescapable conditions were seated at a table on which the treatment apparatus was located. The experimenter uncovered the apparatus and presented instructions to the subject concerning the noise, that there was something she could do to stop it, and the function of the two lights. These instructions were identical to those of Hiroto and Seligman (1975), with the addition of a statement reiterating the function of the two lights: "*You* will have successfully stopped the noise *only* when the green light comes on."

Forty unsignaled trials with an ITI range of 15 to 25 sec and a mean of 17.4 sec followed. In the escapable treatment, pressing the button twice within 5 sec after tone onset terminated the tone, and the green light flashed on briefly. Response latencies were recorded. If the correct response did not occur within 5 sec of tone onset, the tone terminated automatically, the red light flashed, and a response latency of 5 sec was recorded.

In the inescapable treatment condition, exposure to tone duration and pattern was yoked to the subject's counterpart in the escapable condition and was independent of the button. The tone was not escapable, and the red light flashed at the termination of each trial.

Control subjects were not exposed to the treatment apparatus but worked on a questionnaire composed of 100 items selected from the California Psychological Inventory on the basis of relatively neutral content, during which they received yoked exposure to the series of 40 tones. Subjects in this condition received instructions on completing the inventory and were told, "From time to time while you are filling out the questionnaire a noise will come on. Please continue working on the questionnaire."[2]

Test task. Immediately after treatment, subjects in all three groups were seated at a table on which the covered test task apparatus was located and were given instructions concerning it identical to those of Hiroto and Seligman (1975). These instructions were similar to those for the first task, but with no references to lights, since they were not present on the test apparatus.

The test phase consisted of 20 signaled 10 sec escape–avoidance trials. A 5 sec amber light, located at the midpoint of the shuttlebox cover, preceded the 3000 Hz, 100 db tone, with the light's termination coinciding with the onset of the tone. The ITI ranged from 10 to 25 sec, with a mean of 15 sec.

Prior to the first trial the sliding knob was always located at the midpoint of the channel. The appropriate response was moving the knob to one side of the shuttlebox on one trial, moving it to the opposite side on the next, and alternating for the remaining trials. Although the subject was informed only of the possibility of escape, tone avoidance was possible by making the appropriate response after the onset of the warning light and before the onset of the tone. A successful avoidance response (i.e., a response latency of less than 5 sec.) terminated the warning light and prevented exposure to the tone. A successful escape response (i.e., a response latency between 5 and 10 sec) terminated the tone. If the subject failed to avoid or to escape, the tone was terminated automatically and a response latency of 10 sec was recorded.

Dependent Measures

Five test task-dependent measures were used to assess performance. These were: (a) the "conditional probability" of an escape or avoidance response, defined as the percentage of times a successful response (escape or avoidance) was preceded by a successful response; (b) the number of failures to avoid or escape, defined as the number of responses with latencies of 10 sec; (c) success criterion, defined as the number of trials needed to reach either an avoidance or an escape criterion of three consecutive escape or avoidance responses (subjects who failed to reach either were assigned scores of 21); (d) the mean overall response latency for the 20 trials; and (e) the number of trials needed to reach an avoidance criterion of three consecutive avoidance responses. Although partially redundant with (c), (e) was included as an exploratory measure to determine if any differences existed between internals and externals in their acquisition of avoidance responses. Latency was included as a dependent variable, since previous studies have utilized it, although Buchwald, Coyne, and Cole (1978) argue that it may be a poor measure of motivation in a learned helplessness experiment. At least two learned helplessness studies (Douglas & Anisman, 1975; Tennen & Eller, 1977) have failed to find significant main effects for this variable. The other three dependent variables have all been used previously in learned helplessness research. The conditional probability measure is assumed to assess cognitive learning ability, whereas the others index motivational processes (Klein & Seligman, 1976).

Additionally, six questionnaire measures were included at the end of the experiment; these required the subject to rate her treatment experiences on 7-point bipolar scales. These were: (a) her motivation to solve the problem; (b) the extent to which she felt in control of her success and failure; (c) the level of her frustration; (d) the level of her irritation; (e) the level of her depression; and (f) the aversiveness of the tone. The second of these measures was intended as a manipulation check; the others were included as crude exploratory measures of possible differences in subjective state.

Results

Self-Report Measures

A 2 (Internal vs. External Personal Control Orientation) × 3 (Escapable, Control, or Inescapable Treatment Condition) analysis of variance (ANOVA) was performed on each of the self-report measures. Only one significant

[2] At least two other studies (Gatchell & Proctor, 1976; Hiroto & Seligman, 1975) have employed a control group that was told that noise was inevitable. Neither study found interference effects in the control group subjects. Apparently, informing subjects not to try to do anything about the noise precludes formation of perceptions of helplessness (Buchwald, Coyne, & Cole, 1978, p. 184).

effect was revealed. A main effect for treatment was found for subjects' perceptions of control, $F(2, 56) = 3.55$, $p < .035$. Simple effects analysis revealed that subjects in the inescapable condition perceived less personal control than did control condition subjects, $F(1, 56) = 6.23$, $p < .016$, whereas control condition subjects did not differ from escapable condition subjects, $F(1, 56) < 1$, *ns*. This suggests that the experimental manipulation of perceived control over events was successful.

Random Assignment of Subjects

To insure that the assignment of subjects to conditions had been random, the same 2 × 3 ANOVA was performed on subjects' personal control scores. Neither the interaction nor the main effect for treatment condition were significant, both Fs < 1. A main effect for personal control orientation was found, $F(1, 56) = 1248.18$, $p < .001$, indicating that subjects in the internal conditions were significantly more internal than subjects in the external conditions.

Multivariate Clustering

Since the conditional probability measure was assumed to assess cognitive learning ability, it was analyzed apart from the other dependent variables, which were assumed to assess motivation. Also, since the avoidance criterion variable was included as an exploratory measure, and because the use of response latencies has been questioned previously, these variables, along with the number of failures to escape or avoid and the success criterion variables, were subjected to a test of the assumption of homogeneity of the variance–covariance matrices (Bock, 1975, p. 412). This was done in order to determine the appropriateness of analyzing these variables as a single multivariate vector. This test indicated that the four variables together violated the assumption, $\chi_c^2(50) = 104.83$, $p < .001$. Accordingly, a principal components analysis with iterations utilizing the varimax rotation procedure (Nie, Hull, Jenkins, Steinbrenner, & Bent, 1975) was undertaken to determine an appropriate clustering of the four variables. The principal components analysis resulted in two unrotated factors with eigenvalues greater than 1. The remaining two unrotated factors had eigenvalues of .15 or less and accounted for less than 5% of the variance. The first rotated factor accounted for 72.5% of the variance and consisted of the number of failures to avoid or escape and the success criterion variables. The second rotated factor accounted for 27.5% of the variance and consisted of the avoidance criterion and response latency variables. Thus, it was determined that separate multivariate analyses of variance would be performed on the variables represented by these two factors. All subsequent analyses utilized the MANOVA program (Clyde, 1969), and planned comparisons were employed as justified by the specific predictions of the study (Hays, 1973, p. 582; Winer, 1971, pp. 175, 384). The MANOVA program uses least squares analyses when performing univariate analyses of variance.

Performance of Internal Versus External Subjects

Planned comparisons were used to contrast internals to externals in each of the three treatment conditions. The planned comparison F value for each dependent variable is displayed in Table 1. The means for all cells on the performance measures are displayed in Table 2. Figure 1 displays graphically the results of the conditional probability measure, typical of the overall pattern of responses.

Escapable condition. As can be seen from Table 2, there were no significant differences between internals and externals on any of the performance measures in the escapable condition. This supports the prediction that externals perform as well as internals when given explicit response cues.

Control condition. As predicted, when there was no prior experience with explicit response cues, as in the control condition, internals outperformed externals on the conditional probability, number of failures, and

Table 1
Planned Comparison F *Values for the Multivariate and Univariate Contrasts*

Planned comparison and condition	Multivariate contrast *F*		Univariate contrast *F*				
	NF & SC	RL & AC	CP	NF	SC	RL	AC
Internal vs. external subjects							
Escapable	<1	<1	<1	<1	<1	<1.05	<1.15
Control	2.51*	1.53	5.04**	4.27**	4.81**	3.02*	1.82
Inescapable	2.58*	<1	4.10**	4.98**	2.08	<1	<1
Internal subjects							
Escapable vs. control	<1	7.04***	<1	1.13	<1	2.41	9.53***
Inescapable vs. control	10.05***	4.43**	19.28***	20.35***	11.26***	8.57***	4.77**
External subjects							
Escapable vs. control	2.39*	3.44**	5.05**	4.74**	4.48**	1.45	<1
Inescapable vs. control	<1	<1	<1	<1	<1	<1	<1

Note. The degrees of freedom for all multivariate contrasts are 2 and 55, whereas they are 1 and 56 for all univariate contrasts. CP = conditional probability measure. NF = number of failures to escape or avoid measure. SC = success criterion measure. RL = response latency measure. AC = avoidance criterion measure.
* $p < .10$. ** $p < .05$. *** $p < .01$.

success criterion measures.[3] No significant differences were obtained for the response latency and avoidance criterion measures, although the means were in the predicted direction.

Inescapable condition. As predicted for the inescapable condition, where there were explicit cues indicating that subjects failed to control the outcome, internals performed significantly worse than externals did on the conditional probability and number of failures measures. Differences on the other measures were in the predicted direction but were nonsignificant.

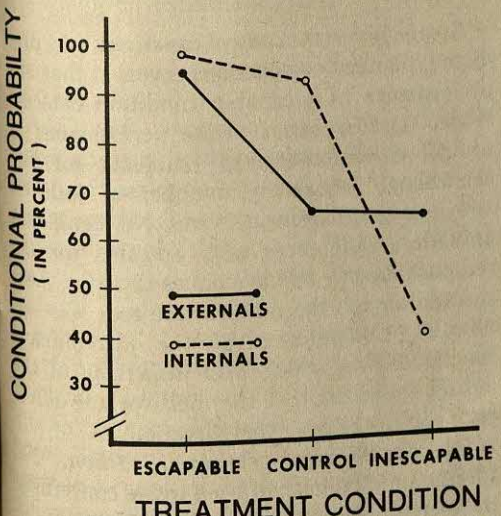


Figure 1. Conditional probability by treatment condition.

Performance of Internals Across Treatment Conditions

Escapable versus control condition. It was also predicted that planned comparisons would not reveal significant differences between escapable condition–internal subjects and control condition–internal subjects. With the exception of the avoidance criterion measure, no significant effects were obtained. The multivariate comparison for the combined avoidance criterion and response latency mea-

[3] Healy (1969) has noted that since univariate confidence intervals are located along a linear axis, whereas bivariate confidence intervals are elliptical, it is possible to reject both univariate null hypotheses while failing to reject the bivariate null hypothesis. This is known as Rao's paradox, and Baron (1978) has published results on its occurrence for a trivariate case.

Table 2
Means for Each Performance Measure

Measure	Treatment condition: Escapable *M*	Control *M*	Inescapable *M*
Internal orientation			
Conditional probability	97.8	93.0	39.1
Number of failures	0.3	1.3	11.0
Success criterion	3.4	4.2	12.0
Response latency	5.9	4.5	7.2
Avoidance criterion	19.4	12.3	17.4
External orientation			
Conditional probability	94.6	65.4	63.8
Number of failures	0.8	5.7	6.2
Success criterion	4.0	9.3	8.7
Response latency	5.0	6.2	6.4
Avoidance criterion	16.6	16.1	16.7

Note. With the exception of the conditional probability measure, low numbers represent better performance. In the internal personal control orientation, escapable $n = 17$, control $n = 12$, and inescapable $n = 9$. In the external personal control orientation, escapable $n = 8$, control $n = 7$, and inescapable $n = 9$.

sures was significant, and inspection of the means indicates that internals in the control condition, contrary to prediction, were superior in performance to the internals in the escapable condition. A brief comment may be in order concerning this unexpected finding. For the other measures of cognitive learning and motivation, explicit response cues indicating that internals were successful in escaping the aversive stimuli did not serve to enhance their subsequent test task performance above that of control condition–internal subjects. However, the explicit cues may have led internals to overlearn an escape response to the extent that avoidance learning on the subsequent task was debilitated. This may occur only under conditions of explicit response cues, but further research is indicated on this topic. Given success, internals may be prone to stereotypic responses and fail to explore the possibilities of even more beneficial behaviors.

Inescapable versus control condition. As predicted, planned comparisons revealed that the performance of inescapable condition–internal subjects was significantly worse than that of control condition–internal subjects for all measures.

Summary of response pattern for internal subjects. Overall, internals displayed the predicted learned helplessness pattern. That is, the performance of internal subjects in the inescapable condition was inferior to that of internals in the control condition. With the exception of the avoidance criterion measure, the performance of internal subjects in the escapable condition was not different from that of internals in the control condition. For the avoidance criterion measure, control condition–internal subjects outperformed escapable condition–internal subjects.

It appears that an explicit response cue indicating that they are escaping an aversive stimulus does not serve to enhance the performance of internal subjects above that of control condition–internal subjects, although it can serve to interfere with subsequent avoidance learning. By contrast, an explicit response cue indicating that they are failing to escape an aversive stimulus does serve to debilitate the performance of internals in comparison to that of control condition–internal subjects.

Performance of Externals Across Treatment Conditions

Escapable versus control condition. As predicted, planned comparisons revealed that the performance of escapable condition–external subjects was superior to the performance of control condition–external subjects for the conditional probability, number of failures, and success criterion measures. No significant univariate differences were revealed for the response latency and avoidance criterion measures, although the multivariate root was significant. Computation of the discriminant function scores, along with inspection of the means, indicated that the multivariate difference was in the predicted direction.

Inescapable versus control condition. No specific predictions had been made concerning the performance of external subjects in the inescapable condition relative to the performance of external subjects in the control con-

dition. To remain consistent with the previous analyses, however, planned comparisons between these groups were conducted. No significant differences were obtained on any of the performance measures. Since planned comparisons are an extremely powerful technique for detecting differences, should they exist, the failure to obtain any significant differences for any of the univariate or multivariate contrasts should underscore the apparent lack of differences between the performances of external subjects in the inescapable and control conditions.

Summary of response pattern for external subjects. It appears that externals did not display learned helplessness, since there were no differences between the control condition and the inescapable condition for any of the dependent measures. As predicted, the performance of externals in the escapable condition was superior to that of externals in the control condition for most of the performance measures. It would appear that an explicit response cue indicating that they are escaping an aversive stimulus does serve to enhance the subsequent performance of external subjects above that of control condition–external subjects. By contrast, an explicit cue indicating that they are failing to escape an aversive stimulus does not serve to debilitate the performance of externals in comparison to control condition–external subjects.

Discussion

Additional conclusions regarding cue explication and the performances of internals and externals would be possible, had the present design included nonexplicated escapable and inescapable conditions, as in Hiroto's (1974) study. Nonetheless, the cue explication hypothesis (Lefcourt, 1967) received substantive support. Following escapable treatment on an apparatus equipped with response cue lights, the performance of externals was significantly enhanced above that of control condition–external subjects. Additionally, these escapable condition–external subjects equalled in performance the escapable condition–internal subjects. It appears that making explicit to the external subject the contingency between response and outcome is one means of cue explication. The finding that inescapable condition–internal subjects performed worse than both control condition–internal subjects and inescapable condition–external subjects when explicit response cues were provided suggests that cue explication may have a negative counterpart for internal subjects. When internals are made aware of the lack of availability of reinforcements, or the lack of contingency between behavior and outcome, their subsequent performance is debilitated to a level inferior to that of externals. This point suggests a potential for behavioral problems in persons with extremely internal control orientations.

Previous researchers (DuCette & Wolk, 1973) have argued that internals are cognitively superior to externals because they take fewer trials to criterion on a skill task. In the present study, internals and externals in the escapable condition did not differ in their subsequent acquisition of the escape-avoidance behavior. Thus, the provision of explicit cue responses can result in externals equalling internals in learning ability. It is important to note that this effect was obtained on a subsequent (albeit highly similar) task that did not utilize explicit cues. This indicates that the motivational and cognitive effects of cue explication during the treatment phase generalized to the second task. Unlike much of the locus of control research that attempts to delineate differences between internals and externals, this study seems to refocus attention on the neglected issue of alleviating or altering the behavioral consequences of an external perception of control. The present results are even more dramatic, considering that they were obtained from individuals with extreme external control orientations. Whether our findings are generalizable to extreme males as well as females is of course an empirical question, although the lack of consistent sex differences in the internal–external literature would suggest such applicability.

The present results also lead us to consider some possible antecedents and consequences of an external expectancy for control. Given that the performance of externals in the control condition was inferior to that of externals

in the escapable condition, it may be argued, as did Hiroto (1974), that one consequence of an external orientation is some degree of pre-extant helplessness for skill tasks. Since externals appeared to be helpless prior to entering the experimental situation, their altered performance in the escapable condition might best be termed "learned effectiveness." Further, since there were no differences in performance between control condition–externals and inescapable condition–externals, it is clear that the pre-extant helplessness of externals in the inescapable condition was not increased by the treatment. Since helplessness is produced (in instrumental induction studies) by failure to control an aversive experience, we propose that one developmental antecedent of externality is repeated failure to control aversive experiences. However, this speculation on externality is qualified by Gregory's (1978) findings that the Rotter (1966) I–E scale and the Mirels (1970) personal fate control scale are predictive only of control over negative outcomes or aversive events. Externality for nonaversive events or positive outcomes would, of course, be expected to follow from different antecedents.

Locus of Control, Cue Explication, and Learned Effectiveness

The present findings suggest a reinterpretation of another study involving locus of control and learned helplessness. Following a pretreatment consisting of contingent or noncontingent feedback on a concept formation task represented as a test of "spatial reasoning ability," Cohen, Rothbart, and Phillips (1976) found what they interpreted as learned helplessness for external subjects on both a Stroop color task and a maze tracing task. A performance decrement for internal subjects was found only on the maze tracing task. In the contingent condition, subjects received accurate feedback concerning the correctness of their choice after each concept formation slide. Cue explication theory would predict that external subjects would perform as well in this condition as internal subjects, as indeed was found. Although the performance of external subjects in the contingent condition was superior to that of external subjects in the noncontingent condition, one cannot determine whether learned helplessness or learned effectiveness occurred. As Buchwald, Coyne, and Cole (1978) note, in the absence of a control condition it is impossible to know whether performance enhancement (or learned effectiveness, as we have termed it) or interference has occurred. Given the procedure employed by Cohen et al. in the contingent condition, it is possible that their external subjects displayed learned effectiveness and, as in the present study, their internal subjects displayed learned helplessness. The finding that internal subjects displayed performance decrements on only the maze tracing task may have been due to its greater similarity to the training task.

Finally, it should be noted that one unpublished study (Benson & Kennelly, Note 2) has found results identical to the present study. Benson and Kennelly treated male and female subjects who had scored at the extremes of Levenson's (1974) locus of control scale with explicit contingent or noncontingent feedback on a concept formation task. Using anagram problems as the test task, they found that relative to control condition subjects, "the performance of internals was impaired by pretreatment with insoluble problems. The performance of externals was not impaired by pretreatment with insoluble problems, but facilitated by pretreatment with soluble problems (p. 7)." Combined, the results of Cohen et al. (1976) and Benson and Kennelly extend and generalize the notion that cue explication can improve significantly the performance of externals.

Conclusions

The present study has implications for research in two areas. The examination of learned helplessness might include systematic investigation of the effects that seemingly minor procedural variations have on the behaviors of subjects with individual differences, such as locus of control. Locus of control research could benefit by studying the effects of explicit cues on the behavior of internal and external locus of control subjects, and from

refocusing some attention on this inadequately researched area.

Reference Notes

1. James, W. H. *Dekalb survey test: Student opinion survey—Form I-E, 1.* Copyright by the author, 1963.
2. Benson, J. S., & Kennelly, K. *A re-examination of the relationship between locus of control and learned helplessness.* Paper presented at the meeting of the Southwestern Psychological Association, Albuquerque, April 1976.

References

Abramson, L. Y., Seligman, M. E. P., & Teasdale, J. D. Learned helplessness in humans: Critique and reformulation. *Journal of Abnormal Psychology,* 1978, *87,* 49–74.

Baron, R. A. Aggression-inhibiting influence of sexual humor. *Journal of Personality and Social Psychology,* 1978, *36,* 189–197.

Bock, R. D. *Multivariate statistical methods in behavioral research.* New York: McGraw-Hill, 1975.

Buchwald, A. M., Coyne, J. C., & Cole, C. S. A critical evaluation of the learned helplessness model of depression. *Journal of Abnormal Psychology,* 1978, *87,* 180–193.

Clyde, D. J. *Multivariate analysis of variance on large computers.* Miami, Fla.: Clyde Computing Service, 1969.

Cohen, S., Rothbart, M., & Phillips, S. Locus of control and the generality of learned helplessness in humans. *Journal of Personality and Social Psychology,* 1976, *34,* 1049–1056.

Dollinger, S. J., & Taub, S. I. The interaction of locus of control expectancies and providing purpose on children's motivation. *Journal of Research in Personality,* 1977, *11,* 118–127.

Douglas, D., & Anisman, H. Helplessness or expectation incongruency: Effects of aversive stimulation on subsequent performance. *Journal of Experimental Psychology: Human Perception and Performance,* 1975, *1,* 411–417.

DuCette, J., & Wolk, S. Cognitive and motivational correlates of generalized expectancies for control. *Journal of Personality and Social Psychology,* 1973, *26,* 420–426.

Feather, N. T. Some personality correlates of external control. *Australian Journal of Psychology,* 1967, *19,* 253–260.

Feather, N. T. Change in confidence following success and failure as a predictor of subsequent performance. *Journal of Personality and Social Psychology,* 1968, *9,* 38–46.

Garrett, A. M., & Willoughby, R. H. Personal orientation and reactions to success and failure in urban black children. *Developmental Psychology,* 1972, *7,* 92.

Gatchell, R., & Proctor, J. Physiological correlates of learned helplessness in man. *Journal of Abnormal Psychology,* 1976, *85,* 24–34.

Gregory, W. L. Locus of control for positive and negative outcomes. *Journal of Personality and Social Psychology,* 1978, *36,* 840–849.

Gregory, W. L., & Nelson II, D. V. Locus of control and cue explication. *Personality and Social Psychology Bulletin,* 1978, *4,* 248–251.

Hays, W. L. *Statistics for the social sciences.* New York: Holt, Rinehart & Winston, 1973.

Healy, M. J. R. Rao's paradox concerning multivariate tests of significance. *Biometrics,* 1969, *25,* 411–413.

Hiroto, D. S. Locus of control and learned helplessness. *Journal of Experimental Psychology,* 1974, *102,* 187–193.

Hiroto, D. S., & Seligman, M. E. P. Generality of learned helplessness in man. *Journal of Personality and Social Psychology,* 1975, *31,* 311–327.

Houston, B. K. Control over stress, locus of control, and response to stress. *Journal of Personality and Social Psychology,* 1972, *21,* 249–255.

Klein, D. C., & Seligman, M. E. P. Reversal of performance deficits and perceptual deficits in learned helplessness and depression. *Journal of Abnormal Psychology,* 1976, *85,* 11–26.

Lefcourt, H. M. Effects of cue explication upon persons maintaining external control expectancies. *Journal of Personality and Social Psychology,* 1967, *5,* 372–378.

Lefcourt, H. M. *Locus of control: Current trends in theory and research.* Hillsdale, N.J.: Erlbaum, 1976.

Lefcourt, H. M., & Ladwig, G. W. The effect of reference group upon Negroes' task persistance in a biracial competitive game. *Journal of Personality and Social Psychology,* 1965, *1,* 668–671.

Lefcourt, H. M., Lewis, L., & Silverman, I. W. Internal versus external control of reinforcement and attention in a decision making task. *Journal of Personality,* 1968, *36,* 663–682.

Lefcourt, H. M., & Wine, J. Internal versus external control of reinforcement and the deployment of attention in experimental situations. *Canadian Journal of Behavioral Science,* 1969, *1,* 167–181.

Levenson, H. Activism and powerful others. Distinctions within the concept of internal–external control. *Journal of Personality Assessment,* 1974, *38,* 377–383.

Maier, S. F., & Seligman, M. E. P. Learned helplessness: Theory and evidence. *Journal of Experimental Psychology: General,* 1976, *105,* 3–46.

Miller, W. R., & Seligman, M. E. P. Depression and learned helplessness in man. *Journal of Abnormal Psychology,* 1975, *84,* 228–238.

Mirels, H. L. Dimensions of internal versus external control. *Journal of Consulting and Clinical Psychology,* 1970, *34,* 226–228.

Nie, N. H., Hull, C. H., Jenkins, J. G., Steinbrenner, K., & Bent, D. H. *Statistical package for the social sciences.* New York: McGraw-Hill, 1975.

Phares, E. J. *Locus of control in personality*. Morristown, N.J.: General Learning Press, 1976.

Price, K. P., Tryon, W. W., & Raps, C. S. Learned helplessness and depression in a clinical population: A test of two behavioral hypotheses. *Journal of Abnormal Psychology*, 1978, *87*, 113–121.

Ray, W. J., & Katahn, M. Relation of anxiety to locus of control. *Psychological Reports*, 1968, *23*, 1196.

Rotter, J. B. Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs*, 1966, *80*(1, Whole No. 609).

Rotter, J. B., & Mulry, R. C. Internal versus external control of reinforcement and decision time. *Journal of Personality and Social Psychology*, 1965, *2*, 598–604.

Sacco, W. P., & Hokanson, J. E. Expectations of success and anagram performance of depressives in a public and private setting. *Journal of Abnormal Psychology*, 1978, *87*, 122–130.

Taub, S. I., & Dollinger, S. J. Reward and purpose as incentives for children differing in locus of control expectancies. *Journal of Personality*, 1975, *43*, 179–195.

Tennen, H., & Eller, S. J. Attributional components of learned helplessness and facilitation. *Journal of Personality and Social Psychology*, 1977, *35*, 265–271.

Turner, L., & Solomon, R. L. Human traumatic avoidance learning: Theory and experiments on the operant-respondent distinction and failures to learn. *Psychological Monographs*, 1962, *76*(1, Whole No. 559).

Watson, D., & Baumal, E. Effects of locus of control and expectation of future control upon present performance. *Journal of Personality and Social Psychology*, 1967, *6*, 212–215.

Winer, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1971.

Wolk, S., & DuCette, J. Intentional performance and incidental learning as a function of personality and task dimensions. *Journal of Personality and Social Psychology*, 1974, *29*, 90–101.

Received August 30, 1978 ■

New JPSP Format Delayed Until April

Recent issues of this journal carried an announcement that the *Journal of Personality and Social Psychology* would appear in January 1980 under new policies and in a new sectioned format. However, in order to publish the backlog of manuscripts accepted under present policies, the first three issues of 1980 will continue to carry articles accepted under Clyde Hendrick's Acting Editorship.

This journal will appear in April as a sectioned journal edited by Melvin Manis, Ivan D. Steiner, and Robert Hogan, as previously announced.

Robert S. Daniel, Chair, APA
Publications and Communications Board

Journal of Personality and Social Psychology
1979, Vol. 37, No. 11, 1993-2002

An Empirical Analysis of the Correlations Between Leadership Status and Participation Rates Reported in the Literature

R. Timothy Stein and Tamar Heller
University of Illinois at Chicago Circle

Many studies of small group leadership have reported a positive relationship between leadership status and the rate of verbal participation. The leadership participation relationship has been attributed to (a) the performance of task leadership behaviors, (b) task skill and knowledge, (c) social status characteristics, and (d) the presence of observers. Hypotheses based on these four explanations were generated and tested using 72 correlations reported in the literature as data points. The relationship of a large number of situational variables to the leadership–participation rate relationship was also examined. In a multiple regression analysis, four variables accounted for 63% of the variance in the leadership–participation rate correlations. The first three explanations were supported. They can be integrated if the moderating effect of task competence and social status on the performance of task leadership behavior is recognized.

A positive relationship between verbal participation rates and leadership status is frequently reported in the small group literature. In reviewing that literature, the authors uncovered 77 correlations between the variables. The mean was .65, and the range was from −.48 (Gustafson & Harrell, 1970) to .98 (Bates, 1952); only 4 were negative. In these studies new groups were formed for the research or for classroom activities. Either observers or the group members rated the participants on leadership. These ratings were measures of leadership status because all the group members, and not just those who emerged to a leadership position, were included. Measures of the relative amounts of talking by the group members were obtained from audiotapes or from observers' records (e.g., the number of words spoken).

Other lab studies have shown a covariation between participation and leadership status. Several authors report higher mean leadership ratings for high participants (e.g., Carter, Haythorn, Shriver, & Lanzetta, 1950; Gintner & Lindskold, 1975). Also, experimenters have manipulated the relative amounts of talking by group members. Parallel shifts in peer judgments about the participants' leadership followed (e.g., Bavelas, Hastorf, Gross, & Kite, 1965; Zdep & Oakes, 1967).

The covariation of leadership status and participation rates has been attributed to (a) the performance of task roles, (b) task skill and knowledge, (c) social status characteristics, (d) the presence of observers, and (e) motivation. Although an extensive literature points to a relationship between leadership and participation, the research described here is the first to compare the explanations to determine how well they account for the rela-

The authors thank Steven Cooley for rating the tasks and the leadership functions, Herbert Stenson for his advice on the statistical analyses, Carolyn Clark for secretarial support, Joyce Rozwadowski for her assistance with literary revisions, and Mary Glenn Wiley and L. Richard Hoffman for their comments on a draft of the paper. The authors acknowledge the use of the library and computer facility at the University of Illinois at Chicago Circle and the library at DePaul University.

A summary of this research was presented to the 86th annual convention of the American Psychological Association in Toronto (Heller and Stein, 1978).

Requests for reprints should be sent to Tamar Heller, Psychology Department, University of Illinois at Chicago Circle, P.O. Box 4348, Chicago, Illinois 60680.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3711-1993$00.75

tionship. Hypotheses were generated from the first four explanations.[1] Additional hypotheses were formulated to include two situational parameters, group size and observer bias. Seventy-two of the correlations reported in the literature were used as data.

First, it is possible that the role performed by leaders requires their participation and wins them status. The most common division of roles in small group interaction is the distinction between task and group maintenance leadership (Hare, 1976, p. 145). Task leadership behaviors are defined as "behaviors of the individual related to his efforts to assist the group in achieving goals toward which the group is oriented" (Bales, 1958). Group maintenance leadership behaviors are "behaviors of the individual related to his efforts to establish and maintain cordial and socially satisfying relations with other group members" (Bales, 1958). The development of task leadership roles is the major thesis of emergent leadership theories (Bales, 1953; Hollander, 1958; Stein, Hoffman, Cooley, & Pearse, 1979; Stogdill, 1959). Studies that have compared leaders to nonleaders in the performance of specific behaviors have clearly indicated that leaders perform more task-related behaviors than do nonleaders (e.g., Carter et al., 1950; Kirscht, Lodahl, & Haire, 1959; Morris & Hackman, 1969). In summarizing this research, Stein and Heller (Note 1) noted that leaders were significantly more active than nonleaders in performing task leadership behaviors in 51 of 76 (67%) statistical tests.[2] The task leadership activities included problem identification, proposing solutions, seeking and giving information and opinions, and initiating structure. The performance of group maintenance activities differentiated leaders from nonleaders in only 5 of 24 tests. The following hypotheses are based on the task facilitation explanation: *Hypothesis 1a: Leadership–participation rate correlations in which task leadership measures are used are higher than those involving maintenance functions. Hypothesis 1b: The more general the task leadership measure, the higher is the correlation with participation rates.*

Several researchers have suggested that the leadership–participation rate relationship might be a reflection of technical competence (e.g., Gintner & Lindskold, 1975). Task skill has been shown to be related to emergent leadership in both field (Clifford & Cohn, 1964; Crockett, 1955) and lab studies (e.g., Heinicke & Bales, 1953; Moment & Zaleznik, 1963). Hemphill (1961) found that members of new groups who believed that they were more competent than others performed a higher number of leadership behaviors. In some studies, the quality and quantity of participation were manipulated (Gintner & Lindskold, 1975; Jaffee & Lucas, 1969; Reilly & Jaffee, 1970; Sorrentino & Boutillier, 1975). Even when confederates suggested the same solution to the assigned problem in different groups, their leadership ratings were dependent upon how much they talked. Thus, the quality of the solution could not account for the leadership–participation rate relationship. However, it is possible that both leadership status and participation rate differences occur because those who offer better solutions make more attempts to influence others. In that case, competency may be a basis for both influence and participation in tasks requiring skill. *Hypothesis 2: The value of the correlation between leadership status and participation is directly related to the skill requirements of the task.*

Third, Blau (1964) has theorized that members who impress others as being capable of providing task assistance are initially permitted more participation time. The personal characteristics that have been tied to these judgments include intelligence, personality traits, and social status (race, sex, and income). Of these variables, only sexual status

---

[1] A hypothesis was not formulated for the motivation explanation because data concerning individual differences on either motivation or the raters' perceptions of motivation were not reported with the leadership–participation rate correlations in the literature.

[2] Some researchers compared leaders to nonleaders on the proportion of their total behavior that fell within certain task behavior categories. In these tests, leaders were found to have a significantly higher proportion in only 5 of 40 (12.5%) comparisons (Stein & Heller, Note 1). These results indicate that leadership status is based on the number of task leadership acts, and not the proportion of one's behavior that is task leadership.

seems to be strongly related to both participation rates and leadership. Comparisons of sexes have shown men to be more verbose and more influential (e.g., Strodtbeck, James, & Hawkins, 1957), and more proactively involved in performing the task (e.g., Strodtbeck & Mann, 1956).

Hypothesis 3: The correlation between leadership status and participation is higher in mixed-sex than in same-sex groups.

Finally, Wilson (1971) has suggested that the leadership–participation rate relationship is an experimental artifact. Members with high assertiveness and ascendancy needs are thought to respond to the presence of an observer with high participation and ingratiating behaviors. The other members rate these people high on leadership because they assume that the observer identifies leadership with dominance, influence, and high participation rates.

Hypothesis 4: The correlation between leadership and participation is lower in groups unaware of being observed.

Hypotheses were formulated for two situational variables, group size and observer bias. Other independent variables were included for exploratory research. Several researchers have found that the differences in participation rates among group members are dependent upon group size (see Hare, 1976, pp. 219–221). After reviewing several studies, Thomas and Fink (1963) concluded that greater size was associated with an increase in organization and division of labor. These findings suggest that (*Hypothesis 5*) *the correlation between participation rates and leadership status is directly related to group size.*

Greater observer dependency on participation rates in formulating leadership judgments has been suggested in three studies (Juola, 1957; Kanungo, 1966; Stein, 1977). Observers, but not the group members, failed to differentiate between task and maintenance functions, which suggests that the observers used a more global criterion in evaluating leadership.

If the observers are so biased, then (*Hypothesis 6*) *the correlations between leadership and participation involving observer leadership ratings are higher than those based on peer ratings or self-ratings.*

Table 1
Correlations of Leadership Status and Participation Rates

Study	Correlations
Bass, 1949	.93, .86
Bass, Wurster, Doll, & Clair, 1953	.77, .60
Bates, 1952	.85
Berkowitz, 1956	.34, .48
Borgatta, 1954	.55, .35, .21, .71, .48, .25, .22, .16
Borgatta & Bales, 1956	.53, −.31, .18, .79, .79, .35, .43, .34, .55
Burke, 1974	.81
Burroughs & Jaffe, 1969	.39, .45, .13, .30, .24
Gustafson & Harrell, 1970	.77, .80, .50, .92, .88, .40, .68, .78, −.02, .66, .67, −.48, .82, .80, .87, −.17, .51, .71, .70, .34, .56, .77, .89, .47, .83, .79, .78, .05
Heinicke, 1950	.91
Jaffe & Lucas, 1969	.63
Juola, 1957	.89, .81
Morris & Hackman, 1969	.46, .65, .52
Slater, 1955	.80, .75, .38, .48, .51, .10
Strodtbeck & Hook, 1961	.69

Method

Dependent Measure

Seventy-two correlations (Pearson or Spearman), reported in the literature between emergent leadership status and the verbal participation rates of group members were used as data for the dependent variable (see Table 1).[3] The use of more than one correlation from a study reflects differences in tasks, group composition, the degree of group consensus on who was leader, the length of interaction covered, or the leadership or participation measures used. Three of the values (from Morris & Hackman, 1969) include both emergent and appointed leaders. No other data of this nature were found for formal leaders. The authors transformed the correlations to Fisher *z* scores for the statistical tests.

The leadership–participation rate correlations are

[3] Four correlations were excluded because of insufficient information about the task and subject population (.96, French, 1950; reported in Bass, 1954) and/or because an interaction time was used that far exceeded that of the other studies (.95 and .94, Norfleet, 1948; and .98, Bates, 1952). The .89 value reported by Stang, Castellaneta, Constantindis and Fortuno (1976) came to the authors' attention too late to be included in the analysis.

not equally stable. The *n*s for the correlations differ, and in some instances median rhos were used. However, the differences in error variance due to different sample sizes are randomly distributed; sample size was uncorrelated with the *z* transformations of the leadership–participation rate correlations ($r = -.12$, *ns*). For this reason, the values did not need to be weighted.

Independent Measures

The study included 28 independent measures related to (a) the type of task, (b) the characteristics of the task, (c) the nature of the group, and (d) the type of leadership measure. The data provided on these variables summarize the methodological and measurement choices made by past researchers. From this summary, possible limitations on the generalizability of the leadership–participation rate relationship can be identified.

Task type. All the correlations are from groups assigned a task. The tasks were categorized according to four types: (a) problem solving, deriving a group solution to a problem ($n = 16$); (b) verbal production, preparing verbal presentations or creating stories ($n = 20$); (c) opinion, discussing issues without having to reach a group decision ($n = 34$); and (d) manipulation, assemblying objects ($n = 2$).

Task characteristics. The task dimensions included the six identified by Shaw (1971, 1973) and three additional measures: (a) skill required, requirements for special training or knowledge; (b) difficulty, including time pressures, skill or knowledge required, number of operations, knowledge of the desired outcome, amount of effort needed, and the degree of the subjects' familiarity with the task; (c) goal clarity; (d) solution multiplicity, the number of possible solutions and the extent to which the adequacy of a solution could be verified; (e) cooperation requirements, requirements for integrated group action; (f) manipulation–intellectual requirements, intellectual operations versus physical operations; (g) population familiarity, the occurrence of the task in the larger society; (h) intrinsic interest, interest created by the nature of the task; and (i) external motivation, consequences of performance (e.g., money, grades, the judgment of esteemed others). Eight-point rating scales were used (after Shaw, 1973). Higher scale values indicate a higher level of the dimension in question (e.g., a more difficult task).

Nature of groups. In order to identify different types of groups, five dichotomous variables were set up. For the first, a score of 1 was assigned if all the group members were males, and 0 for groups of other compositions. In a similar fashion, variables were defined for all female members, members of both sexes, college students, and group members who were aware that they were being observed. Two additional variables were the number of group members, and the group members' knowledge of one another prior to the meeting. A 3-point scale was used for the last variable.

Leadership measures. The characteristics of the leadership measures were assessed with eight variables. One variable indicated (a) whether the measure was of task leadership or maintenance leadership.[4] Other dichotomous variables indicated whether or not the leadership ratings were made by (b) peers or (c) observers, or (d) if they were self-ratings. Two time measures were used: (e) the number of minutes the group had met before the first leadership measure was taken and (f) the number of minutes before the last leadership measure. (g) Whether or not the leadership rating was meaningful to the group members, that is, if the selected leader actually functioned in any capacity, was noted. (h) Finally, the breadth of the leadership measure (i.e., how many different types of behavior it included) was rated on an 8-point scale.

Interrater Reliabilities

Two judges (including the first author) rated the leadership functions and descriptions of the task without knowledge of the correlations between participation and leadership. The Pearson correlations between the judges' ratings ranged from .79 to .97 (see Table 2).

Results

The relationships between the dependent and independent variables are presented in Tables 2 and 3. If a rating scale was used for an independent variable, the scores were correlated with the Fisher *z* transformations of the leadership–participation correlations (see Table 2). If a variable was dichotomous, *t* tests were used. The means are given in Table 3. The correlations that were assigned a score of 1, because the condition indicated by the variable name was fulfilled (e.g., correlations from groups of all males), are called "included cases." The correlations for which the condition was not fulfilled (e.g., correlations from groups other than all male) are labeled as "excluded cases."

---

[4] Some experiments used vague measures, such as "Who was the real leader of the group?" (Morris & Hackman, 1969). However, those authors and others (e.g., Bales, 1953) have shown that subjects name task leaders when asked in a general way to identify their leaders. For this reason, the general leadership measures were included with the task leadership measures. The data provide support for this decision. The mean of the 13 leadership–participation rate correlations in which general leadership measures were used was .74. It was not significantly different from the mean of .67 for correlations involving specific measures of task leadership.

Table 2
Situational Factors: Interrater Reliabilities and Correlations With the Degree of Association Between Participation Rates and Leadership

Variable	Interrater reliability[a]	Correlation	
		All values[b]	Task subsample[c]
Task characteristics			
Skill requirements	.94	.27*	.35**
Task difficulty	.97	.11	.07
Goal clarity	.97	.12	.07
Solution multiplicity	.79	.05	.17
Intrinsic interest	.91	.34**	.54**
External motivation	.97	.24**	.21
Cooperation requirements	.95	.06	.01
Population familiarity	.94	−.07	−.15
Manipulative-intellectual	.95	.19	.40**
Nature of the group			
Knowledge of group members	.87	.11	.22
Group size	—	.39**	.45**
Leadership rating			
Time before first rating	—	.25*	.32*
Total time of session	—	.19	.24
Breadth of leadership measure	.95	.13	.36**

[a] All interrater reliabilities are significant at the .001 level, $df = 20$. [b] $df = 70$. [c] $df = 55$.
* $p < .05$. ** $p < .01$.

Because the first hypothesis stated that the leadership–participation rate relationship was limited to measures of task leadership, all of the analyses were performed on both the total sample of 72 correlations and on a subsample of all the correlations that involved task leadership measures.

Tests of Hypotheses: Theoretical Explanations

Leadership behavior. The data for Hypotheses 1a and 1b support the contention that the performance of task leadership behaviors is a common source of both high participation and leadership status. The mean value of the leadership–participation rate correlations with task leadership measures was .69, whereas the mean for the maintenance measures was only .16, $t(70) = 6.37$, $p < .001$.

The means for subcategories of functions were as follows (the first three categories include maintenance leadership): (a) .017 for negative social interaction, (b) .178 for general maintenance measures (e.g., liking), (c) .180 for positive social interaction, (d) .390 for seeking information and structure, (e) .615 for initiating structure and procedure, (f) .722 for giving information and opinions, (g) .735 for problem identification and solving, (h) .736 for the general measures of leadership, and (i) .841 for general measures of task leadership.

As hypothesized (1b), the association between leadership and participation was significantly correlated with ratings of the breadth of the task leadership measures, $r = .36$, $p < .01$. However, the correlation was only .13 when the entire sample of 72 values was used. This lower value is consistent with the notion that the maintenance behaviors are not the source of the leadership–participation rate relationship.

Task skill. A positive relationship was predicted (Hypothesis 2) between task skill requirements and the association between participation and leadership. The data are supportive. The variables were correlated .27 ($p < .01$) for the total sample and .35 ($p < .01$) for the task subsample.

Sexual status. A significantly higher leadership–participation correlation mean was found for mixed-sex groups than for same-sex

Table 3
Means and t *Values for Dichotomous Variables*

Variable	Included cases		Excluded cases	
	M	*n*	*M*	*t* score
Total sample	.604	72	—	—
Nature of the group				
Male	.505	30	.665	2.29*
Female	.466	8	.619	1.27
Mixed sex	.702	34	.497	3.18**
College	.657	52	.437	2.75**
Aware of observer	.643	26	.581	.88
Leadership measure				
Task (vs. maintenance)	.687	57	.156	6.37***
Peer	.609	58	.582	.30
Observer	.672	14	.586	1.30
Self	.581	9	.607	.25
Future significance	.369	6	.622	1.75

Note. All *t* values were calculated with pooled error terms and with 70 degrees of freedom. The means are correlational values, not Fisher *z* scores.
* $p < .05$. ** $p < .01$. *** $p < .001$.

groups in the total sample, $t(70) = 3.18$, $p < .01$, and the subsample of task measures, $t(55) = 5.68$, $p < .001$. The means for male and female groups were almost identical (see Table 3). Unfortunately, the mixed-sex variable is confounded with type of task. Thirty-one of the 34 correlations for mixed-sex groups were from groups assigned opinion tasks. These 31 correlations constitute most of the data points (34) for opinion groups. Consequently, it is impossible to know which of the two variables is creating the significant result. The confound remained when the maintenance leadership measures were removed.

Demand characteristics. The leadership–participation correlations based on data from subjects aware of being observed were not significantly higher than those from unaware subjects for the total sample, $t(70) = .88$, or for the subsample, $t(55) = .61$.

Tests of the Hypotheses: Situational Parameters

Group size. A higher association between leadership status and participation was predicted for larger groups (Hypothesis 5). The group size ranged from 3 to 18 members, with a mean of 4.8. The moderate correlations between group size and the dependent measure for both the total (.39) and subsample (.45) were significant ($p < .01$).

Observer ratings. The hypothesis (6) that leadership–participation correlations would be higher when based on observer ratings than when based on peer ratings or self-ratings was not supported. Although the means were in the predicted direction, the difference was not significant for either the total sample, $t(70) = 1.30$, or for the subsample, $t(55) = 1.70$.

Exploratory Analyses

As mentioned above, a large number of independent variables, related to the groups, their tasks, and the leadership measures used, were included for exploratory analyses. The authors recognize that because of the nature of such procedures, the findings may be due to chance and may not be real effects. For this reason, these results are presented only as possible relationships that merit further investigation.

Differences due to the type of task were examined by a one-way analysis of variance. Although the resulting F for the entire sample was not significant, $F(3, 86) = 2.02$, $p = .12$, the value for the task subsample was $F(3, 53) = 5.92$, $p < .001$. Using a Duncan multiple range, the mean for opinion groups was found to be higher than the other means (problem solving, verbal production, and manipulation). As indicated above, the groups assigned opinion tasks were predominately mixed-sex groups. Consequently, either sexual composition or the type of task could be creating the results.

The two measures of motivation (intrinsic interest and external motivation) and the extent to which the task was intellectual were correlated with the degree of association between leadership status and participation (see Table 2).

The interchange before the first leadership measure ranged from 1 minute to 4 hours with a mean of 42 minutes. The time prior to the last rating varied from 5 minutes

Table 4
Stepwise Regression Analysis for the Entire Sample and for the Task Subsample

Statistical parameter	All values	Task subsample
Multiple R	.792	.801
R^2	.628	.641
F	28.268*	18.220*
df	4/67	5/51
Constant (Fisher z)	.816	.303

* $p < .01$.

to 24 hours, with a mean of 95 minutes. A low but significant positive correlation was found only between the dependent measure and the time elapsed prior to the first leadership measure for both the total sample and the subsample (Table 2).

Only six of the correlations involved leadership ratings that had practical significance for the group members. They were from studies which used the same methodology (Burroughs & Jaffee, 1969; Jaffee & Lucas, 1969). The six correlations were significantly lower than the others in the task subsample, $t(55) = 3.33$, $p < .001$.

Multiple Regression Analysis

Stepwise multiple regression analyses were performed to determine the amount of variance in the leadership–participation Fisher z scores accounted for by the 28 independent variables. The analyses for the entire sample and for the subsample are summarized in Tables 4 and 5. The B coefficients are for nonstandardized measures; the beta values are for standardized scores.

For the entire set of leadership–participation correlations, the multiple R was .792, accounting for 63% of the variance. The largest coefficient was a negative value for maintenance functions. The inclusion of this variable and the ratings of the breadth of the leadership measures support the task leadership activity explanation. The contribution of the mixed-sex measure supports the view that the leadership–participation relationship is due in part to the sexual composition of the group. The inclusion of the observer ratings supports the observer bias hypothesis. Because the maintenance functions were excluded from the subsample, the strongest predictor for the entire sample was removed from the second analysis. The multiple R was .801. The first three variables to be included were added in the same order as in the total sample analysis. Two additional variables were included, all male groups versus other sex composition groups, and groups of college students versus noncollege subjects.

Conclusions

Three major explanations for the leadership–participation rate relationship were supported by the data analysis: (a) task leadership behaviors are a component of total participation and a basis for leadership status; (b) those with superior task ability make more task-related contributions to the group, which increases their participation; and (c) males are permitted both greater influence and higher participation rates (although the sex differences were confounded with group differences). These explanations can be inte-

Table 5
Regression Beta Weights for the Entire Sample and for the Task Subsample

Variable	All values				Task subsample			
	B	β	F	R^2	B	β	F	R^2
Maintenance functions	−.77	−.68	80.21**	.37	—	—	—	—
Mixed-sex subjects	.37	.41	26.68**	.52	.63	.82	40.56**	.37
Observer ratings	.36	.31	15.06**	.58	.39	.41	22.07**	.48
Breadth of measure	.03	.25	9.09**	.63	.02	.22	5.52*	.57
Male subjects	—	—	—	—	.34	.44	8.51**	.59
College subjects	—	—	—	—	.27	.32	7.03*	.64

* $p < .05$. ** $p < .01$.

grated if the performance of task leadership behaviors is viewed as the major source of the variance common to leadership status and participation rates. Competency and sexual status can be considered as variables that moderate the performance of such behaviors. Behavioral manifestations of competency (such as giving information and proposing solutions, procedures, and structure) are in themselves task leadership behaviors. In addition, members of new groups who feel they are competent make more influence attempts, whereas members who realize that they are less competent make fewer leadership attempts (Hemphill, 1961). The literature on sex differences indicates that males are considered to be more influential, talk more often, and perform more task leadership acts.

The suggestion is made from several points of view that the quality and frequency of members' contributions are correlated. For example, theories of emergent leadership postulate that those who have established their competency are allowed greater influence as the group continues (Bales, 1953; Stogdill, 1959). Hoffman and Clark (1979) report positive correlations between the group members' total participation and the extent to which they contributed to the adoption of the solution.

Some of the explanations of the leadership–participation rate relationship that are based on within-group differences (such as personality traits and motivation) could not be tested in this study. However, the performance of task leadership behaviors can be offered as an alternative to Sorrentino and Boutillier's (1975) assertion that differences in group members' motivation explain the leadership–participation rate relationship. They found participation rates, but not the quality of the solution offered by the confederates, to be related to peer ratings of task leadership. Sorrentino and Boutillier suggested that the high participants received higher leadership ratings because they were perceived to be more motivated. However, in manipulating participation rates, Sorrentino and Boutillier held constant the proportion, but not the frequency, of the types of behavior performed by the high and low participant confederates, thus confounding task behaviors with total participation. Consequently, the higher participants may have been rated higher in task leadership because the group members recognized their performance of more task leadership acts.

Our analyses suggest that the *high* correlation of task leadership and participation may be cautiously generalized to real groups. Almost all of the findings are from formally leaderless groups. The relationship between the two variables is lower if the leadership ratings have practical implications for group members and if the groups are not comprised of college students. These variables were also confounded with task differences, which makes it more difficult to generalize the strength of the leadership–participation rate relationship to a variety of real-world tasks. The degree of association between leadership and participation was also lower when the leadership judgments were not made by observers. In addition, for all but two of the correlations analyzed, less than 45 minutes had passed before the first leadership measure was taken.

Additional research is warranted in several areas:

1. The generalizability of the leadership–participation rate relationship needs to be explored.

2. Predictive research is needed to determine if the performance of task leadership behaviors actually serves as the basis for the common variance between leadership and participation and if it is moderated by sexual status, competence, and motivation.

3. Although it is clear that task leadership status and participation covary, the causal relationship between the variables is not clear. Perhaps high participation precedes emergent leadership status. Or perhaps, as the leadership valence theory (Stein et al., 1979) suggests, the attainment of leadership status occurs almost simultaneously with acts of participation. According to that theory, the leadership status of a member is enhanced by each successful attempt to influence others.

4. The relationship of participation rates to formal leadership status has not been examined. The roles of formal leaders may require that they speak more often than others (Burke, 1974). When the formal roles require

the performance of task leadership behaviors, such behaviors could be the basis for both high participation and judgments of the extent to which an individual actually leads the group.

Reference Note

1. Stein, R. T., & Heller, T. *The relationship of emergent leadership status and high verbal participation in small groups: A review of the literature.* Manuscript submitted for publication, 1979.

References

Bales, R. F. The equilibrium problem in small groups. In T. Parsons, R. F. Bales, & E. A. Shils (Eds.), *Working papers in the theory of action.* Glencoe, Ill.: Free Press, 1953.

Bales, R. F. Task roles and social roles in problem-solving groups. In E. E. Maccoby, T. M. Newcomb, & F. L. Hartley (Eds.), *Readings in social psychology* (3rd ed.). New York: Holt, Rinehart & Winston, 1958.

Bass, B. M. An analysis of leaderless group discussion. *Journal of Applied Psychology,* 1949, *33,* 527–553.

Bass, B. M. The leaderless group discussion as a leadership evaluation instrument. *Personnel Psychology,* 1954, *7,* 470–477.

Bass, B. M., Wurster, C. R., Doll, P. A., & Clair, D. J. Situational and personality factors in leadership among sorority women. *Psychological Monographs,* 1953, *67*(16, Whole No. 366).

Bates, A. P. Some sociometric aspects of social ranking in a small, face-to-face group. *Sociometry,* 1952, *15,* 330–341.

Bavelas, A., Hastorf, A. H., Gross, A. E., & Kite, W. R. Experiments on the alteration of group structure. *Journal of Experimental Social Psychology,* 1965, *1,* 55–70.

Berkowitz, L. Social desirability and frequence of influence attempts as factors in leadership choice. *Journal of Personality,* 1956, *24,* 424–435.

Blau, P. M. *Exchange and power in social life.* New York: Wiley, 1964.

Borgatta, E. F. Analysis of social interaction and sociometric perception. *Sociometry,* 1954, *17,* 7–32.

Borgatta, E. F., & Bales, R. F. Sociometric status patterns and characteristics of interaction. *Journal of Social Psychology,* 1956, *43,* 289–297.

Burke, P. J. Participation and leadership in small groups. *American Sociological Review,* 1974, *39,* 832–842.

Burroughs, W. A., & Jaffee, C. L. Verbal participation and leadership voting behavior in the leaderless group discussion. *The Psychological Record,* 1969, *19,* 605–610.

Carter, L., Haythorn, W., Shriver, B., & Lanzetta, J. The behavior of leaders and other group members. *Journal of Abnormal and Social Psychology,* 1950, *46,* 589–595.

Clifford, C., & Cohn, T. S. The relationship between leadership and personality attributes perceived by followers. *Journal of Social Psychology,* 1964, *64*(1), 57–64.

Crockett, W. H. Emergent leadership in small decision-making groups. *Journal of Abnormal Social Psychology,* 1955, *51,* 378–383.

French, R. L. Verbal output and leadership status in initially leaderless discussion groups. *American Psychologist,* 1950, *5,* 310–311. (Abstract)

Gintner, G., & Lindskold, S. Rate of participation and expertise as factors influencing leader choice. *Journal of Personality and Social Psychology,* 1975, *32,* 1085–1089.

Gustafson, D. P., & Harrell, T. W. A comparison of role differentiation in several situations. *Organizational Behavior and Human Performance,* 1970, *5,* 299–312.

Hare, A. P. *Handbook of small group research* (2nd ed.). New York: Free Press, 1976.

Heinicke, C. M. *The development of structure in small groups of men.* Unpublished master's thesis, Northwestern University, 1950.

Heinicke, C., & Bales, R. F. Developmental trends in the structure of small groups. *Sociometry,* 1953, *16,* 7–38.

Heller, T., & Stein, R. T. Explaining the relationship of leadership to high verbal participation. *Personality and Social Psychology Bulletin,* 1978, *4,* 356. (Abstract)

Hemphill, J. K. Why people attempt to lead. In L. Petrullo & B. M. Bass (Eds.), *Leadership and Interpersonal Behavior.* New York: Holt, Rinehart & Winston, 1961.

Hoffman, L. R., & Clark, M. M. Participation and influence in problem-solving groups. In L. R. Hoffman (Ed.), *The group problem-solving process.* New York: Praeger, 1979.

Hollander, E. P. Conformity, status, and idiosyncrasy credit. *Psychological Review,* 1958, *65,* 117–127.

Jaffee, C. L., & Lucas, R. L. Effects of rate of talking and correctness of decisions on leader choice in small groups. *Journal of Social Psychology,* 1969, *79,* 247–254.

Juola, A. E. Leaderless group discussion ratings: What do they measure? *Education and Psychology Management,* 1957, *17,* 499–509.

Kanungo, R. Sociometric rating and perceived interpersonal behavior. *Journal of Social Psychology,* 1966, *68,* 253–268.

Kirscht, J. P., Lodahl, T. M., & Haire, M. Some factors in the selection of leaders by members of small groups. *Journal of Abnormal and Social Psychology,* 1959, *58,* 406–408.

Moment, D., & Zaleznik, A. *Role development and interpersonal competence.* Boston: Harvard University, Graduate School of Business Administration, 1963.

Morris, C. G., & Hackman, J. R. Behavioral correlates of perceived leadership. *Journal of Personality and Social Psychology,* 1969, *13,* 350–361.

Norfleet, B. Interpersonal relations and group productivity. *Journal of Social Issues*, 1948, *4*(2), 66–69.

Reilly, R. R., & Jaffee, C. L. Influence of some task-irrelevant factors on leader selection. *Psychology Record*, 1970, *20*, 535–539.

Shaw, M. E. *Group dynamics: The psychology of small group behavior*. New York: McGraw-Hill, 1971.

Shaw, M. E. Scaling group tasks: A method for dimensional analysis. *JSAS Catalog of Selected Documents in Psychology*, 1973, *3*, 8. (Ms. No. 294)

Slater, P. E. Role differentiation in small groups. *American Sociological Review*, 1955, *20*, 300–310.

Sorrentino, R. M., & Boutillier, R. G. The effect of quantity and quality of verbal interaction on ratings of leadership ability. *Journal of Experimental Social Psychology*, 1975, *11*, 403–411.

Stang, D. J., Castellaneta, J. A., Constantinidis, G., & Fortuno, C. R. Actual vs. perceived talkativeness as determinants of judged leadership, popularity and likeableness. *Bulletin of the Psychonomic Society*, 1976, *8*, 44–46.

Stein, R. T. Accuracy of process consultants and untrained observers in perceiving emergent leadership. *Journal of Applied Psychology*, 1977, *62*, 755–759.

Stein, R. T., Hoffman, L. R., Cooley, S. H., & Pearse, R. W. Leadership valence: Modeling and measuring the process of emergent leadership. In J. G. Hunt and L. L. Larson (Eds.), *Crosscurrents in Leadership*. Carbondale, Ill.: University of Southern Illinois Press, 1979.

Stogdill, R. M. *Individual behavior and group achievement*. New York: Oxford University Press, 1959.

Strodtbeck, F. L., & Hook, L. H. The social dimensions of a twelve-man jury table. *Sociometry*, 1961, *24*, 397–415.

Strodtbeck, F. L., James, R. M., & Hawkins, C. Social status in jury deliberations. *Journal of Abnormal and Social Psychology*, 1957, *22*, 713–719.

Strodtbeck, F. L., & Mann, R. D. Sex role differentiation in jury deliberations. *Sociometry*, 1956, *19*, 3–11.

Thomas, E. J., & Fink, C. F. Effects of group size. *Psychological Bulletin*, 1963, *60*, 371–384.

Wilson, S. R. Leadership, participation and self-orientation in observed and nonobserved groups. *Journal of Applied Psychology*, 1971, *55*, 433–438.

Zdep, S. M., & Oakes, W. F. Reinforcement of leadership behavior in group discussion. *Journal of Experimental Social Psychology*, 1967, *3*, 310–320.

Received September 25, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 11, 2003–2013

Environmental Determinants of Memory Improvement in Late Adulthood

Ellen J. Langer
Harvard University

Judith Rodin
Yale University

Pearl Beck and Cynthia Weinman
Graduate Center, City University of New York

Lynn Spitzer
Yale University

Two studies were conducted with nursing home residents to determine whether memory could be improved. This was accomplished by increasing the cognitive demand of the environment and then varying the extent to which residents were motivated to attend to and remember these environmental factors. In Study 1, motivation to practice recommended cognitive activities was manipulated by varying the degree of reciprocal self-disclosure offered by interviewers in a series of dyadic interactions. In Study 2, motivation to practice recommended cognitive activities was manipulated by varying whether positive outcomes were contingent on attending to and remembering these activities, which increased in demand over time. Whether as a function of interpersonal (Study 1) or practical (Study 2) incentives, engaging in cognitive activity resulted in improvement on standard short-term memory tests, including probe recall and pattern recall, as well as in improvement on nurses' ratings of alertness, mental activity, and social adjustment for experimental groups relative to controls.

Arguments and counterarguments over evidence purportedly showing age-related intellectual and cognitive decline have been made with great vigor (cf. Baltes & Schaie, 1976; Horn & Donaldson, 1976; Horn & Donaldson, 1977). Moreover, there is disagreement about what the deficits might be, if they do exist. Let us take memory as an example. Adamowicz (1974) found short-term memory decrements at the encoding and postcoding phases but not at the retrieval phase, whereas Anders and Fozard (1973) basically found retrieval deficits. Drachman and Leavitt (1972) suggested that the major cause of memory deficiency for the elderly lies not in a deficit in retrieval but in a disorder of storage. Yet another view is that of Eysenck (1973), who suggested that the decrement is primarily in the decision process and not in retrieval.

In this article we do not wish to question whether there is memory decline with advanced age. Instead, we ask to what the decline may be attributed when it does appear. Thus, rather than looking for either a description of the memory deficit (as in the studies previously cited) or methodological artifacts that may explain away purported differences (cf. Baltes & Schaie, 1976), we suggest another focus of analysis: one in which deficits are seen as real and one in which environmental factors are seen as influencing and masking cognitive abilities of the elderly. Our position is basically that social and environmental factors may impact on memory directly, leading to real deficits that may only look as if they are biologically determined because they are age related.

How might social factors directly impact

This research was supported by Grant BNS 77-24159 from the National Science Foundation to the first author. The authors are grateful to the directors and staffs of the Arden House and Florence Nightingale nursing homes for their permission and cooperation in conducting the two studies.

Requests for reprints should be sent to Ellen Langer, Department of Psychology and Social Relations, Harvard University, 33 Kirkland Street, Cambridge, Massachusetts 02138.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3711-2003$00.75

on cognitive processes? The elderly, especially those residing in institutions, are a population who are by and large confined to a very restricted living space. Such restriction, arising from social system factors and physical ailments orthogonal to direct brain deterioration, may have far-reaching effects on cognitive processes. For these people, every day is like every other day. Moreover, for those residing in institutions, every room is like every other room, and each person seen today is the same as each person seen yesterday and the day before. Living in such highly redundant environments leaves one with little information to process. As a result, the ability to think and remember may diminish. During the earlier part of one's life, it may be an efficient strategy to try to minimize informational input when living in a demanding and overstimulating environment. Over the years one may have successfully taught oneself to tune out much of the environment in order to function more adaptively (cf. Langer, 1978; Langer, Note 1) only to find that in later years such proficiency has become maladaptive.

After years of learning how to avoid expending cognitive effort, a person might find it difficult to know how to reverse the process when faced with little to think about and little to do. Yet it seems that reversal is possible at least in some areas. In earlier research (Langer & Rodin, 1976; Rodin & Langer, 1977) we found that making more demands on the decision-making abilities of institutionalized elderly adults had the effect of improving health and psychological well-being and tended to be related to lower mortality rates as well. In the present studies we consider whether improvements in memory might also result from changes in the social psychological environment.

If cognitive abilities do diminish at all with disuse, then practice in thinking over time may be successful in reversing the debilitations. However, if people have been implicitly led in the ways outlined above to avoid expending cognitive effort, then more than a mere suggestion to expend this effort would seem necessary to motivate them to do so. The following studies were undertaken with these two considerations in mind. In both studies we attempted to motivate subjects to adhere to a recommended course of action over time that asked them to think about various issues. Encouraging this commitment to increase cognitive activity was accomplished in the first study by providing the opportunity for reciprocal self-disclosure between the subject and an interviewer who recommended the particular course of cognitive activity.

According to Janis and Rodin (1979), eliciting self-disclosure is one factor that is critical in helping a person to establish referent power (French & Raven, 1959) that promotes commitment to recommended courses of action. Persons have referent power for those who perceive them as likeable, benevolent, admirable, and accepting. Their motivating power derives from this source. Rodin and Janis (1979) have described how the use of referent power in health care settings promotes internalization of medical recommendations made by health care professionals. Because no coercive inducements are used to motivate the person, Rodin and Janis hypothesized that the use of referent power increases the likelihood that persons will feel personal responsibility in adhering to prescribed regimens. To establish referent power that motivates persons to commit themselves to recommended courses of action, an individual first must encourage them to disclose their personal feelings, troubles, or weaknesses (Janis, in press; Quinlan, Janis, & Bales, in press). Demonstrated willingness to engage in reciprocal self-disclosure is effective in this regard (cf. Jourard, 1971).

We obtained a commitment to increase cognitive activity in the second study by introducing a contingency whereby greater cognitive effort resulted in greater tangible reward. In both studies, we hypothesized that relative to control groups, increased cognitive activity in experimental groups would promote more efficient performance in cognitive areas such as memory.

Study 1

In this study we used a series of reciprocal self-disclosure conversations as an incentive

for adhering to a particular course of recommended cognitive activity. The cognitive activity involved remembering a series of questions and searching the environment and one's long-term memory for information relevant to their answers. A low-self-disclosure group was included to control for the amount of time spent with an interviewer discussing past and present events. This more standard interview procedure, however, was not expected to provide sufficient interpersonal incentive to occasion the kind of commitment necessary to stimulate renewed cognitive activity. To control for changes that may have occurred simply because of multiple administrations of the memory tests, a no-treatment group was added to the design.

Method

Subjects

Fifty-four middle-class ambulatory residents of a private suburban nursing home participated in the study. Residents who were judged by the institution's head social worker to be at least somewhat alert took part in the study. The average length of stay for the sample was 2.8 years. All participants could eat their meals in the dining room without assistance. Subjects were matched on age, length of time in the nursing home, social worker's ratings of alertness, and sex and on this basis were assigned to one of four groups: high reciprocal self-disclosure ($n = 10$, M age $= 80.4$); low reciprocal self-disclosure ($n = 20$, M age $= 79.6$); pretest and posttest only (for no-interview controls, $n = 14$, M age $= 78.9$); and pretest only (for test-instructions subjects, $n = 10$, M age $= 78.8$).[1]

Procedure

All subjects were visited 1 month prior to the experimental treatment period by an experimenter who administered an interview and the memory tests described below. She did not participate in any of the interviews and was blind to experimental conditions to which subjects had been assigned.

In both experimental conditions, subjects were subsequently visited a total of four times by one of five interviewers (four males, one female) during the course of a 6-week period. Each interviewer visited subjects in both experimental conditions. None of them knew about the memory pretest and posttest, and all were blind to the experimental hypothesis regarding memory. Visits were spaced so that there was at least 1 week between the visits. Interviewers were instructed to keep the length of each visit between 30 and 40 minutes. At the beginning of the first session, residents were told that the interviewers would be coming for about 6 weeks.

The interviewers introduced themselves as college students who were interested in getting to know people who were much older than themselves. The interviewer told the subject that he or she would like to get together once a week with the participant to talk. If the person agreed, the interviewer began a discussion. The nature of the discussion was determined by the condition to which the subject had been assigned. At the end of each visit, the interviewer told the subject that he or she would return in approximately 1 week.

High-reciprocal-self-disclosure group. In this condition the interviewer began the first visit by stating that "one of the things I never get to do is to tell someone older than I am the kinds of things I am doing and thinking about." In each of the four visits the interviewer began the visit by talking about events that were occurring in his or her own life (e.g., school, family, jobs) in a largely unstructured way, with the only requirement being that she or he let the resident talk whenever she or he wanted to. At the end of each visit, the experimenter suggested specific topics about which subjects were asked to think. These were to be discussed on the next visit.

Low-reciprocal-self-disclosure group. Here the interviewer began the discussion by stating that "one of the things I never get to hear very much about is what people who are older than I am are doing and thinking about." The questions asked during this and subsequent visits were concerned with current occurrences and events, both personal and public (e.g., what was going on that day, families). The interviewer was instructed to ask questions, comment appropriately, and listen attentively and interestedly. At the end of each visit, the experimenter suggested specific topics about which subjects were asked to think. These were identical to those suggested to high-disclosure subjects and were to be discussed on the next visit. This standard interview procedure was typical of the kinds of interviews older people in nursing homes experience.

No-treatment group. Subjects in this condition were not visited by the interviewers, but received the same memory pretest and posttest as the above two groups.

Dependent Variables

Two types of measures were used to assess the effects of the experimental manipulations. The first measure, a questionnaire, was responded to by the nurses who staffed the floors on which the subjects lived; the nurses were unaware of the experimental hypotheses and treatments. The questionnaires were completed at two different times—once before the experimental manipulations were introduced and again several weeks after the experimental treatments

[1] Only those subjects who were available for each interview session were included in the final sample.

had ended. The questionnaires included 10-point scales that asked for ratings of how happy, alert, dependent, sociable, and active the residents were.

The other measures were two types of memory tests administered 1 month prior to and 1 week after the experimental treatments. To minimize the anxiety that is often felt by the elderly when presented with testlike situations, the tests were not described as such. Instead the experimenter introduced herself as an individual interested in finding out what changes might be made to make the nursing home "a more pleasant place for the people who live here." She then asked subjects for their suggestions as to what changes they would like to see made. After the subjects made their suggestions, the experimenter requested that they give their opinion on some new ideas that the experimenter had come up with. The two memory tests were disguised as a new activity or procedure that might be instituted in the nursing home. The subjects "tried out" these "new activities and procedures" and offered an opinion of each. The order in which the tests were presented was systematically varied.

The two tests used were pattern recall and probe recall. Pattern recall is a test of immediate recall of visual patterns. The subject is required to watch the experimenter tap a set of four blocks in a certain order and to repeat those exact movements immediately afterward. As the test proceeds the series of items (each block pointed to is considered an item) increases in length. The score reflects the length of the series a subject is capable of imitating.

In probe recall, a test that measures short-term memory of paired items, photographs of five people were shown to the subject, and each photograph was paired with a person's name. The photographs were then presented again at various intervals throughout the test. Subjects were required to recall the name associated with each photograph. To make this test relevant, pictures were used of elderly people who, subjects were told, might be coming to the nursing home.

Both tests had two equivalent forms, with split-half reliability and/or test–retest reliability ($rs = .86$ and .88, respectively) allowing both premanipulation and postmanipulation measures to be taken. Approximately one half of the subjects in each condition received Form A of the tests for the pretest and Form B for the posttest, whereas the other half received them in the opposite order. We predicted that greater practice in thinking would result in higher scores on these measures for the high-reciprocal-self-disclosure group relative to the remaining groups.

As a secondary goal of this experiment, we were interested in determining whether making the tests relevant and nonthreatening in this way would actually have an effect on test performance, since test anxiety alone may impair performance. Therefore, we also included for comparison another no-treatment group of similar nursing home residents, who were given the pretest measures only and were told that they were memory tests rather than "new activities."

Results

Analyses of pretest scores on both reca tests and of nurses' ratings revealed no signi icant differences among the three groups a hence demonstrated group comparability the start of the experiment (all $Fs < 1$).

Table 1 presents group means and chan scores on the recall tests as a function of t experimental treatment. Analyses of varian for both tests individually showed significa differences among the groups: for patte recall, $F(2, 41) = 3.25$, $p < .05$; for pro recall, $F(2, 41) = 3.92$, $p < .05$. Individu comparisons revealed that whereas the hi reciprocal-self-disclosure group showed i proved memory relative to controls on b tests, the low-reciprocal-self-disclosure gro was not reliably different from the no-tre ment group. In addition, pretest–posttest co parisons on the pattern recall test showed th there was significant improvement for t high-self-disclosure group, $t(9) = 2.36$, p .05, but no significant change for the low-se disclosure group or for the control group. Si ilar analyses for the probe recall test revea not only that the treatment improved memo for the high-self-disclosure group, $t(9)$ 3.46, $p < .01$, but also that when there was treatment or minimal treatment, memory clined significantly in as short as 6 wee time: for no treatment, $t(12) = 2.29$, p .05; for low self-disclosure, $t(18) = 2.18$, p .05.

A fourth group of residents were also giv the memory tests ($n = 10$). For them t tests were presented as memory tests and we not integrated into the nursing home conte A comparison of this group's scores with t scores for subjects in all the other groups co bined ($n = 44$) was conducted to assess t effects of test context. Subjects in the tes concealed groups scored significantly bett (pattern recall $M = 4.0$; probe recall M = 10.1) than subjects in the test-unconceale group (pattern recall $M = 2.4$; probe recal $M = 5.3$), $t(52) = 2.01$, $p < .05$; $t(52)$ = 2.68, $p < .02$, respectively.

After each interview, the interviewers rate residents for amount spoken during the inter view, percentage of items initiated for discus

Table 1
Study 1: Mean Change Scores (Posttreatment Minus Pretreatment) on Memory Tests for Each Experimental Condition as a Function of Level of Involvement

Group	Memory test					
	Pattern recall			Probe recall		
	Pre	Post	Change	Pre	Post	Change
High reciprocal self-disclosure	4.7	6.7	2.0_a	10.0	15.3	5.3_a
Low reciprocal self-disclosure	4.0	3.3	$-.7_b$	12.2	9.2	-3.0_b
No treatment	4.8	4.6	$-.2_b$	11.5	9.0	-2.5_b

Note. Different subscripts within each test indicate significant differences at $p < .05$.

sion, and how friendly they were to the interviewer. The high-reciprocal-self-disclosure residents spoke on the average for 70% of the interview and initiated 75% of the items for discussion. By comparison, the low-self-disclosure subjects spoke for an average of 45% of the interview and initiated only 30% of the items for discussion. The groups were not rated differently in how friendly they were to the interviewer. There were also no reliable differences between conditions in the length of interviews, and all interviews became longer over the 4 weeks.

The next analyses of variance were performed on the nurses' ratings. These results reveal that the greater practice in thinking initiated by the high-reciprocal-self-disclosure treatment not only influenced short-term memory but produced general social psychological effects that were discernible to nurses, who were unaware of the experimental treatment. As can be seen in Table 2, the high-involvement group was rated as being more aware, more active, and more self-initiating than the no-treatment group (all ps $< .05$). Although the low-involvement treatment appears to have resulted in nurses' judgments of a greater self-initiating attitude, on other measures the low-involvement group was not significantly different from the no-treatment group.

Table 2
Study 1: Nurses' Mean Ratings (Posttreatment Minus Pretreatment Scores) of Residents' Behavior by Experimental Condition

Group	Mood	Awareness	Sociability	Activity	Self-initiating	Alertness	Health
High reciprocal self-disclosure							
Pre	4.5	3.5	5.3	3.0	2.8	6.5	4.3
Post	5.0	5.8	7.0	4.5	5.3	6.8	5.1
Change	$.5_a$	2.3_a	1.7_a	1.5_a	2.5_a	$.3_a$	$.8_a$
Low reciprocal self-disclosure							
Pre	4.8	4.7	5.0	3.3	3.7	5.8	4.7
Post	5.5	5.2	4.5	1.7	5.7	5.8	4.7
Change	$.7_a$	$.5_b$	$-.5_a$	-1.6_b	2.0_a	$.0_a$	$.0_a$
No treatment							
Pre	4.0	4.5	5.3	3.8	3.5	5.8	4.0
Post	4.3	4.5	4.0	3.5	4.0	5.8	4.0
Change	$.3_a$	$.0_b$	-1.3_a	$-.3_b$	$.5_b$	$.0_a$	$.0_a$

Note. Cells bearing different subscripts within each rating are significantly different at $p < .05$.

Discussion

The results of Study 1 suggest that memory can be improved when the elderly are motivated to commit themselves to a recommended course of action that involves increased use of their cognitive abilities. Because people in the high-reciprocal-self-disclosure group spoke more and initiated more topics for conversation than their low-self-disclosure counterparts during the 6-week period of the study, they appear to have been more involved in the substance and conduct of the interviews. This greater involvement apparently led to a greater commitment to follow the interviewer's recommendations to think about a variety of proposed topics for future discussion. Although the low-involvement group also was asked to think about things to be discussed in between the scheduled visits, the character of the interaction was not sufficiently involving to encourage them to make a commitment to do so. Participants were not judged to be more friendly in any one treatment condition, suggesting that increased thinking and not simply greater liking for the interviewer was responsible for the results obtained. In addition, the items on which the high-self-disclosure group increased on the nurses' ratings (aware, active, and self-initiating) also seem to reflect a general increased involvement with the environment.

It is significant to note that the memory tests used were measures of short-term recall, in which there has been strong but by no means unequivocal evidence for age-related decline. The study suggests that such decline is not inevitable and that it can be strongly affected by environmental–social processes. However, we can only infer that the subjects actually followed the recommended cognitive activity as a result of the high-reciprocal-self-disclosure treatment and that this led to the memory improvement. Therefore, a second study was designed to increase, in a very different way, the likelihood that subjects would exercise their cognitive abilities.

Study 2

Whereas the first study provided interpersonal incentives for increased cognitive activity, in the present study practical incentives were employed. An experimental group for whom positive outcomes depended on remembering was compared with a group for whom positive outcomes were not contingent on memory and with a no-treatment control group. We predicted that the contingent group would engage in more cognitive activity that would result in improved memory.

Method

Subjects

Forty-five middle-class residents of a second nursing home participated in this research. The average length of stay for the sample was 3.4 years. There were 14 subjects (M age = 80.62 years) in the contingent group, 16 subjects (M age = 76.19) in the noncontingent group, and 15 subjects (M age = 80.73) in the no-treatment control group. Only those residents who had been judged on an 11-point scale by the institution's head nurse to be at least somewhat alert (contingency $M = 7.77$, noncontingency $M = 7.13$, and no-treatment $M = 8.33$) served as subjects. Though residents varied greatly with respect to their particular medical concerns, to participate in the investigation they had to be healthy enough to be ambulatory. Thus, there were no initial group differences with respect to social class, age, alertness, and general health. Similarly, no initial mood differences were revealed on an 11-point scale of self-reported happiness (contingency $M = 5.21$, noncontingency $M = 5.58$, and no-treatment $M = 5.53$).

The experiment was conducted by seven female experimenters who were blind to the study's hypothesis.

Procedure

The residents in the two experimental conditions were visited a total of nine times over the course of a 3-week period, with various cognitive demands made during each visit. The visits were spaced such that the amount of time in between visits increased over the 3 weeks, thereby increasing the demands on subjects' memories over the course of the study. Those in the no-treatment condition were visited only on the first and last day of the study (when the memory tests were administered). Memory tests (described below) were given 2 days after the ninth visit.

After introducing herself, the experimenter told each resident that she was a student who was interested in learning about life in a nursing home. She explained that with the person's permission, she would like to come several times during the next 3 weeks to discuss certain things with him or her. To insure that the relationship from the resident's per-

spective was time bound, she indicated that she would have to return to school after 3 weeks but that they could correspond if the resident so desired. The nature of the subsequent interaction was a function of the condition to which the resident had been assigned.

Contingent condition. Residents in this condition were told that they were going to be asked several questions during each visit and that they would receive one poker chip for each correct response. The experimenter explained that at the end of the 3 weeks they would be able to redeem the chips for a gift and that the more chips they had the better the gift would be. The experimenter then asked the resident a few questions about that day's meals and activities. The first questions asked were made simple enough to insure that everyone received at least one chip on the first visit (e.g., "Did you eat breakfast today?"). The residents were then asked two questions that required them to seek information from their immediate environment. If they did not know the answer, they were told they would be asked again on the next visit. On the first visit, the questions were "Do you know the names of any of the nurses?" and "Do you know the name of another resident?" Subsequently, each individual was asked three such questions per visit.

The assumption, which proved correct, was that the resident would not know the answer to at least one of the questions when first asked and would therefore have to engage in an information-seeking process before the next visit. The questions were of varying levels of difficulty, but they all concerned life in the institution (e.g., "How many different nurses' and patients' names do you know?", "What activities were there today?", "When will the next cocktail party/bingo/concert take place?", and "Approximately how many male residents are there on this floor?"). Although the information-seeking questions kept changing, residents were asked what they ate for breakfast and lunch that day, what they ate for dinner the night before, and what activities they engaged in during that day on each visit.

At the end of each visit, before departing, the experimenter told the resident, "Please try and remember everything you do today and also find out . . . [e.g., when the next bingo game will take place]. I'll be returning on . . . [day specified] to talk to you and will give you more chips for more answers." Thus, each subject had to remember what information to seek out, to actively seek it, and then to remember it until the experimenter's next visit. Each question was asked on three consecutive visits unless the resident knew the answer sooner. If on the third visit the resident still did not know the information, the question was not asked again.

On those questions for which it was impossible to make an immediate determination of accuracy (e.g., what the person ate for breakfast), the answers were taken at face value, and the residents in this condition received a chip every time they replied to a question. Since persons in all conditions in the study at several points said "I don't remember" in response to questions asked and the answers given were accurate much more often than not in the cases that could be checked, this assumption of accuracy seemed reasonable.

Noncontingent condition. The substance and the form of the visits in this condition were exactly the same as those in the contingent condition. The only difference between the noncontingent condition and the contingent condition was that the residents in the former were told that they would be given chips as mementos of each visit, the number of which would depend on how many the experimenter had with her that day. The residents in this condition were told that the chips would be redeemable for a gift in 3 weeks. Each subject in this condition was yoked to a subject in the contingent condition for number of chips and length of interview.

No-treatment condition. These residents received an initial visit by the experimenters and were asked the same questions as the residents in the contingent and noncontingent conditions. At the time of this first visit, they were told that the experimenter would return to speak to them in 3 weeks and on that occasion, in return for cooperating with the project, they would each receive a gift. This was done to control for the possible effects of anticipating some positive future outcome. They were not visited again until the measures were taken.

Dependent Variables

Four days after the last visit all subjects were given a test of immediate memory and a test of memory for remote events by an experimenter with whom they had no previous contact. The test of immediate memory used was a probe recall measure in which test words varied in terms of how many new words were presented between the first presentation of the test word and its repetition. There were approximately 5 sec in between the presentation of each word. After the test was administered, the experimenter also asked the residents several questions concerning remote past events (e.g., "When did the Great Depression occur?"; "Where is the Empire State Building?"; "How long has television been around?"; and "What were the names of your neighbors before you moved here?"). Subjects in the contingent and noncontingent groups also were asked how much they liked the woman who had visited them, whether they remembered her name, and approximately how many times they had been visited by her.

Responses to questions asked during visits to subjects in the contingent and noncontingent groups allowed for a comparison of memory for recent events between these two groups over time. These questions fell into two categories: memory for events they had experienced (e.g., meals and activities) and memory for information they had to seek out and remember (e.g., how many male residents there were in the nursing home).

In addition to these measures, with informed con-

sent, each resident's medical records were examined and all entries regarding the following aspects of the person's medical condition were recorded: medications, major tranquilizers, vital signs, complaints, and social adjustment.

Results

The major questions addressed in this study were whether giving people a reason to remember would commit them to greater cognitive activity and, if they were committed, whether the beneficial effects would generalize to other memory processes. Before looking at generalization effects, we felt it important to see if the training was effective. That is, did the contingent group show signs of improvement for the things on which the contingency was focused? A comparison of memory for meals and activities over time suggests that the contingency treatment was indeed effective. A repeated measures analysis of variance for meals remembered over time yielded a significant effect for contingency, $F(1, 28) = 4.23$, $p < .05$.[2] Of a maximum of three, the contingent group remembered an average of 2.76 meals, whereas the noncontingent group remembered an average of 2.20. The Group × Time interaction was also significant, $F(8, 224) = 2.66$, $p < .01$ (Winer, 1971). Whereas there was no difference between the two groups on the first visit, the contingent group remembered significantly more meals than the noncontingent group for each visit after the first. The difference between the first and last visits for the contingent group was also significant, $t(13) = 2.94$, $p < .02$. However, there was no difference between the first and last visit for the noncontingent group.

In addition to improved memory of meals over time, subjects in the contingent condition took less time than subjects in the noncontingent condition to find out the required information regarding activities in the nursing home. Obviously, only those subjects who did not know the answers upon being asked the first time were included in this calculation, since we were interested in subjects finding out and remembering information they did not already know. (There was no initial difference among the groups with respect to this knowledge.) Those who knew the answer by the second time they were asked were given a score of 3, those who obtained the information after being asked three times were assigned a weight of 2, and those who only knew the answer after four visits were assigned a weight of 1. The mean weighted total over all the target questions was 12.40 for the contingent group, compared with only 8.25 for the residents in the noncontingent group, $t(28) = 2.62$, $p < .05$.

Attitudes toward the interviewer and the visits also were assessed. Although each group received a total of 9 visits, the contingent group estimated that they had received an average of 13 visits, and the noncontingent group estimated an average of 6 visits. Eighty-two percent of the contingent group *over*estimated the number of visits, whereas 73% of the noncontingent group *under*estimated the number of visits, $\chi^2(1) = 6.60$, $p < .05$. Both groups were also asked to indicate on a 10-point scale how much they liked the person who had visited them. The results indicated that all participants claimed to have liked the experimenter; the mean for the contingent group was 9.92, compared with 7.93 for the noncontingent group, $F(1, 26) = 7.18$, $p < .02$. Residents were also asked if they remembered the name of the person who had visited them. Sixty-nine percent of the contingent group remembered, compared with 53% of the noncontingent group. Although in the predicted direction, this difference was not statistically significant.

Probe Recall Test

To test whether training in remembering recent events would generalize to improvement in immediate memory, scores on this test were compared for the three groups. Out of a possible score of 10, the mean for the contingent group was 7.50, compared with 4.57 and 5.83 for the noncontingent and no-treatment groups, respectively, $F(2, 36) = 6.55$, $p < .01$. A Tukey test showed that the contingent group was significantly different from

[2] All subjects were not available when each measure was taken. Therefore the degrees of freedom vary from measure to measure.

the noncontingent group ($p < .05$) and the no-treatment group ($p < .05$). The latter two were not different from one another. This finding is especially impressive in light of the fact that the residents were not rewarded for answering these questions; nor were the experimenters who administered the test previously known to them.

Remote Events

Immediately after the probe recall test was administered, subjects were asked questions regarding remote past events (e.g., "When was the Great Depression?" and "How long has television been around?"). Again, significant differences emerged. Responses were coded by an individual who was unaware of the conditions to which subjects had been assigned. The mean number of correct answers for the contingent group was 2.46, compared with 1.53 and 1.54 for the noncontingent and no-treatment groups, respectively. A contrast that set the contingent group as greater than the remaining two groups, which were set equal to one another, was significant, $F(1, 38) = 7.42$, $p < .01$.

Medical Chart Data

The social adjustment comments recorded on residents' medical charts by the nursing home staff were rated on a 7-point scale, in which higher numbers indicated improvement from 1 week before the first visit to 1 week after the last measures were taken; a score of 4 signified no change. These ratings were made by an experimenter who had training as a nurse and who was blind to the purposes of the study and to the groups to which the subjects belonged. The mean score for the contingent group was 5.00, showing some improvement, compared with 3.79 for the noncontingent group and 3.87 for the no-treatment group, both of which showed slight decrements on this measure, $F(2, 40) = 5.41$, $p < .01$. A Tukey test indicated that the contingent group was significantly different from the other two groups ($p < .05$ in both cases), which were not different from one another.

Data on residents' vital signs showed the same pattern as above, although the ratings were not significantly different for the three groups (4.31, 4.29, and 4.21 for contingent, noncontingent, and no-treatment groups, respectively). There were also no reliable differences in the amount of medication taken. However, differences did emerge in regard to the amount of complaints about physical discomfort that residents voiced. The mean rating, on a 7-point scale (higher numbers indicate more complaints), for the contingent group on this measure was 5.07, compared with 3.64 and 3.80 for the noncontingent and no-treatment groups, respectively, $F(2, 40) = 8.14$, $p < .001$. Again, a Tukey test showed the contingent group to be different from the other two ($p < .05$ in both cases), which were not different from one another. As in Langer and Rodin (1976), complaints were taken as a positive sign—when one feels one can have an effect on the environment, complaining makes sense; when one has given up, it does not. Subjects on whom demands were made, in turn, appeared to be more likely to make demands on their environment.

Discussion

Those residents for whom outcomes were contingent upon memory showed cognitive improvement relative to those to whom the same amount of attention was given without contingent rewards for cognitive activity. Increasing the use of cognitive abilities in this way, that is, making remembering matter, had several generalized effects on memory. It improved short-term memory, as evidenced by performance on the probe recall test, improved memory for recent events, as evidenced by better recall of meals and activities, and resulted in greater speed in finding and reporting information that was to be ascertained from the environment. It also significantly improved social adjustment.

It appears that having this responsibility for outcomes was subjectively a positive experience. Residents in the contingent condition overestimated the number of visits they had received and also expressed greater liking for the visitor than did the comparison group.

Conclusion

Both studies provided either interpersonal or practical incentives for becoming more involved with the environment and for greater attention to and remembering of environmental events. We found that restructuring the environment to make it more demanding and then motivating people to increase their cognitive activity seem to lead to improvements in memory that are generalizable. It is critical to note that the interviews in no way provided explicit training for the types of memory tests that constituted the dependent variables of greatest interest; nonetheless, performance for experimental subjects showed reliable improvement.

Perhaps one could argue that the improvement in performance for the high-reciprocal-self-disclosure group in Study 1 and the contingency group in Study 2 was due solely to an increase in motivation to perform well and not to improved ability through increased practice. However, the greater recall over time demonstrated in Study 2 suggests that real improvements did take place. Although this alternative explanation can never be ruled out completely in this or any other study, for that matter, that finds performance differences, if one never has the motivation to exercise one's ability, the motivation/ability distinction becomes a moot point. Although this general issue has been most directly argued with regard to racial differences and IQ (e.g., Loehlin, Lindzey, & Spuhler, 1975), the same can be said of age differences and some cognitive abilities. Again, like findings in which race and IQ are considered, the data from Study 1 suggest that the context in which the measures are given, such as whether they are labeled as memory tests, also affects motivational performance. Taken together, these findings suggest that psychologists must be sensitive to the social–environmental factors that can influence not only performance in the testing context but, ultimately, the desire to practice and perform in general.

Though the present studies were not aimed at determining the extent to which memory loss in the elderly is environmentally rather than biologically determined, the present investigations coupled with other studies examining the effects of environment on intelligence in the elderly (e.g., Plemons, Willis, & Baltes, 1978) suggest that the impact of the environment may be profound. The significance of these factors for memory, and cognitive ability more generally, suggests that studies explicitly designed to test for age-related deficits must control for differences in environmental variables.

Reference Note

1. Langer, E. *Old age: An artifact?* Washington, D.C.: National Research Council, Committee on Biology, Behavior, and Aging, in press.

References

Adamowicz, J. Visual short-term memory and aging. *Journal of Gerontology*, 1974, *31*, 39–46.

Anders, T., & Fozard, J. Effects of age upon retrieval from primary and secondary memory. *Developmental Psychology*, 1973, *9*, 411–415.

Baltes, P. B., & Schaie, K. W. On the plasticity of intelligence in adulthood and old age: Where Horn and Donaldson fail. *American Psychologist*, 1976, *31*, 720–725.

Drachman, D., & Leavitt, J. Memory impairment in the aged: Storage versus retrieval deficit. *Journal of Experimental Psychology*, 1972, *93*, 302–308.

Eysenck, M. Retrieval from semantic memory as a function of age. *Journal of Gerontology*, 1973, *30*, 174–180.

French, J. R. P., Jr., & Raven, B. H. The bases of social power. In D. Cartwright (Ed.), *Studies in social power*. Ann Arbor: University of Michigan, 1959.

Horn, J. L., & Donaldson, G. On the myth of intellectual decline in adulthood. *American Psychologist*, 1976, *31*, 701–719.

Horn, J. L., & Donaldson, G. Faith is not enough: A response to the Baltes–Schaie claim that intelligence does not wane. *American Psychologist*, 1977, *32*, 369–373.

Janis, I. L. (Ed.). *Counselling on personal decisions: Theory and research on short-term helping relationships*. New Haven, Conn.: Yale University Press, in press.

Janis, I., & Rodin, J. Attributions, control, and decision making: Social psychology and health care. In G. Stone, F. Cohen, & N. Adler (Eds.), *Health psychology*. San Francisco: Jossey-Bass, 1979.

Jourard, S. *Self-disclosure*. New York: Wiley-Interscience, 1971.

Langer, E. Rethinking the role of thought in social interaction. In J. Harvey, W. Ickes, & R. Kidd (Eds.), *New directions in attribution research* (Vol. 2). Hillsdale, N.J.: Erlbaum, 1978.

Langer, E., & Rodin, J. The effects of choice and enhanced personal responsibility: A field experiment in an institutional setting. *Journal of Personality and Social Psychology*, 1976, *34*, 191–198.

Loehlin, J. C., Lindzey, G., & Spuhler, J. N. *Race differences in intelligence*. San Francisco: Freeman, 1975.

Plemons, J., Willis, S., & Baltes, P. Modifiability of fluid intelligence in aging: A short-term longitudinal training approach. *Journal of Gerontology*, 1978, *33*, 224–231.

Quinlan, D. M., Janis, I. L., & Bales, V. L. Effects of moderate self-disclosure and subsequent telephone contact. In I. L. Janis (Ed.), *Counselling on personal decisions: Theory and research on short-term helping relationships*. New Haven, Conn.: Yale University Press, in press.

Rodin, J., & Janis, I. The social power of health-care practitioners as agents of change. *Journal of Social Issues*, 1979, *35*, 60–81.

Rodin, J., & Langer, E. Long-term effects of a control-relevant intervention. *Journal of Personality and Social Psychology*, 1977, *35*, 897–902.

Winer, B. J. *Statistical principles in experimental design*. New York: McGraw-Hill, 1971.

Received September 26, 1978 ■

Instructions for Submission of Manuscripts

Papers submitted to the *Journal of Personality and Social Psychology* after April 1, 1979 will be considered under new policies approved by the Publications and Communications Board. Effective with the 1980 volume of the journal, the following policies will apply:

The *Journal of Personality and Social Psychology* publishes original papers in all areas of personality and social psychology. It emphasizes empirical reports but may include specialized theoretical, methodological, and review papers. Beginning in April 1980, the journal will be divided into 3 independently edited sections: ATTITUDES AND SOCIAL COGNITION, INTERPERSONAL RELATIONS AND GROUP PROCESSES, and PERSONALITY PROCESSES AND INDIVIDUAL DIFFERENCES. Authors should submit manuscripts to the appropriate section editor according to the section definitions and manuscript preparation instructions provided on page 468 of the March 1979 issue of the journal.

Section editors' addresses are as follows: ATTITUDES AND SOCIAL COGNITION, Melvin Manis, Editor, Research Center for Group Dynamics, Institute for Social Research, University of Michigan, Ann Arbor, Michigan 48106; INTERPERSONAL RELATIONS AND GROUP PROCESSES, Ivan D. Steiner, Editor, Department of Psychology, University of Massachusetts, Amherst, Massachusetts 01003; PERSONALITY PROCESSES AND INDIVIDUAL DIFFERENCES, Robert Hogan, Editor, Johns Hopkins University, Department of Psychology, 33rd and Charles Streets, Baltimore, Maryland 21218.

Journal of Personality and Social Psychology
1979, Vol. 37, No. 11, 2014-2024

When Practice Makes Imperfect: Debilitating Effects of Overlearning

Ellen J. Langer and Lois G. Imber
Harvard University

It was hypothesized that as overlearning leads to "mindlessness," the individual components of a task become relatively inaccessible to consciousness and therefore unavailable to serve as evidence of task competence. This may lead to a decrement in performance if circumstances, for example, a label connoting relative inferiority, lead one to question one's ability. This was tested in the first experiment by varying practice on a task (no practice, moderate practice, and overpractice) by label assigned to subjects (no label, assistant, boss). As predicted, performance decrements resulted for the no practice and overpracticed subjects who were assigned the inferior status label but not for the moderate practice subjects for whom the task components were still salient. In a second experiment it was found that the debilitation could be prevented for an overlearned task by making components of the task salient. Implications for the vulnerability of experts to these performance debilitations are explored.

In a recent article (Langer & Benevento, 1978) it was argued that decrements in performance may result from negative interpersonal contextual factors, like labels that connote inferiority relative to another person (e.g., assistant vs. boss), regardless of the outcome one experiences. It was suggested that, if salient, the inferior label may lead the individual to an erroneous inference of incompetence, resulting in performance decrements.[1] This effect was termed self-induced dependence. It is self-induced in that individuals who are given low status labels relative to other people, for example, may draw unnecessary inferences from those labels that they then generalize to new situations without any external inducement to do so. Self-induced dependence was originally proposed in contrast to learned helplessness (see Seligman, 1975) where a perception of incompetence is clearly externally induced through prior exposure to uncontrollable aversive outcomes.

If people become dependent on labels when performing tasks with which they have had a good deal of prior experience, as was the case in the Langer and Benevento studies, one might assume that they would be even more vulnerable with respect to tasks with which they have had little practice. However, recent research on the "mindlessness of ostensibly thoughtful action" (Langer, 1978, 1979; Langer, Blank, & Chanowitz, 1978; Langer & Newman, 1979) suggests that the reverse may occur. That is, repeated experience with a task actually may *increase* an individual's vulnerability to an experience such as being provided with a label that connotes inferiority.

This research was supported by a grant from the William F. Milton Fund of Harvard University and by Grant MH32946-01 from the National Institute of Mental Health to the first author.

The authors are grateful to Herb Kelman and Ross Rizley for their helpful comments and to Rob Rubin, Marty Richardson, Howard Botwinick, and Paul Levitt for their help in conducting the studies.

Requests for reprints should be sent to Ellen Langer, Department of Psychology, Harvard University, Cambridge, Massachusetts 02138.

[1] Other interpersonal factors that are hypothesized to result in unnecessary performance decrements include no longer performing tasks that others continue to perform, engaging in a demeaning task, or simply allowing someone else to help you.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3711-2014$00.75

When an individual first approaches a task he/she is necessarily attentive to the particulars of the task. With each repetition of the task, less and less attention to those particulars is required for successful completion of that task. As "mindlessness" is achieved, the components of the task may drop out or coalesce to form a whole. Learning, then, in a sense is learning what elements of the task may be ignored (see Langer, 1978). Repeated practice with the task as a whole rather than with the individual parts may lead to a strange turn of events. The person who has overlearned the task, the expert, may be in a position of knowing that he/she can perform the task, without any longer knowing *how* he or she performs it, that is, without knowing the steps or components that make up the performance. If external factors like labels then led the individual to question his or her competence on these overlearned tasks, the individual would have difficulty supplying information about the solution process as evidence of competence and could erroneously infer incompetence.

The present research was designed to assess this relationship between amount of task experience and vulnerability to an inferior label.

In Phase 1 of the paradigm used, baseline measures are taken as pairs of subjects individually perform a task successfully. In Phase 2 subjects perform a different task together, with one in the role of a "worker" and the other as the "boss." In Phase 3, subjects return to the original task. To test the relationship between vulnerability to external factors like labels and amount of task experience, task experience was also varied in the present investigation. In Phase 1 subjects either were given an opportunity to overlearn a novel task or were given only a moderate amount of practice with that task. In Phase 2, a no practice group was introduced, and all subjects performed a second task under one of three labeling conditions (assistant, boss, no label). In Phase 3 all subjects returned to the original task. A curvilinear relationship between the amount of practice and susceptibility to labels connoting relative inferiority was predicted. Since the vulnerability is presumed to be a function of people losing sight of the intervening steps of the task, it was expected that individuals who had no prior task experience and those for whom the task had become overlearned would show the performance decrement in the inferior label conditions. However, it was predicted that since moderate practice groups would be necessarily attentive to the components of the target task, they would not show the decrement.

Experiment 1

The first study employed a 3×3 design where the variables of interest were label (no label, assistant, boss) and amount of practice (no practice, moderate practice, overpractice).

Method

Subjects. One hundred twenty-six adults were recruited from Boston and New York airport lounges to participate in a study concerned with developing methods to improve task performance. Since there were far more women at the airports than men, either waiting for their own planes or for the departures or arrivals of friends and relatives, we recruited only females.[2] They were informed that for our research we needed people to perform different tasks for us, individually and in pairs. Once 1 subject agreed to participate in the study, a 2nd subject, a stranger to the first, was recruited to be her partner. Thus all subjects were run in pairs, and there were 14 subjects per cell. The experiment was conducted by one female and five male experimenters. Experimenters who administered the tasks were blind to the label assignment.

Task. A novel task was employed so that both extremes of task experience, no practice and overpractice, could be compared. It was also necessary for the task to have several parts so that it could be determined whether, as predicted, overpractice results in obfuscation of the task components. To this end a coding task was devised that involves translating English sentences into a different language where every letter is represented by a corresponding symbol and number. Letters were alphabetically arranged in groups (two groups of nine and one of eight letters); each group had a different symbol (a triangle, circle, or square), and each letter in each group had a number from one to nine (eight). (Punctuation marks had only symbols, rather than symbols and numbers). For example, the letter A

[2] Although the subjects in these studies were all females, there is no reason to assume that the results would not generalize to males. Further investigation is required before this can be determined conclusively, however.

had as the corresponding symbol a triangle and the number 1, and the letter J had as the symbol a circle and the number 1. Subjects were given computer cards on which sentences had been typed, and they were required to transcribe these sentences into the new language. They were to write the symbol below each letter and circle the corresponding number in the column for each letter of the sentence. This task permits subjects to chunk the various elements into fewer and fewer units, thereby facilitating overlearning with repeated practice.

For practical purposes it was important to determine a criterion for overpractice in advance. To that end, the task was pretested on 18 adult female subjects. Subjects were instructed to inform us when they felt as though they could perform the task automatically. After coding two sentences, none of the subjects had overlearned the task. After coding six sentences all subjects were performing the task at least 50% more quickly and more accurately than they had done originally: they no longer referred to the coding reference sheet, their performance reached an asymptote, and by this time they had all reported that the task had become automatic. Therefore, the overpracticed group in the present experiment was given six practice sentences, whereas the moderate practice group was given only two.

Procedure. In Phase 1 of the study subject pairs were randomly divided into three groups: no practice, moderate practice, and overpractice on the coding task. After the task was explained and demonstrated, the overpracticed group was instructed to code six sentences, the moderate practice group was instructed to code two sentences, and the no practice group did no coding. This group went immediately to Phase 2 of the study. Each sentence appeared on a separate computer card. While one experimenter administered this task, a second experimenter, who was hidden from view, recorded the amount of time it took subjects to complete the first sentence. The amount of time it took to complete this sentence and the number of errors made were used as baseline measures for the moderate and overpractice groups.

All subjects now participated in Phase 2 of the research. They were each told that the task they would be asked to perform next required a joint effort: One of them would be the assistant and one the boss (one-third of the pairs performed this task without labels indicating relative status). Subjects were asked to examine a picture and find objects that were hidden in it. One person was to find the objects, and the other was to record the objects found. To control for task effects, for half of the pairs the boss did the searching and the assistant the recording, whereas for the remaining half the tasks were reversed. The experimenter handed subjects the envelope with instructions to place the task materials back in the envelope when they had finished. At that time she/he would return to give them further instructions. In the meantime she/he was purportedly going to recruit two more subjects. Subjects chose a slip of paper from the envelope that indicated both whether they would be in the boss, assistant, or no label condition and what their task would be. When the second experimenter, whose presence was unknown to the subject, observed that the task had been completed she/he signaled the first experimenter, who then returned to complete the final phase of the experiment and thus was blind to the labeling condition.

Table 1
Mean Number of Minutes to Complete the Postlabel Coding Task

Condition	Label		
	Assistant	No label	Boss
No practice	7.10_a	5.50_b	5.40_b
Moderate practice	4.50_b	4.54_b	4.72_b
Overpractice	4.40_b	3.28_c	3.10_c

Note. All cell ns = 14. Cells bearing different subscripts are significantly different from each other: subscript a differs from all other cells at $p < .001$, and subscript c differs from b cells at $p < .05$.

In the last phase of the study, subjects were told that an individual effort rather than a joint effort again was required. All subjects then were asked to code another sentence. (This, of course, was the first sentence for the no practice group.) Once again the second experimenter timed their performance. Both speed and accuracy served as dependent measures.

At the end of the task, all subjects were asked to complete a brief questionnaire. They were asked to make a list of the components of the task. Specifically, subjects were told, "To do the coding task, certain steps were involved. For instance, you had to look at the letter to be coded. Please list as many steps as you can think of which you used to perform the coding task." When subjects completed this questionnaire, they were completely debriefed.

Results and Discussion

Before examining the effects of the label, it is important to verify group comparability at the start of the experiment. Therefore scores for all of the groups that performed the prelabel task were compared. As expected, the analysis of variance of prelabel scores (the amount of time it took subjects to code the first sentence and the number of errors they made) showed no differences among the six groups for which this measure was relevant: the no label, assistant, and boss, moderate practice and overpracticed groups. Speed scores ranged from 5.56 to 5.89, and accuracy scores ranged from 2.36 to 3.07.

It was expected, however, that differences would emerge on postlabel measures. A 3 (No Label, Assistant, Boss) × 3 (No Practice, Moderate Practice, Overpractice) analysis of variance was performed on both the speed and accuracy of the postlabel coding task. The analysis performed on speed scores yielded highly significant main effects for practice, $F(2, 117) = 133.82$, $p < .001$, and label, $F(2, 117) = 25.05$, $p < .001$, as well as a highly significant interaction, $F(4, 117) = 3.96$, $p < .001$.[3] Table 1 presents the mean number of minutes it took to complete the postlabel coding task for each group. A Newman–Keuls test was performed to understand better the relationships among these means and thus the nature of these significant effects. Cells bearing different subscripts in the table are significantly different from each other by this test. As may be gleaned from the table, subjects in the no practice/assistant condition took longer to complete the postlabeling task than did subjects in all other conditions ($p < .001$), whereas subjects in the no label/overpracticed and boss/overpracticed group performed most quickly ($p < .05$). It is, of course, not surprising that subjects who have had enough practice to overlearn a task perform that task faster than subjects do who have had very little practice. What is interesting is the comparison of the assistant and the no label groups within each practice condition. As predicted, the label *assistant* had a detrimental effect for both the no practice and the overpracticed groups but had no effect on the moderate practice condition.

The analysis performed on the number of errors made by subjects on this task yielded very similar results. The analysis yielded highly significant main effects for practice, $F(2, 117) = 12.02$, $p < .001$, and for label, $F(2, 117) = 35.85$, $p < .001$, and a highly significant interaction, $F(4, 117) = 9.80$, $p < .001$. The mean number of errors made by each group is shown in Table 2. A Newman–Keuls test revealed that the means that bear different subscripts in the table are significantly different from each other. Once again, in contrast to the moderate practice group, the label assistant seems to have had a detrimental effect for both the no practice and overpracticed groups. Subjects in both of these groups made at least twice as many errors as subjects in the other conditions did. Although 86% of the overpracticed/assistant group showed a decrement in performance from Phase 1 to Phase 3, only from 7% to 14% of the remaining relevant groups performed more poorly, $\chi^2(2) = 6.08$, $p < .05$.

Table 2
Mean Number of Errors Made in Postlabel Coding Task

Condition	Label		
	Assistant	No label	Boss
No practice	5.35_a	2.36_c	2.00_c
Moderate practice	1.64_c	1.93_c	1.50_c
Overpractice	4.93_b	1.07_c	1.14_c

Note. The cell bearing subscript a differs from those bearing subscript c at $p < .001$. Cell b differs from cells c at $p < .01$. Cells a and b do not differ from each other.

Although decrements were expected on the target task, subjects in the assistant groups should not show a decrement in performance on the intervening hidden objects task because this task has been defined as one that is consistent with their label, that is, it is an "assistant" task. The analysis of these scores in fact revealed no group differences. This further suggests that the assistant subjects took the task as seriously as the other groups of subjects did (see Langer & Benevento, 1978).

Although subjects in all assistant groups apparently took the label seriously, only those who had no prior experience with the target task and those who were overpracticed on it were debilitated by the inferior status label. It was hypothesized and indeed found that this label would not adversely affect subjects in the moderate practice group. This group was expected to be protected from the potentially debilitating effects of the label

[3] Because subjects were run in pairs, there may be a loss of degrees of freedom. A more conservative test that uses the degrees of freedom based on the number of pairs within groups shows that each of the *F*s in this study are still significant at the $p < .001$ level.

because their experience with the task, in contrast to that of the no practice group, has taught them the components of the task, but not so much as to obscure them, in contrast to the overpracticed group. After performing the last task, subjects in each group were asked to list all the components or steps they could think of that were involved in performing the coding task. The mean number of steps listed for the no practice and overpracticed groups ranged from 3.07 to 4.43, whereas the means for the moderate practice groups ranged from 6.07 to 6.21. The analysis of variance revealed a significant main effect for the amount of practice, $F(2, 117) = 30.95$, $p < .001$. A Newman–Keuls test confirmed the prediction: The moderate practice groups listed significantly more steps than either the no practice or overpractice groups, which were not different from each other.

If the components of a task do indeed become obscured with practice, this should be revealed not only in the number of steps listed but also in their degree of specificity. To assess this, each subject's list was rated by two blind raters ($r = .92$) on a 10-point scale that ranged from very specific to very global. The analysis revealed a significant group difference, $F(2, 123) = 31.38$; $p < .001$. A Newman–Keuls test showed that the responses of the no practice group ($M = 7.5$) were equivalent in specificity to the overpracticed group ($M = 6.57$) and that both were far more global than the moderate practice group ($M = 3.6$). Typical steps listed by the no practice and overpracticed groups were "Look at the letter and write in the code." In contrast, typical steps listed by the moderate practice group included such things as "Fill in the circle if it is a capital and check for punctuation." Of course, the more specific one is, the more steps are available to be listed. Therefore, although both measures are interesting, they should not be viewed as independent of one another.

The performance decrements found in this study occur for tasks for which the components are not salient (a novel task and an overlearned task), after the individual had experience in an inferior status position. Therefore one may ask whether these decrements can be prevented by making components of the task salient. The next study was undertaken with this in mind.

Experiment 2

This experiment used basically the same paradigm as that described in Experiment 1. Subjects in Phase 1 performed a task successfully, in Phase 2 they performed a different task in one of three label conditions (assistant, no label, boss), and in Phase 3 they returned to the original task. In this experiment, however, only one level of practice was used. A task was selected that subjects, in all likelihood, already performed automatically. To see whether the debilitation could be prevented by making components of the task salient, task components were listed for half of the subjects before they began Phase 1 of the experiment. Thus the study utilized a 2 (Components Salient/Nonsalient) × 3 (Assistant, No Label, Boss) design. It was predicted that relative to the no label condition, the assistant condition would show a decrement in performance, and the boss condition would either show an increment in performance (as in the Langer & Benevento, 1978, study) or would be equivalent to the no label group (as in Experiment 1 reported in this paper). It was predicted, however, that the performance decrement would occur only in the components-nonsalient condition. That is, it was expected that the salience manipulation would prevent the debilitation, since these subjects now would have a set of task components recently used and verified to supply as evidence if they questioned their ability.

Method

Subjects. Seventy-two adult females were recruited to serve as subjects from the lounges of Logan Airport in Boston. They were asked to participate in a study on developing educational methods where subjects would be asked to perform different tasks either individually or with another person. Once one subject agreed to participate, a second person, who was a stranger to her, was recruited to be her partner. Thus, as in Experiment 1, subjects were run in pairs.

Task. The target task was a proofreading task. This was selected because it is a task that almost

ll literate people have performed—for example, hen reading one's term paper for school, in eading over a letter before sending it, and so orth—and because it is a task where many rules nay be employed without the individual's neces- arily knowing what they are. That is, the indi- idual can locate errors without being able to rticulate why the error is an error. Two articles ere chosen from popular women's magazines for his task. Their titles accurately reflect their con- ent: One was entitled "Is your doctor overcharging ou?" and the other was "Energy savers' guide." The articles were edited so that each contained 40 rrors distributed evenly throughout each article. There were 13 spelling errors, 9 errors of capitaliza- ion, 12 punctuation errors, and 6 grammatical er- ors in each article.

Procedure. All subjects were initially told that one task found helpful for developing educational methods is proofreading.

> Virtually everyone has done this task although they do not necessarily call it proofreading. For example, if you ever read over a paper for school before handing it in, you were proofreading. Or if you ever read over a letter you wrote before sending it to see if it said what you intended it to say, you were proofreading. Therefore we would like you to proofread a story for us.

Components-salient condition: For this condition the experimenter went on to say,

> Most people can do this task even though they are not aware of just how they do it. That is, they are not always aware of the rules they are using. For instance, if you knew the word "susan" should be typed "Susan," you would be using the rule that the first letter of proper nouns is always capitalized. We would like you to make a list of *at least* three things which you would need to know in order to correct a story for errors.

After subjects completed their lists, they were told that we would like them to remember these rules when they read the story we were about to give them to proofread.

> When you find an error, try to recall the rule you used, and then circle the error. You are only given 3 minutes to read the story, so do not expect to finish it. Most people who do this task can find six errors in 3 minutes. If you are able to do this, you should consider yourself successful at this task.

The time allotted for this task was intentionally brief in order to prevent overlearning of the components.

Components-nonsalient condition: Subjects in this condition were also told that proofreading is a task that most people can do. In addition, they were told,

> In order to proofread a story for errors, you should at least glance at each word in the story to see if it is correctly written. For instance, if the word "susan" appeared, you would know it is an error because the word should read "Susan." First, we would like you to make a list of at least three ways in which proofreading is used. We would like you to read this story and circle any errors that you find. You are only given 3 minutes to read the story, so do not expect to finish it. Most people who do this task can find six errors in 3 minutes. If you are able to do this, you should consider yourself successful at this task.

Thus both groups were led to expect success on the task. Instructions encouraged both to pay attention to the story, both were given an example of an error, and both were asked to make task-relevant lists. The difference was that rules for determining errors were generated only by the components-salient group. The task was pretested on a similar group of subjects to ensure that all experimental subjects would be successful at the task, that is, that each subject would find at least six of the errors.

After subjects successfully completed this task, they were told that the second task required a joint effort. One person was to solve a cryptogram task while the other person used the stopwatch provided to keep track of the solution times. The cryptogram task consisted of four lists of five words each that were written in number form. Each number had a corresponding letter (A = 1, B = 2, and so on). The subjects' task was to transcribe the numbers into their alphabet equivalents with the use of a code. Two-thirds of the subject pairs from each of the two conditions described above were told that the task thus "required" someone to be the boss and someone to be the assistant. One-third of the subject pairs performed this task without reference to relative status (no label condition). Task effects were controlled for by varying the label assigned to the particular task (timing vs. solving). That is, for example, for half of the bosses the task was solving and for half the task was timing. After one subject in each pair had been asked to indicate the task she wanted to perform, the pair was informed which task was the boss task and which was the assistant task. Subjects worked together on the cryptogram task for approximately 5 minutes and then were informed that their joint participation was concluded. All subjects then were asked to proofread another story. As in Phase 1 of the study, they were given 3 minutes to locate as many errors in the story as they could. These scores, in comparison with the prelabel proofreading scores, comprised the primary dependent measure.

At the end of the proofreading task, subjects were asked to complete a brief questionnaire. They were asked to make a list of the components of the

Table 3
Mean Number of Postlabel Proofreading Errors Correctly Located

Condition	Label		
	Assistant	No label	Boss
Components salient	14.33_a	13.99_a	14.41_a
Components nonsalient	8.92_b	13.00_a	17.42_c

Note. All cell *n*s = 12. Cells bearing different subscripts are significantly different from each other: subscript b differs from all other cells at $p < .005$, and subscript c differs from all a cells at $p < .05$.

proofreading task. Specifically, subjects were asked to "make a list of rules that you know about writing which helped you to identify errors in the stories." The time it took each subject to make the list was recorded. When subjects had completed the questionnaire, they were completely debriefed.

Results

An analysis of the prelabel scores revealed no difference among the groups. Thus there was group comparability at the start of the experiment. Although the prelabel scores were equivalent across the groups (ranging from 13.33 errors found to 14.67), a very different picture emerges when one examines the postlabel scores. Table 3 shows the mean number of proofreading errors located by each of the six experimental groups. The analysis of these scores yielded a highly significant main effect for label, $F(2, 66) = 8.09$, $p < .001$, and a highly significant interaction, $F(2, 66) = 7.74$, $p < .001$.[4] A Duncan test was computed to reveal the nature of this interaction. As may be ascertained from the subscripts in the table, the major hypothesis of the study was supported. The components-nonsalient group replicated the Langer and Benevento (1978) finding such that, relative to the no label condition, the assistant group showed a severe decrement in performance, whereas the boss group showed a facilitation effect. However, the salience manipulation was successful in wiping out the differences among the groups. The scores for the components-salient condition are virtually identical to the prelabel scores. The correlation for the components-salient groups between their performance on the prelabel and postlabel stories was .83. Similarly, the correlation between prelabel and postlabel stories for the nonlabeled subjects in the components-nonsalient group was .95. Although these groups found as many errors in the second story as in the first, the assistants in the components-nonsalient group found on the average 5.75 *fewer* errors in the second story. The bosses in this group found 2.67 *more* errors on the average than they had in the prelabel story.

Of those subjects in the components-nonsalient/assistant group, 92% showed a decrement from Phase 1 to Phase 3. In contrast to this only 25%–33% of the remaining groups showed a decrement in performance. Both 2×2 comparisons are significant [salient/nonsalient by assistant/boss: $\chi^2(1) = 4.48$, $p < .05$; salient/nonsalient by assistant/no label: $\chi^2(1) = 3.89$, $p < .05$].

To test again that subjects who were labeled *assistant* took the task in Phase 2 seriously, scores for the intervening cryptogram task were compared. As predicted, there were no differences among the groups. Assistants solved the puzzles as quickly as the remaining groups. (Solution times ranged from 43.18 sec to 47.37 sec).

Further evidence that the salience manipulation was responsible for restoring subjects to their prelabel performance comes from the number of rules subjects reported using in the proofreading task and the time it took them to compose this list. Subjects in the salience condition, if they were keeping the rules in mind while performing the task as instructed, should have been able to list more rules more quickly than the components-nonsalient group could. The analysis of the number of components listed (salient $M = 3.56$ vs. nonsalient $M = 2.78$, $t = 2.98$, $p < .005$) and the analysis of time to compose these lists (salient $M = 44.56$ vs. nonsalient $M = 73.81$, $t = 6.39$, $p < .001$) suggest that this was the case.

One might argue that subjects' ability to generate components of the activity both

[4] An analysis using the degrees of freedom based on the number of pairs within groups shows that this more conservative test lowers the p value for these measures to .01.

Table 4
Between-Subjects Correlation Coefficients for Rules Listed at End of Experiment (Yes/No) by Proportion of Errors Found

Type of rule	Errors found/total errors 1	2	3	4
Components-salient condition				
1. Spelling	.69* (.72)	−.40 (.08)	−.20 (−.08)	.22 (.15)
2. Capitalization	−.30 (.01)	.68* (.60)	−.01 (.15)	−.40 (−.32)
3. Punctuation	.00 (.14)	−.01 (−.10)	.67* (.70)	−.46 (−.49)
4. Grammar	.37 (.34)	−.05 (−.18)	−.34 (−.26)	.82* (.63)
Components-nonsalient condition				
1. Spelling	−.35	.07	−.13	.01
2. Capitalization	.09	−.13	.04	.14
3. Punctuation	−.05	.10	−.16	−.14
4. Grammar	.34	.13	.21	−.25

Note. All cell *n*s = 36. Numbers in parentheses represent correlations between rules listed at the start of the experiment by proportion of errors found.
* $p < .001$.

before and after the task suggests that counter to the position presented here, the components are not inaccessible. However, although subjects in the components-nonsalient condition listed rules after the task, that does not mean that they used those rules (i.e., identified errors as a function of realizing why they were errors) while performing the proofreading task (see Dweck & Gilliard, 1975). In fact, it is important to demonstrate that they did *not* use the rules, whereas the components-salient group *did*. To do this, the next analysis was conducted to see the relationship between the ability to list certain rules and the actual use of those rules. Since all subjects made a list at the end of the last proofreading task, this list was used for the first comparison. Table 4 shows the correlation coefficients for the type of rule listed by the proportion of errors found for each type of error. Correlation coefficients bearing asterisks are significant at $p < .001$. As may be seen in the table by looking at the diagonals, the relationship between listing and finding particular errors was significant for the components-salient group, but there was no such relationship for the components-nonsalient group. That is, for example, subjects who listed spelling found more spelling errors than subjects who did not list spelling in the components-salient condition, and this was not the case for the components-nonsalient group.

As may be seen in the top portion of Table 4, the correlations between rules listed at the beginning of the experiment and the proportion of errors found (these correlations are reported in parentheses in the table) are also significant at $p < .001$. This strongly suggests that the salience manipulation was effective in providing subjects with rules to use while they performed the proofreading task.

To explore further the relationship between rules listed and errors found, a second correlational analysis was performed on *within*-subjects scores. This analysis yielded a coefficient that represented, for every subject, the relationship between rules listed and the proportion of errors of each type that were found. These correlation coefficients were then transformed into Z scores, and the resulting analysis yielded a significant effect for salience, $F(1, 66) = 71.12$, $p < .001$. The mean correlations between the types of rules listed and the types of errors found ranged from −.29 to −.01 for the components-nonsalient conditions and from .77 to .86 for the components-salient groups. Clearly there was a strong relationship between rules listed and the rules

used for the components-salient group that was absent for the components-nonsalient group.

Discussion

Both experiments provide strong support for the assumption that overlearning may lead to increased vulnerability to labels that connote relative inferiority. In the paradigm used, the effect of the label reveals itself in the final phase of the experiments. Subjects in the second phase of both studies are performing what is deemed by the label *assistant* to be a low status task. There is no reason, therefore, for them to question their ability to perform it. However, the original and final tasks are not assistant tasks. To the extent that subjects accept the validity of the label, they may question their competence with regard to tasks that may now be perceived as psychologically inconsistent with that label. Such questioning will not necessarily have negative consequences, if the individual can convince himself or herself that he or she in fact can perform the task. This evidence must be lacking with respect to a novel or unfamiliar task. If, as the present studies suggest, overlearning results in the obfuscation of the individual components of the task, evidence also will be lacking in this case. Therefore only groups with moderate experience should be able to proceed with the task unhampered by the label. Experiment 1 showed this to be true. Performance decrements resulted after the label assignment for the no practice and the overpracticed groups but not for the moderate practice groups.

Since the overpracticed group was able to generate task components when asked to do so, it is important to consider why this information did not protect them from the label's influence. The present research suggests two reasons for this. First, the overpracticed subjects did not list very specific task components. One might reasonably assume that global knowledge of what a task consists of should not be very confidence inducing. If one wanted to teach the task to someone else, for example, these global components alone would be inadequate. Second, the untreated overpracticed group in Experiment 2 showed no relationship between the steps generated and the steps they actually used in performing the task. Being able to list components that *could* possibly be used to complete a task should not be as confidence inducing as knowing steps that *were* experienced successfully.

The second experiment provides evidence that the decrement in performance that may occur on overlearned tasks may be prevented. It will be recalled that this experiment compared two overpracticed groups: a group that, except for the tasks used, basically was treated just as the overpracticed group in the first experiment had been and a second group for whom the components of the task were made salient. The performance decrement occurred again for subjects bearing the inferior label in the untreated group. Making the components of the task salient, however, apparently inoculated the comparison group of subjects against the potential effects of the label.

One may question how the salience manipulation in Experiment 2 could have been effective in light of the assumption, for which all of us can summon anecdotal support, that renewed attention to that which is overlearned is disruptive (cf. Kimble & Perlmutter, 1970; Langer, 1978). The salience manipulation made subjects attend to an overlearned task, yet their performance was not hampered. The reason for this is likely to lie in the fact that subjects were asked to generate steps necessary to perform the task *before* performing it. Subjects in the components-salient group showed a strong relationship between the components they generated initially and the components they attended to while performing the task, whereas subjects in the group left to their own devices (the components-nonsalient group) showed no relationship between the components they generated afterward and the components they attended to while performing the task. This suggests that the salience manipulation led subjects in some sense to follow a new set of rules for themselves. Therefore, if they were approaching the task with a new set of rules (thereby changing the task to a nonoverlearned task), attention to those rules (which was revealed in the high correlation

etween the rules cited and the ones actually sed) should not be disruptive. And of course erformance on an overlearned task also ould not be disrupted if it simply were per-ormed automatically without bringing com-etence into question. The proficiency mani-ested by subjects in the overpracticed/no abel and boss groups in Experiment 1 bears his out. Thus although negative external actors like pejorative labels may be quite ebilitating, they, of course, are not neces-arily so. What about positive external fac-ors? The data from the present studies do ot permit any clear conclusions with respect o the label "boss." More research on situa-ional and individual difference variables is ecessary before any conclusions about posi-ive labels can be drawn.[5]

In addition to exploring the potentially acilitating effects of positive external influ-nces like labels, additional research is also equired to understand more fully the rela-ionship between particular task character-stics and performance debilitation of this ort. The present analysis would suggest that asks that have only one step, tasks whose erformance requires the articulation of steps (e.g., learning the alphabet), or tasks that are so complex as to preclude complete mas-tery would not render the individual vulnera-ble to external influences. Vulnerability may be evident for many, if not most, other tasks, however.

It is clear that overlearning is adaptive, since it frees limited attention to be paid elsewhere. It would seem from their daily interactions that most people in the world are aware that overlearning serves this function, since they appear to be constantly adding familiarity, predictability, and structure to their lives, which facilitate overlearning. While acknowledging the adaptive function mindlessness or overlearning may serve in freeing limited capacity, the present article supports previous work (Langer, 1978, 1979; Langer, Blank, & Chanowitz, 1978; Langer & Newman, 1979; Langer, Note 1; Langer & Weinman, Note 2) in demonstrating the ways in which mindlessness may also be mal-adaptive.

Although the detrimental effects of over-learning may be pervasive, it is also likely that some subject populations are more susceptible to these effects than others are. For example, the elderly are a group who for many reasons (see Langer, 1978, 1979, Note 1) probably have more experience with overlearned tasks than the nonelderly do. Children probably have more experience with no practice tasks relative to nonchildren. Thus these two groups should be most susceptible to external factors, such as pejorative labels, that may lead one to question one's competence. Although the groups may look similar, they are of course quite different in that the elderly can in fact do the tasks in question. Similarly, people who chronically occupy low status jobs are likely to be particularly susceptible. However, most interesting, perhaps, would be the inclusion of experts on this list.

Although one probably does not want to discourage overlearning or the attainment of mindlessness because of its general advantage in fast and efficient performance, educating people as to the potential side effects would seem fruitful. However, the present research suggests another, more direct way in which potential performance decrements may be prevented. Stated generally, focusing on process

---

[5] In the original Langer and Benevento studies, the label "boss" had a facilitating effect such that relative to the no label conditions, these groups performed significantly better. Although this was also true for Experiment 2 in the present investigation, it was not true for the first experiment. Here subjects labeled *boss* within each level of practice were equivalent to the no label groups. The fact that the "boss" group across all four studies did not show a decrement in performance suggests that the majority of subjects in this condition are not questioning their ability to perform the task, for such questioning for the overlearned and no practice groups would reveal to them an inability to find evidence to support the label. If one goes back to the original studies (Langer & Benevento, 1978), one finds that whereas the boss groups on the average performed better than the no label groups, the proportion of subjects who performed this way is the same for both groups. Although most subjects are unquestioningly accepting the label and trying to live up to it, it would appear that a few are not so accepting. The relative proportions from experiment to experiment of this latter minority group would determine whether the aggregate measures reveal equivalence, facilitation, or perhaps even debilitation.

rather than on outcome may reduce vulnerability to debilitation of this kind.

Reference Notes

1. Langer, E. J. *Old age: An artifact?* Washington, D.C.: National Research Council, in press.
2. Langer, E. J., & Weinman, C. *Mindlessness, confidence, and accuracy*. Unpublished manuscript, Harvard University, 1979.

References

Dweck, C. S., & Gilliard, D. Expectancy statements as determinants of reactions to failure: Sex differences in persistence and expectancy change. *Journal of Personality and Social Psychology*, 1975, *32*, 1077–1084.

Kimble, G., & Perlmutter, L. The problem of volition. *Psychological Review*, 1970, *77*, 361–384.

Langer, E. J. Rethinking the role of thought in social interaction. In J. Harvey, W. Ickes, & R. Kidd (Eds.), *New directions in attribution research*. Hillsdale, N.J.: Erlbaum, 1978.

Langer, E. J. The illusion of incompetence. In L. Perlmutter & R. Monty (Eds.), *Choice and perceived control*. Hillsdale, N.J.: Erlbaum, 1979.

Langer, E. J., & Benevento, A. Self-induced dependence. *Journal of Personality and Social Psychology*, 1978, *36*, 886–893.

Langer, E. J., Blank, A., & Chanowitz, B. The mindlessness of ostensibly thoughtful action: The role of "placebic" information in interpersonal interaction. *Journal of Personality and Social Psychology*, 1978, *36*, 635–642.

Langer, E. J., & Newman, H. M. The role of mindlessness in a typical social psychological experiment. *Personality and Social Psychology Bulletin*, 1979, *5*, 295–298.

Seligman, M. E. P. *Helplessness*. San Francisco: Freeman, 1975.

Received November 20, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 11, 2025-2043

Idiosyncratic Weighting of Trait Information in Impression Formation

Thomas M. Ostrom
Ohio State University

Deborah Davis
University of Nevada, Reno

An analysis is offered of the role of unequal weighting in the averaging model of information integration. A distinction is made between unequal weighting at the normative level (which has been referred to as "differential weighting") and unequal weighting at the level of the individual subject (which we refer to as "idiosyncratic weighting"). Two studies are reported that examine the prevalence of idiosyncratic weighting in the trait-judgment impression formation task. Whereas most past research on the question of unequal weighting in this task involved averaging responses across both subjects and stimulus replications, the present studies were analyzed at the level of an individual subject's repeated responses to separate stimulus replications. Clear evidence of idiosyncratic weighting was obtained from about 50% of the 120 subjects; only 20% of the subjects indicated absolutely no tendency toward unequal weighting. There was no evidence that idiosyncratic weighting was restricted to just a subset of stimuli, since all of the 20 stimulus replications showed idiosyncratic weighting effects. In contrast to previous findings, negative traits did not always receive more weight than positive traits. In more than 20% of the instances of unequal weighting, the more positive trait was accorded a higher weight.

Information integration theory (Anderson, 1974) offers an approach for understanding how people combine stimulus information when making judgments and decisions. The theory seeks to determine the nature of the integration rule (e.g., adding, averaging, multiplying) employed by people in various response domains. In addition, it provides a way to determine whether all stimulus items in the domain contribute equally to the overall judgment or whether they carry different weights.

Integration theory provides no a priori basis for predicting which integration rule or weighting assumption is correct for any particular response domain. It does, however, provide an array of conceptual alternatives along with a diagnostic methodology (functional measurement). With a comprehensive series of studies, it is possible to determine which conceptual alternatives best describe a response domain. Hence, one of the great strengths of integration theory is its ability to uncover different integration rules for different response domains.

In order to apply integration theory to a response domain, it is necessary to specify two features of that domain, the population of information or stimulus items and the nature of the subjective judgment continuum. It is necessary to specify these two features because a shift in either may affect the integration rule or parameter values. For example, one integration rule may apply when traits are combined with traits (e.g., averaging), and a different rule might apply when traits are combined with adverbs (e.g., multiplication). Parameter values may also vary as the nature of the judgment continuum shifts. Whereas *friendly* may have a higher scale value than

This research was supported by National Science Foundation Grant GS-38604. The authors are grateful to Sarah Boysen for her assistance with collection of the data.

Requests for reprints should be sent to Thomas M. Ostrom, 404C W. 17th Avenue, Columbus, Ohio 43210.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3711-2025$00.75

energetic on the favorability continuum, it is likely that *energetic* would have the higher scale value on an activity continuum.

Integration theory has been used extensively to analyze the adjective combination task first introduced by Asch (1946) to study impression formation processes. The stimuli used in the impression formation task are personality trait adjectives, and the response continuum refers to overall impression favorability. At the operational level, most investigators use the 555 traits scaled by Anderson (1968) as their domain of stimuli. The judgment scale usually has seven or more intervals and is anchored with the bipolar terms of favorable–unfavorable or like–dislike.

It is widely assumed that for the impression formation response domain, people employ an averaging integration rule and assign approximately equal weights to all stimuli in the stimulus domain. Anderson has offered this view of an equal weighting model in a number of sources (Anderson, 1974, p. 261; Anderson, 1976, pp. 681–682; Anderson & Lopes, p. 69; Leon, Oden, & Anderson, 1973, pp. 301–302; Oden & Anderson, 1971, p. 159). The main objective of the present article is to examine the assumption that all personality trait adjectives, as items of person information, carry approximately equal weight in the impression formation task.

This assumption of equal weighting has been questioned by some investigators (e.g., Birnbaum, 1974). In the following review of research bearing on the equal weighting question, we note that existing data are not conclusive regarding the source or strength of differences in trait weights. One of the problems with this research is that most of the studies are based on group averages. This paper explores the possibility that such group data may "average out" idiosyncratic differences in trait weights that exist separately for each individual.

Differential Weighting

Two kinds of deviations from equal weighting have been examined in previous research; one is related to the lability of a trait's weight, and the other concerns the variation in weight between different traits.

Weight lability. The lability of a trait's weight reflects the extent to which the weight is affected by situational or contextual variables. There can be little question as to whether "differential" weighting occurs in the impression formation domain in this sense of the term.

Characteristics of the judgment task are known to affect the functional weight given any particular item. For example, communicator credibility (Rosenbaum & Levin, 1968) and serial position (Anderson, 1965) can act as situational determinants of trait weight.

Weight can also be influenced by the semantic relationship that a trait holds with other items in the set. The weight parameter is affected by the level of redundancy (Schmidt, 1969) and inconsistency (Anderson & Jacobson, 1965) among traits describing the same person. Also, there are traits that possess two distinct semantic meanings (polysemous homographs)—traits such as *discriminating* and *sensitive*—that may change their scale value as well as their weight, depending on other stimuli in the information set. Although each of the above features can lead to differential weighting, they all derive from the relationship between the semantic definitions of the different traits and can in principle be specified a priori. They do not represent instances in which particular traits carry different weights when in isolation from one another.

Weight differences between traits. The weights of most, if not all, traits are clearly labile. However, this does not mean that traits differ in their natural or context-free weights. When variables known to influence lability are held constant, it is possible that trait weights are either approximately equal or substantially different from one another. It is this sense of the term *differential weighting* that is being rejected when the integration process in impression formation is said to conform to an equal-weight averaging model.

Several possible determinants of trait weight have been examined. Both trait novelty (Wyer, 1970) and trait ambiguity (Kaplan, 1971, 1975; McKillip, Barrett, & Dimiceli, 1978; Schümer, 1973; Wyer, 1974b; Wyer & Watson, 1969) have received some attention.

In neither case, however, do the findings unequivocally support the conclusion of differential weighting in the integration process.

Other research has been directed toward sources of differential weighting that are associated with trait scale value. It is possible, for example, that traits with extreme scale values carry more weight than do neutral traits (e.g., Warr, 1974; Warr & Jackson, 1975). These findings are difficult to interpret, however, because of confounding between negativity, ambiguity, and extremity (Warr & Jackson, 1976; Wyer, 1973).

Unlike the above potential sources of differential weighting, research findings have unequivocally established the existence of a negativity effect. Traits with a negative valence tend to carry more weight than those with positive valence (e.g., Birnbaum, 1974; Hodges, 1974; Oden & Anderson, 1971). This negativity effect is thought to exist for a number of response domains (Kanouse & Hanson, 1972).

Although the negativity findings represent a genuine limitation to the equal weighting assumption, they may not be especially serious. Anderson (1974, p. 261) has speculated, for example, that the magnitude of this negativity effect is not very large and that it may be restricted to only extremely negative stimuli. Another difficulty with this research is that it only employed a limited sample of positive and negative traits. No attempt was made to see how representative the sample was of the entire trait population. The possibility exists that most traits have equal weights, but that a few negative traits carry a higher weight and a few positive traits a lower weight. One objective of the present research was to examine these questions about the negativity effect.

Idiosyncratic Weighting

Most work on the issue of differential weighting has been at the level of *normative* weights. That is, the research has examined how much weight is given traits by the typical or average subject. Little consideration has been given to the problem of whether or not equal weighting applies to each person individually. It is possible that because of a person's idiosyncratic experiences with different trait adjectives, he or she may assign widely differing weights to the adjectives. However, one person's pattern of weights may be unrelated to another's. One person may weight *friendly* twice as much as *intelligent*, whereas another might give *friendly* only half the weight of *intelligent*. If we were to obtain normative weights by averaging over these two persons (as is the case in most differential weighting research), we would find that the two traits received the same weight. Differential weighting of traits at the level of a single subject (what we term *idiosyncratic weighting*) may emerge as equal weighting when measured at the normative level.

Only one published study has examined the possibility of idiosyncratic differences in trait weights. In this study (Anderson, 1962), the stimulus sets consisted of three traits, one from each of the three factors in the design. Each factor consisited of a negative, neutral, and positive trait to form its three levels. This $3 \times 3 \times 3$ design resulted in a total of 27 stimulus sets. Subjects judged all 27 sets once a day for 5 days. The resulting data were analyzed to determine whether significant interaction variance emerged. This design was repeated for 12 subjects, with every 2 subjects receiving a different stimulus replication (i.e., six stimulus replications of the three-factor design, with 2 subjects per replication). The test for differential weighting (or nonadditivity) was performed on the pooled (i.e., 20 degrees of freedom) interaction. Of the 12 subjects, 3 were found to produce significant nonadditivity.

Even though the total number of subjects was small, this study suggests that about 25% of the people show differential weighting. That percentage figure would, no doubt, have been higher had each of the two-way interactions and the three-way interaction been separately tested for each subject. It would be possible for a subject to have a significant two-way interaction but not to have the pooled interaction reach significance.

On the other hand, it is possible that the 25% figure is actually an overestimate of the number of people displaying differential weighting. Two of the three subjects with sig-

energetic on the favorability continuum, it is likely that *energetic* would have the higher scale value on an activity continuum.

Integration theory has been used extensively to analyze the adjective combination task first introduced by Asch (1946) to study impression formation processes. The stimuli used in the impression formation task are personality trait adjectives, and the response continuum refers to overall impression favorability. At the operational level, most investigators use the 555 traits scaled by Anderson (1968) as their domain of stimuli. The judgment scale usually has seven or more intervals and is anchored with the bipolar terms of favorable–unfavorable or like–dislike.

It is widely assumed that for the impression formation response domain, people employ an averaging integration rule and assign approximately equal weights to all stimuli in the stimulus domain. Anderson has offered this view of an equal weighting model in a number of sources (Anderson, 1974, p. 261; Anderson, 1976, pp. 681–682; Anderson & Lopes, p. 69; Leon, Oden, & Anderson, 1973, pp. 301–302; Oden & Anderson, 1971, p. 159). The main objective of the present article is to examine the assumption that all personality trait adjectives, as items of person information, carry approximately equal weight in the impression formation task.

This assumption of equal weighting has been questioned by some investigators (e.g., Birnbaum, 1974). In the following review of research bearing on the equal weighting question, we note that existing data are not conclusive regarding the source or strength of differences in trait weights. One of the problems with this research is that most of the studies are based on group averages. This paper explores the possibility that such group data may "average out" idiosyncratic differences in trait weights that exist separately for each individual.

Differential Weighting

Two kinds of deviations from equal weighting have been examined in previous research; one is related to the lability of a trait's weight, and the other concerns the variation in weight between different traits.

Weight lability. The lability of a trait's weight reflects the extent to which the weight is affected by situational or contextual variables. There can be little question as to whether "differential" weighting occurs in the impression formation domain in this sense of the term.

Characteristics of the judgment task are known to affect the functional weight given any particular item. For example, communicator credibility (Rosenbaum & Levin, 1968) and serial position (Anderson, 1965) can act as situational determinants of trait weight.

Weight can also be influenced by the semantic relationship that a trait holds with other items in the set. The weight parameter is affected by the level of redundancy (Schmidt, 1969) and inconsistency (Anderson & Jacobson, 1965) among traits describing the same person. Also, there are traits that possess two distinct semantic meanings (polysemous homographs)—traits such as *discriminating* and *sensitive*—that may change their scale value as well as their weight, depending on other stimuli in the information set. Although each of the above features can lead to differential weighting, they all derive from the relationship between the semantic definitions of the different traits and can in principle be specified a priori. They do not represent instances in which particular traits carry different weights when in isolation from one another.

Weight differences between traits. The weights of most, if not all, traits are clearly labile. However, this does not mean that traits differ in their natural or context-free weights. When variables known to influence lability are held constant, it is possible that trait weights are either approximately equal or substantially different from one another. It is this sense of the term *differential weighting* that is being rejected when the integration process in impression formation is said to conform to an equal-weight averaging model.

Several possible determinants of trait weight have been examined. Both trait novelty (Wyer, 1970) and trait ambiguity (Kaplan, 1971, 1975; McKillip, Barrett, & Dimiceli, 1978; Schümer, 1973; Wyer, 1974b; Wyer & Watson, 1969) have received some attention.

In neither case, however, do the findings unequivocally support the conclusion of differential weighting in the integration process.

Other research has been directed toward sources of differential weighting that are associated with trait scale value. It is possible, for example, that traits with extreme scale values carry more weight than do neutral traits (e.g., Warr, 1974; Warr & Jackson, 1975). These findings are difficult to interpret, however, because of confounding between negativity, ambiguity, and extremity (Warr & Jackson, 1976; Wyer, 1973).

Unlike the above potential sources of differential weighting, research findings have unequivocally established the existence of a negativity effect. Traits with a negative valence tend to carry more weight than those with positive valence (e.g., Birnbaum, 1974; Hodges, 1974; Oden & Anderson, 1971). This negativity effect is thought to exist for a number of response domains (Kanouse & Hanson, 1972).

Although the negativity findings represent a genuine limitation to the equal weighting assumption, they may not be especially serious. Anderson (1974, p. 261) has speculated, for example, that the magnitude of this negativity effect is not very large and that it may be restricted to only extremely negative stimuli. Another difficulty with this research is that it only employed a limited sample of positive and negative traits. No attempt was made to see how representative the sample was of the entire trait population. The possibility exists that most traits have equal weights, but that a few negative traits carry a higher weight and a few positive traits a lower weight. One objective of the present research was to examine these questions about the negativity effect.

Idiosyncratic Weighting

Most work on the issue of differential weighting has been at the level of *normative* weights. That is, the research has examined how much weight is given traits by the typical or average subject. Little consideration has been given to the problem of whether or not equal weighting applies to each person individually. It is possible that because of a person's idiosyncratic experiences with different trait adjectives, he or she may assign widely differing weights to the adjectives. However, one person's pattern of weights may be unrelated to another's. One person may weight *friendly* twice as much as *intelligent*, whereas another might give *friendly* only half the weight of *intelligent*. If we were to obtain normative weights by averaging over these two persons (as is the case in most differential weighting research), we would find that the two traits received the same weight. Differential weighting of traits at the level of a single subject (what we term *idiosyncratic weighting*) may emerge as equal weighting when measured at the normative level.

Only one published study has examined the possibility of idiosyncratic differences in trait weights. In this study (Anderson, 1962), the stimulus sets consisted of three traits, one from each of the three factors in the design. Each factor consisited of a negative, neutral, and positive trait to form its three levels. This $3 \times 3 \times 3$ design resulted in a total of 27 stimulus sets. Subjects judged all 27 sets once a day for 5 days. The resulting data were analyzed to determine whether significant interaction variance emerged. This design was repeated for 12 subjects, with every 2 subjects receiving a different stimulus replication (i.e., six stimulus replications of the three-factor design, with 2 subjects per replication). The test for differential weighting (or nonadditivity) was performed on the pooled (i.e., 20 degrees of freedom) interaction. Of the 12 subjects, 3 were found to produce significant nonadditivity.

Even though the total number of subjects was small, this study suggests that about 25% of the people show differential weighting. That percentage figure would, no doubt, have been higher had each of the two-way interactions and the three-way interaction been separately tested for each subject. It would be possible for a subject to have a significant two-way interaction but not to have the pooled interaction reach significance.

On the other hand, it is possible that the 25% figure is actually an overestimate of the number of people displaying differential weighting. Two of the three subjects with sig-

energetic on the favorability continuum, it is likely that *energetic* would have the higher scale value on an activity continuum.

Integration theory has been used extensively to analyze the adjective combination task first introduced by Asch (1946) to study impression formation processes. The stimuli used in the impression formation task are personality trait adjectives, and the response continuum refers to overall impression favorability. At the operational level, most investigators use the 555 traits scaled by Anderson (1968) as their domain of stimuli. The judgment scale usually has seven or more intervals and is anchored with the bipolar terms of favorable–unfavorable or like–dislike.

It is widely assumed that for the impression formation response domain, people employ an averaging integration rule and assign approximately equal weights to all stimuli in the stimulus domain. Anderson has offered this view of an equal weighting model in a number of sources (Anderson, 1974, p. 261; Anderson, 1976, pp. 681–682; Anderson & Lopes, p. 69; Leon, Oden, & Anderson, 1973, pp. 301–302; Oden & Anderson, 1971, p. 159). The main objective of the present article is to examine the assumption that all personality trait adjectives, as items of person information, carry approximately equal weight in the impression formation task.

This assumption of equal weighting has been questioned by some investigators (e.g., Birnbaum, 1974). In the following review of research bearing on the equal weighting question, we note that existing data are not conclusive regarding the source or strength of differences in trait weights. One of the problems with this research is that most of the studies are based on group averages. This paper explores the possibility that such group data may "average out" idiosyncratic differences in trait weights that exist separately for each individual.

Differential Weighting

Two kinds of deviations from equal weighting have been examined in previous research; one is related to the lability of a trait's weight, and the other concerns the variation in weight between different traits.

Weight lability. The lability of a trait's weight reflects the extent to which the weight is affected by situational or contextual variables. There can be little question as to whether "differential" weighting occurs in the impression formation domain in this sense of the term.

Characteristics of the judgment task are known to affect the functional weight given any particular item. For example, communicator credibility (Rosenbaum & Levin, 1968) and serial position (Anderson, 1965) can act as situational determinants of trait weight.

Weight can also be influenced by the semantic relationship that a trait holds with other items in the set. The weight parameter is affected by the level of redundancy (Schmidt, 1969) and inconsistency (Anderson & Jacobson, 1965) among traits describing the same person. Also, there are traits that possess two distinct semantic meanings (polysemous homographs)—traits such as *discriminating* and *sensitive*—that may change their scale value as well as their weight, depending on other stimuli in the information set. Although each of the above features can lead to differential weighting, they all derive from the relationship between the semantic definitions of the different traits and can in principle be specified a priori. They do not represent instances in which particular traits carry different weights when in isolation from one another.

Weight differences between traits. The weights of most, if not all, traits are clearly labile. However, this does not mean that traits differ in their natural or context-free weights. When variables known to influence lability are held constant, it is possible that trait weights are either approximately equal or substantially different from one another. It is this sense of the term *differential weighting* that is being rejected when the integration process in impression formation is said to conform to an equal-weight averaging model.

Several possible determinants of trait weight have been examined. Both trait novelty (Wyer, 1970) and trait ambiguity (Kaplan, 1971, 1975; McKillip, Barrett, & Dimiceli, 1978; Schümer, 1973; Wyer, 1974b; Wyer & Watson, 1969) have received some attention.

In neither case, however, do the findings unequivocally support the conclusion of differential weighting in the integration process.

Other research has been directed toward sources of differential weighting that are associated with trait scale value. It is possible, for example, that traits with extreme scale values carry more weight than do neutral traits (e.g., Warr, 1974; Warr & Jackson, 1975). These findings are difficult to interpret, however, because of confounding between negativity, ambiguity, and extremity (Warr & Jackson, 1976; Wyer, 1973).

Unlike the above potential sources of differential weighting, research findings have unequivocally established the existence of a negativity effect. Traits with a negative valence tend to carry more weight than those with positive valence (e.g., Birnbaum, 1974; Hodges, 1974; Oden & Anderson, 1971). This negativity effect is thought to exist for a number of response domains (Kanouse & Hanson, 1972).

Although the negativity findings represent a genuine limitation to the equal weighting assumption, they may not be especially serious. Anderson (1974, p. 261) has speculated, for example, that the magnitude of this negativity effect is not very large and that it may be restricted to only extremely negative stimuli. Another difficulty with this research is that it only employed a limited sample of positive and negative traits. No attempt was made to see how representative the sample was of the entire trait population. The possibility exists that most traits have equal weights, but that a few negative traits carry a higher weight and a few positive traits a lower weight. One objective of the present research was to examine these questions about the negativity effect.

Idiosyncratic Weighting

Most work on the issue of differential weighting has been at the level of *normative* weights. That is, the research has examined how much weight is given traits by the typical or average subject. Little consideration has been given to the problem of whether or not equal weighting applies to each person individually. It is possible that because of a person's idiosyncratic experiences with different trait adjectives, he or she may assign widely differing weights to the adjectives. However, one person's pattern of weights may be unrelated to another's. One person may weight *friendly* twice as much as *intelligent*, whereas another might give *friendly* only half the weight of *intelligent*. If we were to obtain normative weights by averaging over these two persons (as is the case in most differential weighting research), we would find that the two traits received the same weight. Differential weighting of traits at the level of a single subject (what we term *idiosyncratic weighting*) may emerge as equal weighting when measured at the normative level.

Only one published study has examined the possibility of idiosyncratic differences in trait weights. In this study (Anderson, 1962), the stimulus sets consisted of three traits, one from each of the three factors in the design. Each factor consisited of a negative, neutral, and positive trait to form its three levels. This $3 \times 3 \times 3$ design resulted in a total of 27 stimulus sets. Subjects judged all 27 sets once a day for 5 days. The resulting data were analyzed to determine whether significant interaction variance emerged. This design was repeated for 12 subjects, with every 2 subjects receiving a different stimulus replication (i.e., six stimulus replications of the three-factor design, with 2 subjects per replication). The test for differential weighting (or nonadditivity) was performed on the pooled (i.e., 20 degrees of freedom) interaction. Of the 12 subjects, 3 were found to produce significant nonadditivity.

Even though the total number of subjects was small, this study suggests that about 25% of the people show differential weighting. That percentage figure would, no doubt, have been higher had each of the two-way interactions and the three-way interaction been separately tested for each subject. It would be possible for a subject to have a significant two-way interaction but not to have the pooled interaction reach significance.

On the other hand, it is possible that the 25% figure is actually an overestimate of the number of people displaying differential weighting. Two of the three subjects with sig-

nificant interaction variance received the same stimulus replication. It is conceivable that this particular stimulus replication contained inconsistent traits, redundant traits, or polysemous homographs. Anderson (1962) did not report attempting to avoid stimulus combinations of this sort.

There is a second reason that the 25% figure might be an overestimate. There is no way to verify that the response scale that these three subjects used to make their judgments was linearly related to the underlying subjective continuum. It is possible that these three subjects assigned equal weight to all stimuli but that in reporting their subjective impressions they did not use an equal-interval response scale. For example, a statistical interaction of the form implying that negative traits carry more weight than positive traits could be produced by a nonlinear response scale if the categories at the negative end of the scale were wider (i.e., covered a greater range of the subjective continuum) than the categories at the positive end.

The research in the present paper was undertaken to provide a more thorough test of the prevalence of idiosyncratic weighting in the trait judgment domain. Two studies are reported, both of which have the same methodological features. The studies were designed to maximize the likelihood of detecting idiosyncratic weighting. The design used by Anderson (1962) was improved upon in six ways.

The stimulus sets used by Anderson contained three traits. According to the averaging model, however, the observable effects of weight differences due to a single trait decrease as the number of other traits in the set increases. To maximize sensitivity to differential weighting, the present studies used sets composed of only two traits.

Second, there is a need to increase the number of stimulus replications (Anderson, 1962, used only six). The use of an interaction to test the presence of differential weighting is contingent upon having the two traits with a difference in weight present in the same factor of the design. That is, the two must both be on Factor A or on Factor B. If two highly weighted stimuli are on Factor A and two low weighted stimuli are on Factor B, the averaging model predicts no interaction, and so the effect of the weight differences would go undetected. Increasing the number of stimulus replications increases the likelihood that at least some traits with differential weights will be assigned to the same factor in the design. The present research employed 10 different stimulus replications in the first study and 10 new replications in the second.

Third, most research applying integration theory to the trait-judgment domain has used equal weighting instructions in which subjects are explicitly told that each trait is accurate and equally important and that the subject should pay full attention to each of them. To the extent that subjects actively bear those instructions in mind while making their judgments, it is possible that differential weighting is suppressed in this line of research. Some support for this notion is offered by Anderson and Jacobson (1965), who obtained slightly stronger evidence of inconsistency discounting under naturalistic instructions than under equal weighting instructions. The first study in this paper deleted any mention of equal weighting, and the second study manipulated weighting instructions as a factor.

Fourth, it is possible that the design-wide structure of the stimulus combinations could lead subjects to adopt an equal weighting set regardless of the explicit instructions. Anderson (1962) presented subjects with all 27 combinations in the three-factor design. That meant that each trait appeared nine times, once with every combination of all the traits from the other two factors. Each trait appeared in one third of all the test sets presented. It is possible that such extensive repetition induces a mechanistic judgment set in which the subject actively suppresses any differential weighting that he or she might naturally bring to the task. In the present studies, 40 test sets are presented, and no adjective appears in more than 2 of the 40.

Fifth, if differential weighting applies only to some stimuli and not to others, it is important to design the study so as to maximize the likelihood of identifying those stimulus-subject combinations that display differential weighting. Whereas the basic interaction used

in Anderson (1962) was a 3 × 3 (or four degrees of freedom) interaction, the present study used a 2 × 2 (or one degree of freedom) interaction. In the Anderson design, it would be difficult to detect a significant interaction that was due to only one of the six stimuli entering into a 3 × 3 interaction. Although the effect of differential weighting in that design might affect only one of the four components of the interaction, statistically it would be spread out over the four degrees of freedom and, when tested on that basis, might not produce overall significance. In a 2 × 2 design, however, differential weighting contributed by a single trait would be wholly contained in the one-degree-of-freedom interaction term.

Lastly, the Anderson (1962) study employed only 12 subjects. It is difficult to generalize about an estimate of the percent of people displaying differential weighting from such a small sample. In the present paper, 40 subjects were used in the first study and 80 in the second, providing greater confidence in the generality of the findings.

Experiment 1

In the first experiment, each subject was given 40 two-trait stimulus sets to judge on a likeability scale once a day for five days. The stimulus sets contained 10 replications of a 2 × 2 design in which each factor of the design contained a moderately favorable and a moderately unfavorable trait. Each stimulus replication provided one opportunity for differential weighting to emerge, offering 10 opportunities for each subject. Such a design not only allowed an analysis of the prevalence of idiosyncratic weighting but permitted a determination of whether significant interactions emerge from many or only a few stimulus replications. Further, a qualitative analysis of the pattern of means for significant interactions can reveal the proportion of unequal weight due to negative traits carrying more weight than positive traits.

Method

Subjects

Subjects were 40 introductory psychology students, 19 males and 21 females, at Ohio State University, who participated in partial fulfillment of course requirements.

Stimuli

The stimulus trait adjectives were arbitrarily selected from Edward's (Note 1) Ohio State rescaling of Anderson's (1968) personality trait adjectives. Subjects judged 40 stimulus trait pairs, 4 pairs from each of 10 stimulus set replications. The 4 pairs from each replication represented the four cells of a 2 × 2 factorial design. Each factor of this design had two levels of trait likeableness, a moderately favorable (M+) trait and a moderately unfavorable (M−) trait. For each replication there is one M+M+ trait pair, two M+M− trait pairs, and one M−M− trait pair. Mean scale values for the M+ and M− traits were 5.5 and 2.5, rated on a scale of likeability ranging from 1 (dislike very much) to 7 (like very much). In selecting traits for each replication, care was taken to eliminate all instances of redundancy and inconsistency.

Procedure

When subjects arrived for the first day of the experiment they received the following written instructions:

> This experiment is concerned with impression formation. What we are interested in is how people form impression of others on the basis of very limited information. . . .
>
> In this particular experiment, you will be seeing pairs of traits that describe various persons. You will be asked to tell how much you would like each person. Imagine that each of the two traits was contributed by a different person who knows him (her) well.
>
> Read each pair carefully, try to imagine the type of person being described, and rate how much you like the person, using the scale given below each pair of traits. Sometimes this may seem hard, but just act naturally and do the best you can.

All subjects judged the same 40 stimulus sets on each of 5 successive days. Stimulus sets were presented in a booklet, 1 set to a page. There were five random orders of stimuli. The five booklet orders were Latin square counterbalanced across the 5 days, so that each subject received each booklet order and all booklet orders appeared equally often over the 5 days. Ratings were made on a 21-point scale ranging from 10 (like very much) to −10 (dislike very much).

Results

Group Data

Most studies that have tested for equal weighting in the adjective judgment task have

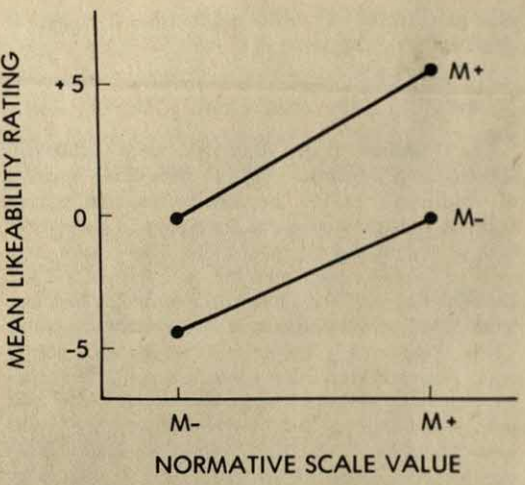


Figure 1. Mean likeability rating of adjective sets composed of M+ and M− (moderately favorable and moderately unfavorable) traits, averaged over days, stimulus replications, and subjects in Experiment 1.

averaged over both stimulus replications and subjects. In analyzing the present group data, the question of equal weighting can be examined both on this overall basis and within each stimulus replication.

If traits are equally weighted, there should be no interaction between the M+ and M− levels of the A factor and the M+ and M− levels of the B factor. The present group data brings considerable power to the test of that interaction; each cell mean is the average of 2,000 ($10 \times 5 \times 40$) observations. The interaction (see Figure 1) was significant, $F(1, 39) = 10.35$, $p < .005$, and its pattern is similar to that obtained by previous investigators. It can be explained by assuming that negative traits carry more weight than positive traits.

Whether the obtained overall interaction represents a serious challenge to the equal weighting assumption is contingent upon several factors, including its magnitude, the proportion of stimulus replications that contribute to it, the presence of similar versus differential interaction patterns across stimulus replications, and whether response scale nonlinearity could account for the significant effects.

The magnitude of the overall interaction was not very substantial. The interaction contributed only one percent to the total between-cells variance. With fewer observations going into each cell mean, this interaction could well have gone undetected.

It is conceivable that the overall interaction was due to only a few of the 10 stimulus replications. In line with Anderson's (1974) suggestions, it could be the result of the one or two stimulus replications that contained the most extremely negative traits from the M− range. There is a second advantage to examining each of the stimulus replications separately. It is possible that the pattern displayed in Figure 1 does not hold for all significant stimulus replications. If the majority of the significant interaction replications showed the pattern in Figure 1 and were averaged with others that showed different patterns (e.g., convergence rather than divergence to the right of the graph), the observed overall interaction could still have emerged.

Separate analyses were performed on each of the 10 stimulus replications, and only 2 were found to contain significant interaction variance, $Fs(1, 1521) > 6.11$, $ps < .05$.[1] The percent of between-cells variance contributed by the interaction was 3.40 and 3.80 for the first and eighth replications, respectively. The solid line portions of Figure 2 display the interactions for those 2 stimulus replications. Both replications showed the same pattern—one that could be explained by assuming that negative traits carry more weight than positive traits do.

A significant interaction could be produced by weight differences between the two traits on the A factor, on the B factor, or on both. Each stimulus replication, then, provides two opportunities for differential weighting to occur. The fact that only two of the replications had significant interactions means that near equal weighting was observed in between 16 and 18 of the 20 opportunities. At this normative level of analysis, it appears that equal weighting was characteristic of 80% to 90% of the trait pairs. The overall negativity effect was produced by at most 20% of the stimulus pairs.

[1] The error term for these analyses was the interaction between subjects (40) and stimulus persons (40).

The data provide some support for the interpretation that differential weighting is restricted to highly negative traits. The lowest marginal mean was computed for each stimulus replication to provide an index of which replications contained the most negative trait. Among the 10 stimulus replications, the 2 replications for which the interaction was significant were the first and fifth lowest.

Response scale linearity. Significant interactions of the form portrayed in Figures 1 and 2 could be obtained even if all traits were subjectively given equal weights. This would occur if subjects used wider categories at the negative end of their response scale than at the positive end. In such a case, subjects' overt responses would not be linearly related to their subjective responses. Two qualitative tests were devised for the purpose of examining the response scale explanation.

Response scale nonlinearity can be dismissed if the qualitative pattern satisfies a *test of disordinality*. The interaction portrayed in Figure 1 is termed an ordinal interaction because both the lower and upper lines are ordinally related to the horizontal axis in the same direction (i.e., the sign of both slopes is the same). A disordinal interaction would be one in which the slope of one line is positive and the slope of the other is negative. If the means of a 2 × 2 design can be plotted so that disordinality emerges, there exists no continuous monotonic transformation of the response scale that will reduce that interaction to zero. Consequently, the presence of disordinality supports an unequal weighting interpretation. Although none of the stimulus replications in the analysis of group data (see Figure 2) displayed such a pattern, it is useful to point out the implications of such a pattern for response scale linearity prior to discussing the individual data.

The *intersection test* involves comparing the data pattern of a significant interaction for one stimulus replication against a base line provided by a nonsignificant (i.e., interaction $F \leq 1$) stimulus replication. If the judgments contributing to both interactions came from comparable portions of the response scale, no continuous monotonic quadratic transformation of the response scale (e.g., using wider

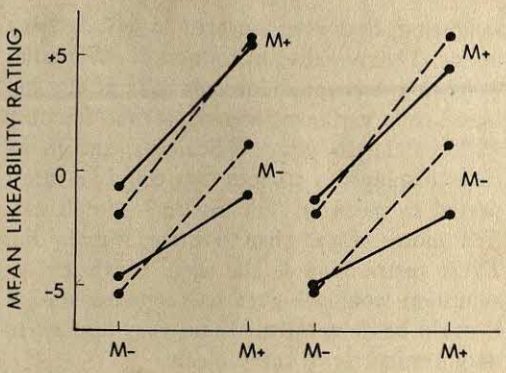


Figure 2. Use of the intersection test to eliminate scale nonlinearity as an explanation of the interactions obtained when averaging over days and subjects in Experiment 1. (M+ = moderately favorable. M− = moderately unfavorable. The significant interactions for Stimulus Replications 1, $F(1, 156) = 6.35$, and 8, $F(1, 156) = 6.12$, are in solid lines and the nonsignificant control interaction from Stimulus Replication 3, $F(1, 156) = 1.00$, is in dashed lines.)

categories at the negative than at the positive end) could simultaneously reduce both interactions to zero. The intersection test provides a way of insuring that both the significant interaction and the nonsignificant control interaction involved responses from similar portions of the rating scale. The control stimulus replication must be selected so that when the two interactions are graphed together, the lower lines of each intersect one another and the upper lines of each intersect one another.

This intersection test was applied to both significant stimulus replications for the group data. It was found that Stimulus Replication 3 ($F = 1.00$), in which the interaction accounted for only 0.43% of the between-cells variance, provided such a control for both significant replications (see Figure 2). It can be concluded, then, that the two significant interactions portrayed in Figure 2 represent genuine instances of unequal weighting and should not be dismissed on the grounds of simple response scale nonlinearity.

Individual Data

The analyses of group data verified the presence of differential weighting in the trait-judgment paradigm. The deviation from equal

weighting, however, appeared to be relatively minor. The overall contribution of differential weighting accounted for only 1% of the between-cells variance, was significant for only 20% of the stimulus replications, and in all cases displayed a pattern that could be interpreted as meaning that negative stimuli carried more weight than positive stimuli did. These restrictions on the equal weighting assumption would be even less consequential if it could be demonstrated that they applied to only a minority of the subjects.

Setting alpha at .05 and testing each 2 × 2 stimulus replication interaction against the subject's own Days (5) × Stimulus Persons (40) interaction (on 156 degrees of freedom), a full 82.5% of the subjects showed evidence of nonadditivity for at least 1 of the 10 stimulus replications (see Table 1).

One difficulty with the use of 10 stimulus replications lay in the increasing role of chance as the number of replications increased. Assuming that 5 subjects out of 100 would produce a significant interaction for any given stimulus replication simply by chance, that percentage increases to 40.1 when each subject responds to 10 different stimulus replications. The role of chance returns to 5% when the interaction variance pooled across all replications (i.e., on ten degrees of freedom) for each subject is tested. This approach is statistically comparable to the pooled test for nonadditivity employed by Anderson (1962). Whereas only 25% of Anderson's subjects showed nonadditivity, Table 1 shows that 60% of the subjects in the present study produced significant overall nonadditivity. These results indicate that when the experiment is designed to be maximally sensitive to the effects of differential weighting, a substantial number of people produce statistically significant interactions in the trait-judgment task.

As noted in the presentation of the group data, a number of questions arise in interpreting significant interactions. First, the possibility must be examined that the interactions resulted from response scale nonlinearity rather than from unequal weighting. Second, it is possible that unequal weighting occurred for only a few stimulus replications (e.g., the two that were significant in the group analysis). Third, it may be that all significant interactions are the result of a single stimulus characteristic (e.g., extremely negative traits having a greater weight than all others), in which case all significant interactions for all persons would have the same pattern (as in Figures 1 and 2). The prediction of idiosyncratic weighting, in regard to the second and third above questions, is that all stimulus replications should be involved and that a variety of interaction patterns should be obtained for each replication.

Scale linearity. Although we discarded scale nonlinearity as an explanation for the two significant stimulus replications for the group data, such an explanation could still be viable for a majority of the individuals. Consequently, the pattern of significant and nonsignificant stimulus replication interactions was examined for the 24 subjects (60% of our sample) who displayed overall nonadditivity.

Of these 24 subjects, 14 satisfied the disordinality test by displaying at least one significant disordinal interaction. Of the 10 remaining subjects, 8 satisfied the intersection test for at least one significant interaction. For the remaining 2 subjects (Subjects 28 and 33), it was not possible to dismiss the nonlinearity explanation, since in both cases they showed the same interaction pattern for all 10 stimulus replications. That pattern was consistent with the interpretation that negative stimuli carry more weight than positive stimuli do. However, even after deletion of these 2 subjects, 55% of the sample showed genuine evidence of differential weighting.

Stimulus replications. The possibility remains that such widespread differential weighting is due to a limited number of stimulus replications. Two of the three subjects with significant interaction variance in the Anderson (1962) study were presented with the same stimulus replication.

Out of 400 interactions tested in the present study, 106 were significant. Table 2 shows the distribution of significant interactions over the 10 stimulus replications. It can be seen that the significant interactions were not restricted to just a few stimulus replications. One hundred percent of the replications were involved,

Table 1
Analyses of Individual Data for Experiment 1

Subject number	Pooled interaction Fs[a]	Stimulus replication interactions[b]									
		1	2	3	4	5	6	7	8	9	10
1	1.04										
2	1.51							*			
3	3.36							*			
4	2.06*		*	*	*						
5	1.19			*							
6	4.05**				**					**	
7	3.33**							**		**	**
8	3.15**								**		
9	.95										
10	1.93*							**			
11	1.09	*								*	
12	.05										
13	7.03**	**			**	**		**	**		**
14	3.97**	**	*	*		*					
15	3.98**	*				**	*	**			*
16	2.86**	**			*				**		
17	3.47*				**	*			**		
18	1.92*		*						*		
19	1.22		*								
20	3.33**		**						**		
21	3.47**		*	**	**						
22	2.63**	*						*	**		
23	.67										
24	5.93**	**		**					**		*
25	1.19			*							
26	1.37										*
27	18.39**	*	**	**	**		**	**	*	**	**
28	1.73						**		*		
29	2.27*			**							
30	6.09**	**		**	*	**	**				
31	18.87**	**		**	**			**	**	**	**
32	1.09						*				
33	25.21**	**	**	**	**	**	**	**	**	**	**
34	.53										
35	1.25										
36	5.50**	**			*			**	**		*
37	2.14*			*		*	*		*		**
38	.67										
39	6.13**				*		**		**		**
40	4.06**							**		*	**

[a] $df = 10, 156$.
[b] $df = 1, 156$.
* $p < .05$. ** $p < .01$.

and all contributed approximately equally. For the distribution of interactions over the 10 stimulus replications, $\chi^2(9) = 5.89$, $p > .75$.

An average of 26.5% of the subjects had a significant interaction for the typical stimulus replication. The number of subjects (out of 40) with significant interactions ranged from a low of 7 to a high of 15 over stimulus replications. Even for the most additive of the stimulus replications, 17.5% of the subjects had a significant interaction. In contrast to the findings of the group data, then, it would appear that the pattern of unequal weighting observed in the individual analyses cannot be dismissed as due to only a few stimulus combinations.

Idiosyncratic weighting. The prediction of

Table 2
Number of Subjects in Each Stimulus Replication With a Significant Interaction for Experiment 1

Stimulus replication	Number of subjects with significant interaction	Number interpretable as $w_p > w_n$	Number interpretable as $w_n > w_p$
1	12	1	11
2	8	3	5
3	12	5	7
4	12	6	5
5	8	1	7
6	8	1	7
7	12	1	10
8	15	0	15
9	7	1	6
10	12	2	9
M	10.6	2.1	8.2

Note. $N = 40$. The terms w_p and w_n refer to the weights given the more positive and the more negative traits, respectively.

idiosyncratic weighting derives from the reasoning that trait meaning and importance is acquired on the basis of each individual's personal linguistic experiences with the word and its referents. It follows from this that for most stimulus replications, some people should weight the negative traits more highly and others should weight the positive traits more highly. Also, when examining the significant stimulus replications for each subject, there should be some subjects who show both patterns of trait weighting. Since the group data showed evidence only of negative traits carrying more weight than positive, we must first ask whether there is any evidence for the opposite pattern in the individual data.

Both ordinal and disordinal interactions can be coded in terms of whether the more positive or the more negative trait is being given the most weight. Of the 103 codable [2] interactions in the present study, 82 (79.6%) were found to have a pattern consistent with the interpretation that negative traits carry more weight than positive traits do. The remaining 20.4% showed exactly the opposite pattern, that is, positive traits carried more weight than negative traits did.

By chance alone, it would be expected that half of the 400 interactions would show a negativity effect and half a positivity effect. At the .05 level of significance, this means that 10 tests should show significant negativity and 10 tests significant positivity. The observed frequencies of 82 and 21 were both significantly greater than the expected frequencies, χ^2s(1) = 531.69 and 12.41, ps < .001, respectively. Unlike previous research at the normative level, clear evidence for a positivity effect is found for some subjects. Table 2 shows that in all but one stimulus replication, there existed at least one subject who weighted the positive traits more than the negative traits. An illustration of this diversity of interaction patterns is provided in Figure 3 for Stimulus Replication 7.

Although there might be individual differences in the propensity to assign greater weight to negative or to positive traits, the notion of idiosyncratic weighting leads us to expect that a substantial number of people will display both weighting patterns over different stimulus replications. Of 23 subjects who provided more than one significant stimulus replication interaction, 9 (or 39%) showed both weighting patterns. This is illustrated for Subject 36 in Figure 4. The remaining subjects split 12 and 2 in showing exclusively higher negative weighting and higher positive weighting, respectively.

Discussion

In contrast to the group data, the individual data indicated that differential weighting in the trait-judgment paradigm was widespread. No stimulus replication was immune and no single pattern of differential weighting (e.g., negative traits receiving more weight) prevailed exclusively. It appears, then, that when a more sensitive experimental design than that used by Anderson (1962) is employed, clear differences emerge between the analyses of group judgments and the analyses of indi-

[2] Three disordinal interactions could not be coded in these terms because the two lines intersected. When such "double disordinality" is present, it is impossible to determine the relative scale values of the stimulus traits for the stimulus replication.

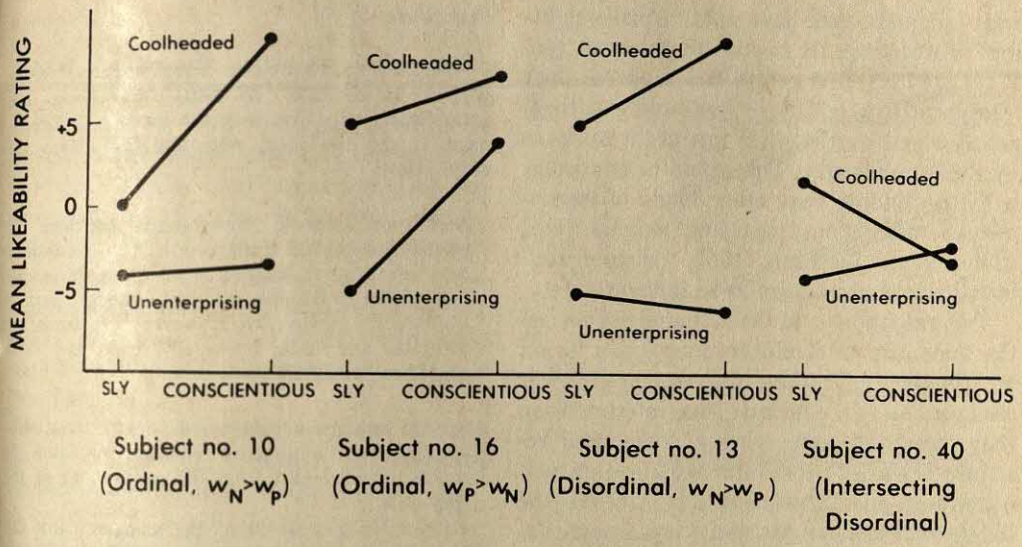


Figure 3. Examples of four different significant interaction patterns obtained for Stimulus Replication 7 in Experiment 1. (The terms w_p and w_n refer to the weights given the more positive and the more negative traits, respectively.)

vidual judgments. Whereas the equal weighting assumption holds for most stimuli at the normative level, idiosyncratic weighting is the dominant feature at the individual level.

Experiment 2

Experiment 2 was undertaken for three reasons. Since the outcome of Experiment 1 stood in direct contradiction to previous assumptions regarding equal weighting of items in the trait-judgment task, a need to replicate the findings was evident. Experiment 2 used the same individual-based design as Experiment 1 but employed 10 new stimulus replications and 80 new respondents.

Whereas Experiment 1 used "naturalistic" instructions, the majority of studies with the

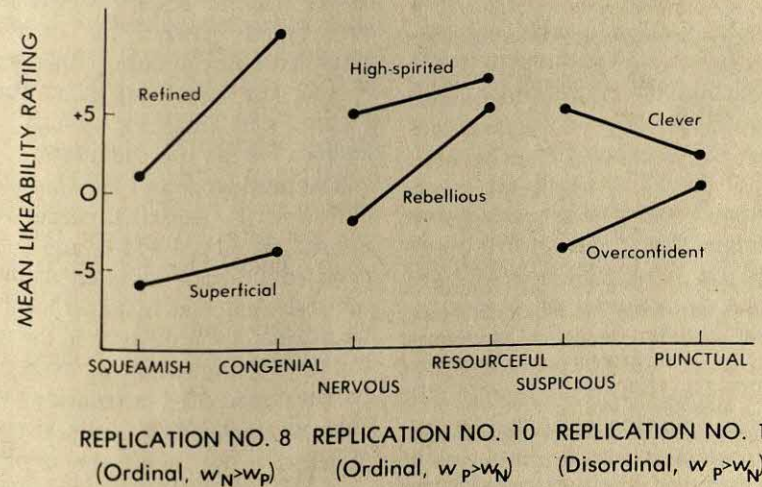


Figure 4. Examples of three different interaction patterns obtained from Subject 36 in Experiment 1. (The terms w_p and w_n refer to the weights given the more positive and the more negative traits, respectively.)

trait-judgment task have used "equal weighting" instructions. It is possible to argue (see Wyer, 1974a) that people can suppress their idiosyncartic weighting of traits and give them nearly equal weight when instructed to do so by the experimenter. This would be consistent with the finding that other kinds of experimenter instructions regarding weights (e.g., Anderson & Jacobson, 1965, "discounting" instructions) are known to be influential.

Two respondents in the first study produced the same form of significant interaction for all 10 stimulus replications. Consequently, it was not possible to eliminate the interpretation that these two were using a noninterval response scale. A procedural modification was introduced in Experiment 2 to reduce the likelihood that subjects would use a noninterval response scale. A total of 20 extreme anchor stimuli preceded and were interspersed among the 40 test stimuli in this study. This leads subjects to use the interior portions of the scale when rating the stimuli (Simpson, Ostrom, & Sloan, 1973) and thereby reduces the influence of any noninterval response tendencies that may intrude into either the positive or negative extremes of the scale. The use of such anchor stimuli has become widely adopted in the trait-judgment paradigm.

Method

Subjects

Subjects were 80 introductory psychology students from Ohio State University, 40 males and 40 females, who participated in partial fulfillment of course requirements.

Stimuli

New stimulus traits were randomly selected from the Edwards (Note 1) list. Subjects judged 40 test stimulus trait pairs from 10 randomly constructed stimulus set replications, where each set consisted of 4 trait pairs produced from a 2×2 factorial design. The two levels of likeability for the two trait factors were M+ and M−, as in Experiment 1. No semantically inconsistent or redundant trait pairs were allowed within a stimulus replication.

In order to adequately anchor the response scale, 12 pairs of extreme anchor stimuli preceded the 40 pairs of test stimuli, and 8 pairs of anchor stimuli were evenly interspersed among the 40 test sets. Half of the anchor pairs were positive, and half were negative.

Procedure

Subjects were given either "equal weighting" ($N = 40$) or "naturalistic" ($N = 40$) instructions. The naturalistic instructions were the same as in Experiment 1, and the equal weighting instructions are given below.

> Each trait is equally important in describing the person. Sometimes, of course, the two words may seem inconsistent. That is to be expected, because each of the two people may see a different part of the person's personality. However, both traits are accurate, and each is equally important. You should pay equal attention to both.

After the experimenter answered any questions, subjects completed a practice booklet containing 15 stimulus pairs sampled from the entire range of the rating scale.

Subjects judged all 60 of the stimulus pairs (20 anchor and 40 test) in the experimental booklets on each of 5 successive days. Ratings were made on a 21-point scale ranging from 0 (dislike very much) to 20 (like very much). Five random orders of test stimuli were Latin square counterbalanced across days. Anchor stimulus pairs always appeared in the same order.

Results

The second experiment differed from the first in two important ways: (a) It employed both equal-weight and naturalistic instructions and (b) it introduced extremely positive and negative anchor stimuli in order to stabilize the response scale. No reliable differences were found between the equal-weight and naturalistic instructions, either for the group data or the individual data. The results in Figure 5 and Table 3 show highly comparable findings for the two conditions.

The introduction of extreme anchor stimuli did have its expected effect. Whereas the means for the M+M+ and M−M− cells deviated 5.0 scale units on the average from the scale midpoint in Experiment 1, they deviated only 3.9 scale units in the second study. Further, the adoption of anchor stimuli apparently succeeded in reducing the number of subjects employing a nonlinear response scale. Whereas in the first study 2 of 40 subjects were identified as possibly using nonlinear scales, none of the 80 subjects in the second study were so identified.

Group Data

As in the first study, the overall interaction was significant, $F(1, 78) = 27.97$, $p < .001$, and as Figure 5 shows, was not significantly reduced under equal weighting instructions, $F(1, 78) < 1$. The total proportion of the between-cells variance contributed by the interaction was .31%. Separate analyses were done for each instruction condition, and the interaction was significant for both, Fs$(1, 39) > 13.87$, ps $< .001$, with neither contributing more than .40% of the between-cells variance (see Table 3).

When tested separately, 4 of the 10 stimulus replications (1, 4, 6, and 10) produced significant interactions, Fs$(1, 3042) > 3.97$, ps $< .05$.[3] Of the 20 stimulus pairs in this study, between 60% and 80% showed no significant departure from equal weighting at this normative level. This is slightly lower than was observed in the first study (for which the comparable estimate was between 80% and 90%). The observed increase in number of significant interactions was due to the doubling of sample size. When each instruction condition was tested separately, only two replications were significant in each (4 and 6).

Three of the four replications satisfied the intersection test and could not therefore be dismissed on the grounds of scale nonlinearity. All significant replications had the same pattern as Figure 5, indicating that when averaged over all subjects, negative stimuli appeared to carry more weight than positive stimuli. There was, however, only modest evidence that the significant stimulus replications contained the most negative traits. Using the lowest marginal mean as an index of each replication's most negative trait, the four significant replications ranked second, third, fourth, and seventh most negative.

Overall, the results for these group data substantially replicate the group results for Experiment 1.

Individual Data

Over three quarters (77.5%) of the subjects had at least one significant interaction. Although this figure was quite comparable to

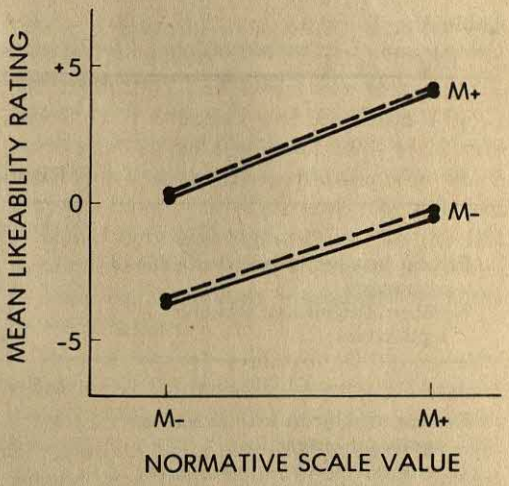


Figure 5. Mean likeability ratings averaged over days, stimulus replications, and subjects in Experiment 2. (M+ = moderately favorable. M− = moderately unfavorable. Naturalistic instructions results are in dashed lines, and equal weighting instructions results are in solid lines.)

that for Experiment 1, there was a reduction in the percentage of subjects who showed significant pooled interaction variance (on 10 and 156 degrees of freedom), from 60% to 46%. Even though this figure is slightly below 50%, it is still substantially higher than a baseline provided either by chance, $\chi^2(1) = 286.30$, $p < .001$, or by the 25% level observed in Anderson's (1962) study, $\chi^2(1) = 19.27$, $p < .001$.

The discrepancy between the two studies is reduced somewhat when those subjects whose interactions might have been the result of nonlinear response scales are eliminated from consideration. Whereas 2 subjects were deleted on these grounds in the first study (reducing the figure to 55%), none were eliminated from the second study (leaving the figure at 46%). In examining the scale nonlinearity interpretation for the 37 nonadditive subjects in the second study, 22 satisfied the disordinality test, and 15 the intersection test.

Stimulus replications. Once again, all stimulus replications contributed approximately equally to the significant interactions obtained

[3] The error term for these analyses was the interaction between stimulus persons (40) and subjects (40) nested within instruction conditions (2).

Table 3
Comparison of Experiments 1 and 2 for Group and Individual Data

Item	Experiment 1	Experiment 2		
		Naturalistic instructions	Equal-weight instructions	Combined
Group data				
Percent between cells variance due to interaction	1.00	.26	.40	.31
Number of significant stimulus replications	2	2	2	4
Individual data				
Percent of subjects with at least one significant interaction[a]	82.5	72.5	82.5	77.5
Percent of subjects with significant pooled interactions[b]	60.0	45.0	47.5	46.25
Percent of stimulus replications with at least one significant interaction	100	100	100	100
Total number of significant interactions	106	57	76	133
Percent of all significant interactions in which:				
$w_n > w_p$	79.6	74.1	78.4	76.6
$w_p > w_n$	20.4	25.9	21.6	23.4

Note. The total number of subjects in Experiment 1 was 40. Experiment 2 had 40 subjects in each of the two instruction conditions. The terms w_p and w_n refer to the weights given the more positive and the more negative traits, respectively.
[a] Chance = 40.1%. [b] Chance = 5%.

at the individual level, $\chi^2(9) = 13.69$, $p > .10$. This occurred, however, despite a substantial decrease in the average percentage of subjects showing a significant interaction for the typical stimulus replication. The figure dropped from 26.5% to 16.6%, with a range, in the second study, from 7.5% to 27.5%.

The decrease in the second study may have resulted from the introduction of 20 extreme anchor stimuli. The anchor stimuli were intended to eliminate some interactions that were due to scale nonlinearity. Since they increased the number of stimulus persons to be judged each day by 50%, however, they may also have induced a more additive integration set by making the task more boring.

Idiosyncratic weighting. Of the 128 codable[4] interactions, 76.6% showed a pattern interpretable as meaning that negative traits carried more weight than positive traits, and 23.4% showed a pattern favoring positive traits. This was very close to the split observed in Experiment 1 (see Table 3). Also, as in Experiment 1, both the number of significant negativity and positivity interactions exceeded chance, $\chi^2(1) = 1624$, $p < .001$, and $\chi^2(1) = 5.13$, $p < .05$, respectively. Instances of both weighting patterns appeared in 8 of the 10 stimulus replications, the remaining 2 containing only interactions in which the negative trait received more weight.

As in the first study, a substantial number of subjects displayed both weighting patterns. Out of the 37 subjects with more than one significant stimulus replication interaction, 35.1% showed both patterns. All but 1 of the remaining 24 subjects gave the negative traits more weight than the positive traits.

Discussion

The results of Experiment 2 replicated the idiosyncratic weighting findings of Experiment

[4] Five disordinal interactions were not codable, four because the lines intersected (see Footnote 2) and one because the marginal means were exactly identical on one of the factors in the 2 × 2 design.

1 and demonstrated as well that idiosyncratic weighting is obtained under both naturalistic and equal weighting instructions. The absence of differences between instruction conditions at the level of group analyses is consistent with several previous studies finding no differences in trait judgments between naturalistic and equal weighting instructional conditions (Anderson & Jacobson, 1965; Gollob & Lugg, 1973; Lampel & Anderson, 1968; Wyer, 1974b). Whereas people appear able to increase differential weighting tendencies when given "discounting" instructions (Anderson & Jacobson, 1965; Kaplan, 1973), they did not reduce their differential weighting tendencies when specifically instructed to weight all traits equally.

General Discussion

The data reported in this article support the conclusion that it is incorrect to describe the integration process in the trait-judgment paradigm as following an equal weighting averaging rule at the level of the individual subject. This conclusion is in contradiction to previous descriptions of the trait-judgment task, descriptions that were based upon the one previous direct test of individual differences in trait weights (Anderson, 1962). The difference in outcomes between this and the Anderson (1962) study can be attributed to the far greater sensitivity of the present experimental design.

Implications for Information Integration Theory

Generality of idiosyncratic weighting. The trait-judgment paradigm is not the only social judgment domain in which differential weighting occurs at the individual level. Significant nonadditivity has been obtained in several other areas. Leon, Oden, and Anderson (1973) found it for 38% of 16 subjects who judged the "badness" of a group of people guilty of committing various crimes; Anderson (1972) obtained it for 83% of 6 subjects who judged, on the basis of behavior items, the severity of psychiatric disturbances; Troutman and Shanteau (1976) observed it for an average of 28% of 20 subjects who judged the quality of disposable diapers in one replication and infant car seats in the others; and Shanteau and Anderson (1969) obtained it for an average of 25% of 20 subjects over four stimulus replications in which subjects rated their preferences for different food and beverage combinations. In the last mentioned study, 65% of the subjects showed at least one significant interaction over the four replications.

This extensive evidence of differential weighting at the individual level was obtained despite the fact that none of the above studies incorporated all of the precautions for maximizing sensitivity employed in the present research. Yet none of these percentage figures are reasonably close to the five percent chance level (although due to small sample sizes, some may not statistically differ from it). It would appear, then, that at present there exists *no* social judgment domain that could plausibly be regarded as an instance where an equal weight averaging integration rule held.

The absence of any verified instance of equal weighting encourages speculation that there may, in fact, exist no social judgment domain for which equal weighting holds. Such a state of nature would follow from the premise that items of social information acquire meaning and importance on the basis of highly individualistic experiences with the signs and symbols conveying that information. Certainly on these grounds, it would seem that any social information conveyed lexically would be susceptible to idiosyncratic weighting differences. A more promising domain where equal weighting may prevail would involve purely sensory stimuli responded to through a nonsemantic mode (as is often done in cross-modality matching).

Normative versus individual levels of analysis. The present studies show that findings obtained at the group (or normative) level do not necessarily replicate for each member of that group. Consequently, it is appropriate to question whether other empirical findings obtained at the group level in the trait-judgment task (e.g., set size, serial position, and inconsistency discounting effects) hold for all indi-

viduals. There are no adequate data presently available that allow us to estimate what percentage of people employ an averaging rule (as opposed to a summative or multiplicative rule) in the trait-judgment task. This possibility of individual differences in integration rule can be illustrated by a study in another social judgment domain (Leon, Oden, & Anderson, 1973). They conducted analyses on individual subject data and found significant set size effects for some subjects and not for others.

The difference in outcome between group and individual responses should not be dismissed as reflecting mere individual differences around a group average. Instead, the two types of data should be regarded as fundamentally different levels of analysis, each appropriate to different scientific objectives.

Data from an individual, when obtained over repeated observations, are definitive for understanding the integration rule used by that person. If the objective is to establish that the typical person uses a particular integration rule, it is necessary to show that most people, when studied as individuals, employ that integration rule. Such a generalization cannot be made on the basis of group data.

This is not meant to dismiss the importance of the multitude of findings previously obtained at the group level. Group data inform us as to the integration rule that characterizes the modal societal reactions made in a particular response domain. At this level of analysis, information integration theory provides estimates of the normative weights and scale values of information items in that domain. It also establishes a normative integration rule for the subject group. This is clearly a descriptive enterprise that need have no implications for how individuals subjectively integrate information. There are many problem areas for which it is of direct interest to study the integration rule underlying the modal group response, areas such as stereotyping, advertising, political preferences, and jury decision making. For example, it is useful to know that jurors in aggregate can presume innocence (Ostrom, Werner, & Saks, 1978), even though this presumption may not hold for all of the jurors considered individually.

Ambiguity of the "parallelism" test. Up to this point in the paper we have chosen to explain deviations from parallelism in terms of unequal trait weights. But to do so requires us to assume that the theoretical scale value of each trait is unaffected by the nature of the other trait in the pair. Unlike trait weight, the scale value is viewed as nonlabile. This assumption has been made in most integration theory interpretations of differential weighting research and has been defended as a legitimate strategy for model building (Himmelfarb, 1975).

Some recent studies (see Ostrom, 1977) have attacked that assumption, and have established the plausibility of scale value lability. The cognitive representation and accompanying evaluative response evoked by a trait appears to be directly affected by the evaluative tone of the context traits. If such meaning shift processes do occur in the trait-judgment domain, the interpretation of deviations from parallelism becomes much more difficult. Significant interactions could be due to differential weighting, scale value shifts (with equal weighting), or to both.

Most of the significant interactions observed in the present studies had an ordinal or disordinal, nonintersecting pattern. These could all be interpreted either in terms of differential weighting or in terms of scale value shift. However, there is one logically possible interaction pattern that, if obtained, could not be explained by a differentially weighted averaging model, namely, a pattern that is both disordinal and intersecting (see the right panel of Figure 3). Given that people are using an averaging integration rule, this pattern could only be obtained if the scale values of the row traits reversed their relative positivity, depending on which column trait they were paired with. Although such patterns were observed in the present studies (three in the first and four in the second), they represented only 2.93% of all significant interactions.

The major findings of the present studies could be explained as easily in terms of scale value shift as they could by differential weighting. Namely, it could be concluded that a majority of the subjects displayed scale value shifts, that all stimuli were capable of undergoing such shifts, and that these shifts

are highly idiosyncratic. For any particular trait pair, some people show greater shift in a negative context, others in a positive context, and other people show no shift at all.

Limitations of a differential weighting averaging model. No fundamental difficulty is created for information integration theory if idiosyncratic weighting proves to be the state of nature in social judgment (or even all judgment) domains. Its objective is, after all, the specification of an integration rule, and determination of whether equal or differential weighting occurs. The problems created are more ones of experimental convenience.

The equal weight averaging model has a remarkably useful feature. When it holds, and when subjects are responding on an interval judgment scale, no interaction should be statistically detectable when orthogonally composed stimulus sets are being judged. This has allowed researchers to draw three important conclusions when such parallelism emerges: (a) An equal weight additive model is appropriate (discarding both unequal weight averaging models and multiplicative models), (b) the interval property of the response scale is validated, and (c) the marginal means offer estimates of stimulus scale values on an interval scale. The empirical procedures needed to discriminate between an averaging model and a multiplicative model, to validate a response scale, and to obtain interval estimates of scale values are much more complicated in the absence of equal weighting.

Another disadvantage is that the trait judgment task cannot be used as a validational baseline for investigating other stimulus domains. For example, Lampel and Anderson (1968) had subjects judge persons described by a photograph and several traits. The traits were factorially constructed, and since no interaction was obtained, the authors concluded that the response scale was validated and that all traits received equal weight. These two conclusions allowed an unambiguous interpretation of an interaction that was obtained between trait valence and photograph attractiveness. The possibility of a nonlinear response scale and the possibility of differential trait weights could both be ruled out because of the absence of a Trait × Trait interaction. This allowed the investigators to conclude (by a process of elimination) that there was a differential weighting of the photographs in which weight was inversely related to photograph attractiveness. The present finding of idiosyncratic weighting would rule out the use of the trait-judgment domain for such purposes in future research at the individual level.

Implications for Person Perception Research

The stimulus domain studied in this research was fairly narrowly defined, being restricted solely to person characteristics in the form of personality trait adjectives. Even that domain, however, was not fully represented in the traits sampled for use in the present studies (from Anderson, 1968). The traits did not include slang terms or adjectives solely descriptive of mood or feeling states (e.g., Bush, 1973). It seems unlikely, however, that either slang or feeling terms would be more characterized by equal weighting than were the sampled adjectives. In fact, slang terms may well be even more prone to idiosyncratic weighting, given respondents who are selected from two or more different social groupings.

There is a wide variety of information items that have been used in person perception research, including personal attitudes, hobbies and interests, demographic characteristics, group memberships, and behavioral acts and intentions. It is, of course, possible (although we do not regard it as probable) that one or more of these categories may represent an equal weighting domain.

Although the present studies were not tailored to investigate the determinants of trait weight, the idiosyncratic weighting findings suggest two possible avenues of exploration. There was some suggestion of individual differences in whether or not weight was associated with scale value. Nearly two-thirds of the subjects who had more than one significant stimulus replication showed the same weighting pattern (either negativity or positivity) for each significant replication. Conceivably, such a dispositional tendency could be related to other individual differences in positivity or negativity in impression judg-

ments (see Kaplan, 1973). A second approach would be to relate trait weight to the personal constructs (Kelly, 1955; Rosenberg, 1977) each individual characteristically uses to describe important others in his or her social world.

Reference Note

1. Edwards, J. D. *Revised likeableness ratings of 554 personality trait adjectives.* Unpublished manuscript, Ohio State University, 1967.

References

Anderson, N. H. Application of an additive model to impression formation. *Science,* 1962, *138,* 817–818.

Anderson, N. H. Primacy effects in personality impression formation using a generalized order effect paradigm. *Journal of Personality and Social Psychology,* 1965, *2,* 1–9.

Anderson, N. H. Likeableness ratings of 555 personality trait words. *Journal of Personality and Social Psychology,* 1968, *9,* 272–279.

Anderson, N. H. Looking for configurality in clinical judgment. *Psychological Bulletin,* 1972, *78,* 93–102.

Anderson, N. H. Information integration theory: A brief survey. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology* (Vol. 2). San Francisco: Freeman, 1974.

Anderson, N. H. How functional measurement can yield validated interval scales of mental quantities. *Journal of Applied Psychology,* 1976, *61,* 677–692.

Anderson, N. H., & Jacobson, A. Effect of stimulus inconsistency and discounting instructions in personality impression formation. *Journal of Personality and Social Psychology,* 1965, *2,* 531–539.

Anderson, N. H., & Lopes, L. L. Some psycholinguistic aspects of person perception. *Memory and Cognition,* 1974, *2,* 67–74.

Asch, S. E. Forming impressions of personality. *Journal of Abnormal and Social Psychology,* 1946, *41,* 258–290.

Birnbaum, M. H. The nonadditivity of personality impressions. *Journal of Experimental Psychology,* 1974, *102,* 543–561.

Bush, L. E., II. Individual differences multidimensional scaling of adjectives denoting feelings. *Journal of Personality and Social Psychology,* 1973, *25,* 50–57.

Gollob, H. F., & Lugg, A. M. Effect of instructions and stimulus presentation on the occurrence of averaging responses in impression formation. *Journal of Experimental Psychology,* 1973, *98,* 217–219.

Himmelfarb, S. On scale value and weight in the weighted averaging model of integration theory. *Personality and Social Psychology Bulletin,* 1975, *1,* 580–583.

Hodges, B. H. Effect of valence on relative weighting in impression formation. *Journal of Personality and Social Psychology,* 1974, *30,* 378–381.

Kanouse, D. E., & Hanson, L. R. *Negativity in evaluation.* Morristown, N.J.: General Learning Press, 1972.

Kaplan, M. F. Context effects in impression formation: The weighted average versus the meaning-change formulation. *Journal of Personality and Social Psychology,* 1971, *19,* 92–99.

Kaplan, M. F. Stimulus inconsistency and response dispositions in forming judgments of other persons. *Journal of Personality and Social Psychology,* 1973, *25,* 58–64.

Kaplan, M. F. Evaluative judgments are based on evaluative information: Evidence against meaning change in evaluative context effects. *Memory and Cognition,* 1975, *3,* 375–380.

Kelly, G. A. *A theory of personality: The psychology of personal constructs.* New York: Norton, 1955.

Lampel, A. K., & Anderson, N. H. Combining visual and verbal information in an impression formation task. *Journal of Personality and Social Psychology,* 1968, *9,* 1–6.

Leon, M., Oden, G. C., & Anderson, N. H. Functional measurement of social values. *Journal of Personality and Social Psychology,* 1973, *27,* 301–310.

McKillip, J., Barrett, G., & Dimiceli, A. J. Trait ambiguity and impression formation: Sufficiency tests of the meaning change model. *Journal of General Psychology,* 1978, *98,* 161–171.

Oden, G. C., & Anderson, N. H. Differential weighting in integration theory. *Journal of Experimental Psychology,* 1971, *89,* 152–161.

Ostrom, T. M. Between-theory and within-theory conflict in explaining context effects in impression formation. *Journal of Experimental Social Psychology,* 1977, *13,* 492–503.

Ostrom, T. M., Werner, C., & Saks, M. J. An integration theory analysis of jurors' presumptions of guilt or innocence. *Journal of Personality and Social Psychology,* 1978, *36,* 436–450.

Rosenbaum, M. E., & Levin, I. P. Impression formation as a function of source credibility and order of presentation of contradictory information. *Journal of Personality and Social Psychology,* 1968, *10,* 167–174.

Rosenberg, S. New approaches to the analysis of personal constructs in person perception. In J. K. Cole and A. W. Landsfield (Eds.), *Nebraska Symposium on Motivation* (Vol. 24). Lincoln: University of Nebraska Press, 1977.

Schmidt, C. F. Personality impression formation as a function of relatedness of information and length of set. *Journal of Personality and Social Psychology,* 1969, *12,* 6–11.

Schümer, R. Context effects in impression formation as a function of the ambiguity of test traits. *European Journal of Social Psychology,* 1973, *3,* 333–338.

Shanteau, J. C., & Anderson, N. H. Test of a conflict model for preference judgment. *Journal of Mathematical Psychology,* 1969, *6,* 312–325.

Simpson, D. D., Ostrom, T. M., & Sloan, L. R. An-

choring effects of trait range in impression formation. *Bulletin of the Psychonomic Society*, 1973, *2*, 383–384.

Troutman, C. M., & Shanteau, J. Do consumers evaluate products by adding or averaging attribute information? *Journal of Consumer Research*, 1976, *3*, 101–106.

Warr, P. B. Inference magnitude, range, and evaluative direction as factors affecting relative importance of cues in impression formation. *Journal of Personality and Social Psychology*, 1974, *30*, 191–197.

Warr, P., & Jackson, P. The importance of extremity. *Journal of Personality and Social Psychology*, 1975, *32*, 278–282.

Warr, P., & Jackson, P. Three weighting criteria in impression formation. *European Journal of Social Psychology*, 1976, *6*, 41–50.

Wyer, R. S. Information redundancy, inconsistency, and novelty and their role in impression formation. *Journal of Experimental Social Psychology*, 1970, *6*, 111–127.

Wyer, R. S. Category ratings as 'subjective expected values': Implications for attitude formation and change. *Psychological Review*, 1973, *80*, 446–467.

Wyer, R. S. *Cognitive organization and change: An information-processing approach*. Potomac, Md.: Erlbaum Associates, 1974. (a)

Wyer, R. S. Changes in meaning and halo effects in personality impression formation. *Journal of Personality and Social Psychology*, 1974, *29*, 829–835. (b)

Wyer, R. S., & Watson, S. F. Context effects in impression formation. *Journal of Personality and Social Psychology*, 1969, *12*, 22–23.

Received November 27, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 11, 2044-2048

The Effect of Situational Contingencies on Female Retaliative Behavior

Deborah Capasso Richardson and Sandy Bernstein
University of Georgia

Stuart P. Taylor
Kent State University

Most studies that have found sex differences in aggression have reported that males are more aggressive than females. Recent evidence, however, suggests that the expectation of female nonaggressiveness may be unwarranted. The present study attempted to reconcile these differences by considering the contingencies of female aggression. Thirty females competed in a task designed to measure aggression (a) alone, (b) in the presence of a silent observer, or (c) in the presence of a supportive observer. Results indicated that as provocation increased, women in the private condition responded more aggressively than did women in the public condition. Also, women who responded in the presence of an audience were more aggressive when the observer was supportive than when she was silent. It is concluded that the usual findings of female nonaggressiveness may be attributable to women's expectations of disapproval for aggressive behavior.

Traditionally, women have been viewed as the nonaggressive sex. Most of the studies that have found sex differences in aggression have reported that males are more aggressive than females. These results have been obtained with children (Liebert & Baron, 1972; Pederson & Bell, 1970; Santrock, 1970; Shortell & Biller, 1970) as well as with adult subjects (Buss, 1966; Doob & Gross, 1968; Epstein, 1965; Youssef, 1968), using various methodologies and different aggression paradigms.

It might appear, then, that females are adhering to a "norm of passivity." According to Maccoby and Jacklin (1974):

> Aggression in general is less acceptable for girls, and is more actively discouraged in them, by either direct punishment, withdrawal of affection, or simply cognitive training that "that isn't the way girls act." Girls then build up greater anxieties about aggression, and greater inhibitions against displaying it. (p. 234)

This response inhibition hypothesis proposes that the female's socialization history causes her to inhibit the display of aggression. Maccoby and Jacklin (1974) suggest that sex differences in aggression may arise because females are reinforced for nonaggressive behavior. Since aggression in general is less acceptable for girls and is more actively discouraged in them by socialization agents, girls have greater anxieties about aggression as well as greater inhibitions against displaying it. Frodi, Macaulay, and Thome (1977), in their review of the literature on sex differences in aggression, note that in the course of normal socialization, women learn to respond to provocation with aggression anxiety and that in our culture aggressive behavior from women is not approved of and receives negative evaluations. Thus, they conclude that "through avoidance or anticipatory arousal of aggression anxiety or guilt, women avoid acting aggressively" (p. 654).

Other evidence, however, suggests that the expectation of female nonaggressiveness may be unwarranted. Taylor and Epstein (1967) found that sex differences in aggression dis-

Requests for reprints should be sent to Deborah Richardson, Department of Psychology, University of Georgia, Athens, Georgia 30602.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3711-2044$00.75

appeared when women were confronted with increasing provocation by a male opponent in a reaction time experiment. Using the same paradigm, Richardson (Note 1) found no sex differences in response to continually high provocation. In addition, Frodi et al. (1977) concluded that evidence indicates that women do *not* show consistently lesser tendencies to physically aggress than do men.

It is apparent that experimental evidence concerning the aggressiveness of women is contradictory. These differences might be reconciled, however, by looking at the environmental contingencies of female aggression. The variability of findings may result from the presence or absence of elements in the experimental situation that evoke conformity to sex role expectations (Frodi et al., 1977). The socialization process that women in our culture undergo has taught them that in many or most situations, aggression will not be approved of and will receive negative sanctions (Costrich, Feinstein, Kidder, Marecek, & Pascale, 1975). Thus women have learned to inhibit their aggressive behavior. At times, however, in spite of childhood socialization and adult sex role restrictions, women do behave aggressively. This negation of traditional sex role requirements can be explained once certain situational factors have been inspected.

Using a competitive reaction time task, Taylor and Epstein (1967) found that when faced with increasing provocation from a male opponent, women would deliver shocks to their opponent at an intensity equivalent to that set by male participants for male opponents. Since the male opponent violated social expectations by delivering relatively high shocks to a woman, the female participants may have felt that they, also, were excused from the usual sex role requirements. Frodi et al. (1977) note that sex differences in aggression seldom appear when aggression is permissible behavior for women. That is, when aggression is justified or when women are acting anonymously—as a deindividuated member of a group—sex differences in aggression are not found.

The contention that women may drop their stereotypical feminine role and behave in a relatively masculine manner also receives support beyond the area of aggression. In a study conducted by Kidder, Bellettirie, and Cohn (1977), reward allocations were made by male and female participants in either a public or a private condition. In the public condition, each sex conformed to its role expectations. When assured of anonymity, however, both sexes abandoned their traditional sex-typed roles.

The present study was designed to investigate the effects of situational contingencies on so-called female nonaggressiveness. Female participants competed against a male opponent in a reaction time task designed to measure aggression. The participants competed in one of three conditions: (a) private —alone in the experimental room; (b) public—in the presence of a silent female observer; or (c) supportive other—in the presence of a female observer who offered social support for retaliating against the opponent.

It was hypothesized that women in the public condition would behave less aggressively in response to provocation than would women in the private condition. The presence of a silent female observer in the public condition was expected to make salient the traditional sex role expectations for *female* nonaggressiveness (e.g., Frodi et al., 1977) and result in behavior that conformed to the participants' perceptions of what the observer expected of them. Women in the supportive other condition, not having to rely on their perceptions of the observer's expectations and knowing that the observer did not expect nonaggressive behavior, were expected to behave more aggressively than would women in the public condition. It was also hypothesized that women in all three conditions would respond more aggressively as provocation by their male opponent increased.

Method

Participants

The participants were 30 female undergraduates enrolled in introductory psychology classes at a large midwestern university. Participation was in partial fulfillment of course requirements. Ten women were randomly assigned to each of the three experimental conditions.

Procedure

Before initiating the procedure, participants were informed that they would be competing in a reaction time task with the person in the next room, that there would be a slight electric shock involved, and that they had the right to leave the experiment at any time without penalty. Participants were then encouraged to ask questions before signing the informed consent statement, which also served as their record of participation.[1]

The female experimenter seated the participant at the task board (described by Taylor, 1967) and attached a concentric shock electrode to her wrist.

In the public and supportive other conditions, a female observer (confederate) arrived at the door of the participant's cubicle and reminded the experimenter that she was a student whose adviser had recommended that she observe an experiment in progress. After obtaining the participant's consent, the experimenter seated the observer slightly behind and to the side of the participant and then left the room for 5 minutes, presumably to attach the opponent's shock electrode.

In the private condition, no observer was present, but the procedure was otherwise identical to the other two conditions.

The participant's "unpleasantness" threshold was then determined. That is, intensity of shock was gradually increased until the participant reported it to be "definitely unpleasant." This was the maximum intensity administered and was designated as Level 5; 90% of the maximum was designated as Level 4; 80% as Level 3; 70% as Level 2; and 60% of the maximum was designated as Level 1. The participant also overheard a tape recording of a male voice telling the experimenter when the shock was unpleasant for him. Thus, she was led to believe that she was to be competing with a male opponent, although no opponent was actually involved in this experiment.

The participant then heard a tape recording of the task instructions. She was told that at the beginning of each trial she should select any one of the five intensities of shock she wished her opponent to receive at the end of the trial if his reaction time was slower than hers and that she would receive the shock her opponent set for her if she was slower than her opponent.

The trials began immediately after the completion of the task instructions. Each trial consisted of four specific events: (a) the onset of the *set* light, which instructed the participant to press one of the five buttons corresponding to the intensity of shock she wished her opponent to receive; (b) the onset of the *press* light, which instructed her to press the reaction time key; (c) the onset of the *release* light, signaling to the participant that she should remove her finger from the reaction time key as quickly as possible; and (d) the onset of one of the five small red lights at the top of the task board, which indicated the intensity of shock that had been set for the participant by her opponent and, if the opponent was faster, the shock as well.

Each participant competed in 25 trials. These consisted of four blocks of 6 trials each plus an extra trial to measure the participant's reaction to Trial 24. The average feedback settings for Blocks 1 through 4 were 1.5, 2.5, 3.5, and 4.5, respectively. Wins and losses and feedback/shock delivered to the participant were programmed by the experimenter. The participant was led to believe that she had won 50% of the trials within each block.

In the supportive other condition, the observer initiated conversation with the participant during the first block of trials, speaking in a friendly, spontaneous manner and sympathizing with her whenever she lost a trial. As the attack increased in the later blocks of trials, the observer attempted to encourage the participant to reciprocate by commenting upon the male opponent's behavior ("he went up there a little bit," "he gave a three!"); indicating that if she were competing with the opponent, she would retaliate ("I wouldn't let him get away with that!", "I don't think I'd let him step all over me"); and suggesting that the participant reciprocate ("Another 4! Maybe you should give it back"). At no time did the observer suggest initiating attacks or setting higher intensity shocks than the opponent. The observer in the public condition remained silent throughout the experimental session.

Following the competition task, women in the public and supportive other conditions were asked to rate the observer on a series of 6-point bipolar adjective scales.

Results

Aggression was measured by the magnitude of shock the participants set for administration to their opponent. Participants were required to select a shock intensity on the first trial without any knowledge of the aggressive intentions of the opponent. As expected, there were no significant differences among conditions in the initial trial, $F(2, 27) = .20$, *ns*.

A 3 (Group) × 4 (Blocks) analysis of variance was performed on mean shock intensities set during the four blocks of trials in order to examine changes in shock settings over blocks. There was no significant main effect for group, $F(2, 27) = 2.24$, *ns*. However, the interaction between groups and blocks was significant, $F(6, 81) = 4.66$, $p <$

[1] This statement indicated that the participant understood (a) the purpose of the experiment, (b) any risks that might be involved, (c) that she could leave the experiment at any time without penalty, and (d) that the details of the experiment would be explained after participation was complete.

.0005. According to Newman–Keuls analyses, as the trials progressed, women in the private and supportive other conditions responded in an increasingly more aggressive manner than did women in the public condition ($p < .05$). Figure 1 demonstrates this relationship and the similarity of the behavior of participants in supportive other and private conditions.

The mean shock intensities set by women in the private condition in Blocks 1 through 4 were 1.52, 1.97, 2.68, and 3.20, respectively. Women in the supportive other condition set mean shock intensities of 1.60, 1.90, 2.79, and 3.57 in Blocks 1 through 4, respectively. However, women in the public condition evidenced less increase by setting mean shock intensities of 1.73, 1.58, 2.12, and 2.03 in Blocks 1 through 4. Further analysis revealed that the differences in mean shock settings were significant only in the final blocks of trials.

As predicted, there was a significant main effect of blocks, $F(3, 81) = 36.22$, $p < .001$, indicating that shock settings, in general, increased over blocks. The mean shock intensities set in Blocks 1 through 4, respectively, were 1.62, 1.85, 2.53, and 2.93.

Following their participation in the competition task, women in supportive other and public conditions were asked to rate the observer on a series of 6-point adjective scales with a midpoint of 3.5. Analysis of this data by t tests indicated that the observer in the supportive other condition was perceived to be significantly more friendly (4.90 vs. 3.90), good-humored (5.50 vs. 4.60), happy (4.40 vs. 3.60), honest (4.60 vs. 3.40), and sociable (4.70 vs. 3.60) than was the observer in the public condition. Also, the observer in the public condition was perceived as being significantly less revengeful (1.40 vs. 3.40), less competitive (2.40 vs. 4.10), less aggressive (2.30 vs. 4.00), and more passive (3.60 vs. 2.00) than was the observer in the supportive other condition (all $ps < .05$).

Discussion

The present study attempted to delineate the effects of situational contingencies on female aggression. It was assumed that women would respond to provocation less aggressively in the presence of a silent audience than when they were either alone or in the presence of a supportive audience. Both hypotheses were supported. Women in the private condition responded to the male opponent's provocation by setting higher shocks than the women who participated in the presence of a silent observer. Similarly, women who were encouraged by the audience retaliated at higher levels than did those who received no such social support.

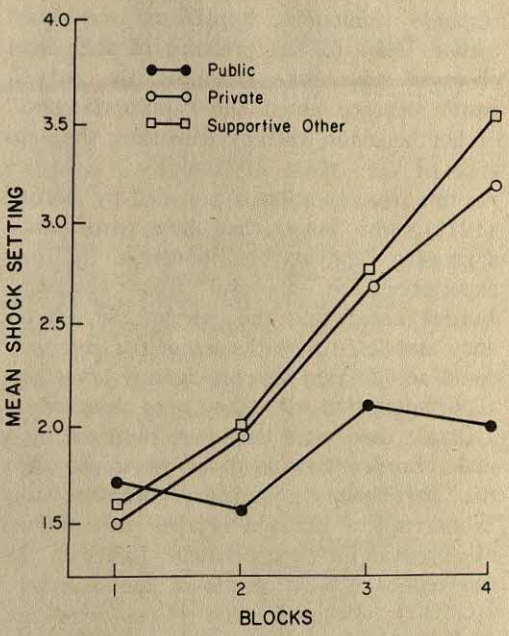


Figure 1. Mean shock settings as a function of group and block.

As predicted, situational contingencies influenced female retaliative behavior, thereby bringing into question the common assumption of female nonaggressiveness. Indeed, it appears that the aggressiveness of women was largely determined by the contingencies present. When reciprocal responding was clearly called for, as in the supportive other condition, women retaliated. When their response was relatively anonymous, as in the private condition, women retaliated. The only situation in which nonaggressiveness was evidenced was in the presence of a silent observer in the public condition. In this case, one might expect that the women were responding to the assumed expectations of the audience.

The relatively low level of aggression in the public condition lends support to the

response inhibition hypothesis mentioned earlier. That is, the presence of the silent observer may have increased the participant's concern about the "appropriateness" of her behavior, thereby enhancing the salience of the "norm of passivity." Evidence for this interpretation is provided by Borden (1975), who found that male participants' aggressive behavior was influenced by their expectations of approval from observers. Borden noted that the "generalized expectancy associated with the sex of the observer" could account for his participants' lower levels of aggression when they were observed by a female than when they were observed by a male. Borden concluded, as we might, that the participants' aggressive behavior was "apparently a function of their expectations of approval for such behavior, based on the inferred or explicit values of the observer" (p. 567). Although such values were not explicit in the present study, participants did attribute characteristics consistent with the "norm of passivity" to the silent observer. They perceived her to be relatively nonrevengeful, noncompetitive, nonaggressive, and passive. The participants in the public condition apparently adapted their retaliative responding in accordance with this expectation.

Similarly, behavior of participants in the supportive other condition was consistent with expectations of the observer. The contingencies in that situation reinforced reciprocation of the opponent's behavior. Thus, expectations were no longer for nonaggressive responding and the female participants could reject the "norm of passivity" and still respond in a manner that would conform to the apparent values of the observer. It was not necessary for these women to inhibit their aggressive behavior due to fear of negative sanction for "out-of-role" behavior.

Reference Note

1. Richardson, D. C. *Sex differences in aggression: A new perspective.* Paper presented at the meeting of the Southeastern Psychological Association, Atlanta, March 1978.

References

Borden, R. Witnessed aggression: Influence of an observer's sex and values on aggressive responding. *Journal of Personality and Social Psychology, 1975, 31,* 567–573.

Buss, A. H. Instrumentality of aggression, feedback, and frustration as determinants of physical aggression. *Journal of Personality and Social Psychology,* 1966, *3,* 153–162.

Costrich, N., Feinstein, J., Kidder, L. H., Marecek, J., & Pascale, L. When stereotypes hurt: Three studies of penalties for sex-role reversals. *Journal of Experimental Social Psychology,* 1975, *11,* 520–530.

Doob, A. N., & Gross, A. E. Status of frustrator as an inhibitor of horn-honking responses. *Journal of Social Psychology,* 1968, *76,* 213–218.

Epstein, R. Authoritarianism, displaced aggression and social status of the target. *Journal of Personality and Social Psychology,* 1965, *2,* 585–589.

Frodi, A., Macaulay, J., & Thome, P. R. Are women always less aggressive than men? A review of the experimental literature. *Psychological Bulletin,* 1977, *84,* 634–660.

Kidder, L. H., Bellettirie, G., & Cohn, E. S. Secret ambitions and public performance: The effects of anonymity on reward allocations made by men and women. *Journal of Experimental Social Psychology,* 1977, *13,* 70–80.

Liebert, R. N., & Baron, R. A. Some immediate effects of televised violence on children's behavior. *Developmental Psychology,* 1972, *6,* 469–475.

Maccoby, E. E., & Jacklin, C. N. *The psychology of sex differences.* Stanford, Calif.: Stanford University Press, 1974.

Pederson, F. A., & Bell, R. Q. Sex differences in preschool children without histories of complications of pregnancy and delivery. *Developmental Psychology,* 1970, *3,* 10–15.

Santrock, J. W. Paternal absence, sex typing, and identification. *Developmental Psychology,* 1970, *2,* 264–272.

Shortell, J. R., & Biller, H. B. Aggression in children as a function of sex of subject and sex of opponent. *Developmental Psychology,* 1970, *3,* 143–144.

Taylor, S. Aggressive behavior and physiological arousal as a function of provocation and the tendency to inhibit aggression. *Journal of Personality,* 1967, *35,* 297–310.

Taylor, S., & Epstein, S. Aggression as a function of the interaction of the sex of the aggressor and the sex of the victim. *Journal of Personality,* 1967, *35,* 474–486.

Youssef, Z. I. The role of race, sex, hostility, and verbal stimulus in inflicting punishment. *Psychonomic Science,* 1968, *12,* 285–286.

Received December 4, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 11, 2049-2058

Emotion Recognition: The Role of Facial Movement and the Relative Importance of Upper and Lower Areas of the Face

John N. Bassili
University of Toronto
Toronto, Canada

In order to investigate the role of facial movement in the recognition of emotions, faces were covered with black makeup and white spots. Video recordings of such faces were played back so that only the white spots were visible. The results demonstrated that moving displays of happiness, sadness, fear, surprise, anger and disgust were recognized more accurately than static displays of the white spots at the apex of the expressions. This indicated that facial motion, in the absence of information about the shape and position of facial features, is informative about these basic emotions. Normally illuminated dynamic displays of these expressions, however, were recognized more accurately than displays of moving spots. The relative effectiveness of upper and lower facial areas for the recognition of these six emotions was also investigated using normally illuminated and spots-only displays. In both instances the results indicated that different facial regions are more informative for different emotions. The movement patterns characterizing the various emotional expressions as well as common confusions between emotions are also discussed.

The communication of emotions through facial expression has received considerable attention since Darwin's (1872) analysis of this phenomenon in humans and animals. A variety of issues have been investigated through the years: Can observers recognize emotions accurately? Are facial expressions genetically prewired? If so, can it be demonstrated that they are universally recognizable? What is the relative contribution of contextual information and facial expressions to the recognition of emotions? Do different emotions define points in a multidimensional space, or are they organized into distinct categories?

Underlying these questions is the problem of delineating the facial patterns that give rise to the judgment of different emotions. This direction has variously been referred to as "psychophysical approach" (Frijda, 1968) or "judgment approach" (Ekman, Friesen, & Ellsworth, 1972). It involves the description of information in the face that leads observers to specific judgments of emotions.

Several studies have investigated this problem by presenting subjects with photographs of emotional expressions and analyzing the relationships between components of the expressions and judgments made by observers (Ekman, Friesen, & Tomkins, 1971; Frijda, 1968; Frijda & Philipszoon, 1963; Frois-Wittmann, 1930; for a review see Ekman et al., 1972). Two features of such experiments are notable. The first is that judgment studies of emotional expression have relied on static representations of facial expressions (photographs).[1] Several authors have criti-

This research was supported by Canada Council Grant S76-1324. The author wishes to thank Bert Forrin for his comments on earlier versions of this article and Brian Kowalchuk for his assistance in the collection of data.

Requests for reprints should be sent to John N. Bassili, Department of Psychology, Scarborough College, University of Toronto, West Hill, Ontario M1C 1A4, Canada.

[1] Ekman and Friesen's (1978) Facial Action Coding System is a notable exception to this assertion. Empirical research is needed, however, to establish correspondences between the system and specific emotions (see Ekman & Oster, 1979, for a brief review).

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3711-2049$00.75

cized the use of such stimuli. For example, Bruner and Taguiri (1954) point out that "it is rare to the vanishing point that judgment ever takes place on the basis of a face caught in a state similar to that provided by a photograph snapped at 20 milliseconds" (p. 638). Nevertheless, Ekman et al. (1972) have argued that frozen representations of the apex of facial movement provide an economical record of emotional expressions. Moreover, many studies using such records have demonstrated the *sufficiency* of static information for the recognition of emotions (see Ekman et al., 1972; Izard, 1971, for reviews).

The second notable feature of judgment studies concerns the type of description of facial information yielded by the analysis of static stimuli. Such descriptions have invariably analyzed the face in terms of the position and shape of features and wrinkles. For example, Frois-Whittmann (1930) analyzed the face in terms of brow, upper lid, lower lid, nostril, and so forth. Other researchers (e.g., Ekman & Friesen, 1975; Frijda & Philipszoon, 1963; Leventhal & Sharp, 1965) have also offered such feature-based analyses of the face. Moreover, Ekman, Friesen, and Tomkins (1971) state that in devising their Facial Affect Scoring Technique, "a decision was made to describe the appearance of the face primarily in terms of wrinkles, of tension or relaxation in specific features, and of positions of features" (p. 40).

Feature-based descriptions derived from static stimuli ignore several forms of facial information relevant to the judgment of emotions. For example, Ekman and Friesen (1975) discuss situations where the rate at which an emotion is expressed modifies the meaning of the expression. One such case is the very brief expression of disgust implying that the person is not actually feeling the emotion but may be referring to something disgusting.

Another kind of facial information overlooked in the analysis of static stimuli concerns the structured deformations of the surface of moving faces. In a recent experiment, Bassili (1978) argued that because facial muscles are fixed in a certain spatial arrangement, the deformations of the elastic surface of the face to which they give rise during facial expressions may be very informative in the recognition of emotions.

In order to investigate facial movement independently of feature-based information that could be captured in a static record of the face, Bassili (1978) covered faces with black makeup and placed white spots in a random order over them. Using a technique pioneered by Johansson (1973), the faces were videotaped and played back on a television monitor adjusted for maximum contrast and low brightness so that only the white spots were visible. The results of Bassili's (1978) experiment demonstrated that observers were more accurate in recognizing several basic emotions on the basis of facial movement alone than would be expected by chance. However, accuracy rates increased significantly when the faces of actors were fully visible.

Bassili's (1978) experiment suffered from two problems. First, because the actors had not been trained in the expression of emotions, recognition rates, even under conditions where the face was normally illuminated, were not always very high (e.g., 31% accuracy in the recognition of expressions of happiness, when 16.7% accuracy could be expected by chance). Second, so many white spots were attached to the faces (around 100) that the outline of features could occasionally be detected. Since no control conditions exhibiting the blackened faces statically at the apex of the expressions were included in the experiment, it is not possible to state with certainty that the better-than-chance recognition rates were caused by motion information alone.

The present experiment investigates the role of facial motion using actors trained in the expression of emotions, as well as conditions minimizing the presence of feature-based information in the records of expressions. The criterion for the latter aim is that the emotions should not be recognizable if the array of white spots is frozen at the apex of expressions. Control conditions using such static arrays were included in the experimental design.

Another issue investigated in the present experiment involves the division of the face

into an upper and lower area. Ekman et al. (1972) have reviewed research investigating the relative importance of different facial areas in judgments of emotions. Of the seven studies relevant to this issue, two suggested that the mouth area is dominant (Dunlap, 1927; Ruckmick, 1921), two found that no facial region is dominant (Coleman, 1949; Frois-Wittman, 1930), and three suggested that different areas are important for different emotions (Hanawalt, 1944; Nummenmaa, 1964; Plutchik, 1962).

These contradictions were probably caused by the fact that few factors were standardized across experiments; the number of actors, their training, and the type of record employed, as well as the emotions portrayed, varied across experiments. For example, Dunlap's (1927) conclusion regarding the superiority of the mouth area was based mainly on judgments of happiness. This finding does not contradict results from experiments suggesting that different areas are important for different emotions (e.g., Hanawalt, 1944, also found the bottom of the face to be dominant in judgments of happiness).

Since the present experiment benefits from recent developments in research on the recognition of emotions (i.e., the use of several trained actors, as well as the choice of six categories of emotions that have received considerable attention in the past), it was decided to investigate again the relative importance of the upper and lower face in judgments of emotions. Moreover, this would afford a comparison of different facial areas under conditions where only facial motion serves as information.

Method

Overview

Six actors were video-recorded while expressing happiness, sadness, fear, surprise, anger, and disgust. The first independent variable (illumination) consisted of either normally illuminated faces or faces covered with black makeup and white spots. A second variable (area) was cross-cut with the illumination variable and consisted of full faces, upper faces, or lower faces. Since six actors expressed each of six emotions in two illumination and three area conditions, a total of 216 moving displays were shown to subjects. Following these displays, subjects were shown 36 static displays depicting the apex of the six emotions expressed by each of the actors in the spots-only/full face mode. Following each display, subjects chose from among the six labeled emotions expressed in the display.

Subjects

Ten male and 10 female volunteers drawn from an introductory course in psychology served as subjects in the experiment. They were paid $5 for their participation.

Stimulus Materials

Twenty students enrolled in a laboratory course in social psychology were trained in the expression of emotions. The students were provided with photographs of emotional expressions prepared by Ekman and Friesen (1975). Each of the emotions—happiness, sadness, fear, surprise, anger and disgust—was represented in two photographs. Underneath the photographs was a written description of the facial features critical to an expression. Students were given a week to practice their skills at duplicating the expressions shown in the photographs. At the next meeting, students paired up and helped each other perfect their expressions. Each student was then video-recorded while expressing the six emotions. The quality of each expression was later rated by all students in the class with the exception of the actor. The six actors receiving highest ratings (four males and two females) were retained for further class research and were later approached for participation in the present experiment.

Each actor was first recorded prior to the application of makeup. The expression of the six emotions was recorded under three conditions: full face, upper half of the face, and lower half of the face. The division of the face into upper and lower regions was accomplished by having the actor hold a rectangular piece of cardboard in front of the area to be masked. The horizontal edge of the mask was placed slightly below the bridge of the nose. Next, the actors were made up in black, and about 50 white spots, 8 mm in diameter, were affixed to their faces in a haphazard or quasi-random arrangement. The recording procedure was then repeated.

These recordings yielded 216 single displays of an emotion, each lasting about 3 sec. The order of the expressions was randomized and divided into two blocks of 108 displays. The order of presentation of the two blocks was counterbalanced across subjects. These presentations were preceded by 36 practice trials and were followed by 36 control trials. The control displays consisted of a 30-sec static presentation of the apex of each of the six emotions expressed by the six actors under spots-only/full face conditions. The order of these displays was also randomized.

Procedure

Subjects took part in the experiment in pairs and were separated by a large partition. They were seated 4 ft. (1.2 m) from two 11-in. (28-cm) Sony television monitors. One monitor was adjusted for normal viewing, whereas the other had its brightness and contrast levels adjusted so that only the white spots were visible in spots-only conditions. The experimenter switched the monitors on and off for the appropriate displays.

Each display was followed by a 10-sec interval for responses. The responses were made on a short questionnaire containing 300 sets of the six emotion labels. Subjects simply checked which of the six labels corresponded best to the emotion portrayed in a display. They were warned at the outset of the experiment that only white spots that had been placed on the faces of actors would be visible in certain displays, whereas other displays would only show the upper and lower half of the face. Moreover, the practice trials contained an example of each of the conditions.

Results and Discussion

The results can be presented most clearly if the following three issues are considered in order: comparison of accuracy rates in moving and static spot displays; analysis of accuracy rates, using a 2 (Sex) × 2 (Illumination) × 3 (Facial Area) × 6 (Emotion) repeated measures analysis of variance (ANOVA); and analysis of movement patterns and confusions. The accuracy rates used in the statistical analyses were computed by giving subjects a score representing the proportion of displays of an expression, as portrayed by the six actors, that they recognized accurately.

Comparison of Moving and Static Displays

One of the main questions addressed by the present research concerns the sufficiency of motion information for the recognition of emotions. To answer this question, the accuracy rates yielded by spots-only/full face displays were compared with those yielded by static displays of the same faces at the apex of expressions. A 2 (Sex) × 2 (Motion) × 6 (Emotion) ANOVA with repeated measures on the last two factors was used for this analysis. It revealed that moving displays yielded on the average ($M = 52.7\%$) significantly higher recognition rates than static displays did ($M = 29.5\%$), $F(1, 90) = 192.29$, $p < .01$. The effect of sex and emotion will receive attention in the next section. Table 1 shows the mean accuracy rates relevant to the present analysis.

Since the ANOVA also received a significant interaction between motion and emotion, $F(5, 90) = 11.37$, $p < .01$, tests of simple main effects were performed to verify the effect of motion on each emotion individually. These tests indicated that the recognition rate in the case of all six emotions under movement was superior to the recognition rate under static conditions, $F(1, 18) > 9.4$, $p < .01$ in all cases. The recognition rate in each static condition was also compared with expected chance performance. These comparisons revealed that only in the case of happiness and surprise was recognition of static displays significantly more accurate than chance, $t(19) = 2.83$ and $t(19) = 7.59$, respectively, $p < .01$, one-tailed in both cases.

Accuracy Rates in Overall Factorial Design

The next two questions addressed by the present research concern the comparison of normally illuminated moving faces with faces where only facial motion is visible, and the

Table 1
Recognition Accuracy in Moving and Static Displays

Emotion	Happiness	Sadness	Fear	Surprise	Anger	Disgust	*M*
Face							
Moving	90.0	44.2	43.3	88.3	39.2	53.3	59.7
Static	30.8	25.8	12.5	47.5	16.7	21.7	29.4
Difference	59.2	18.4	30.8	40.8	22.5	31.6	
M	60.4	35.0	27.9	67.9	27.9	37.5	

Note. Numbers represent recognition accuracy. The accuracy rate expected by chance is 16.7.

lative effectiveness of the upper and lower gions of the face for the recognition of the rious emotions.

A 2 (Sex) × 2 (Illumination) × 3 (Facial rea) × 6 (Emotion) repeated measures NOVA was used to analyze subjects' accuracy ites. All main effects were significant. Female ubjects ($M = 68.1\%$) were on the average ore accurate than male subjects ($M = 1.6\%$), $F(1, 18) = 6.85$, $p < .025$. Displays f normally illuminated faces ($M = 77.3\%$) ielded higher recognition rates than displays f moving spots did ($M = 52.3$), $F(1, 18) = 70.31$, $p < .001$. Full face displays ($M = 4.4\%$) were recognized more accurately than ere displays of the bottom of the face ($M = 4.9\%$), which were in turn recognized more ccurately than displays of the top of the face $M = 55.1\%$), $F(2, 36) = 50.40$, $p < .001$ multiple comparisons by Scheffé tests yielded values smaller than .01 in all cases). Finally, the main effect due to emotions, $F(95, 0) = 20.50$, $p < .001$, indicated that surprise ($M = 82.2\%$) and happiness ($M = 9.6\%$) were recognized more accurately than ere sadness ($M = 60.8\%$), anger ($M = .9\%$), fear ($M = 57.6\%$), and disgust $M = 49.7\%$) ($p < .01$ by Scheffé tests in 1 cases).

Three two-way interactions were also significant. An Illumination × Emotion interaction indicated that the superiority of normally illuminated displays over spots-only displays was evident in anger, sadness, disgust and fear but not in surprise and happiness, $F(5, 90) = 58.94$, $p < .001$. Moreover, an Illumination × Facial Area interaction indicated that the superiority of normally illuminated displays was particularly pronounced under conditions where the full face was presented, $F(2, 36) = 5.47$, $p < .01$. Finally, a Facial Area × Emotion interaction indicated that the lower region of the face was superior to the upper in the recognition of happiness and disgust, the upper region was superior to the lower in the recognition of anger and fear, and facial region made little difference in the recognition of surprise and sadness, $F(10, 180) = 10.59$, $p < .001$.

The three-way interaction between illumination, facial area, and emotion was also significant, $F(10, 180) = 4.98$, $p < .001$. It reflected the fact that a facial area dominant in a normally illuminated display was not necessarily dominant in a spots-only display (see Figure 1).

The results yielded by this analysis are relevant to several issues addressed by past

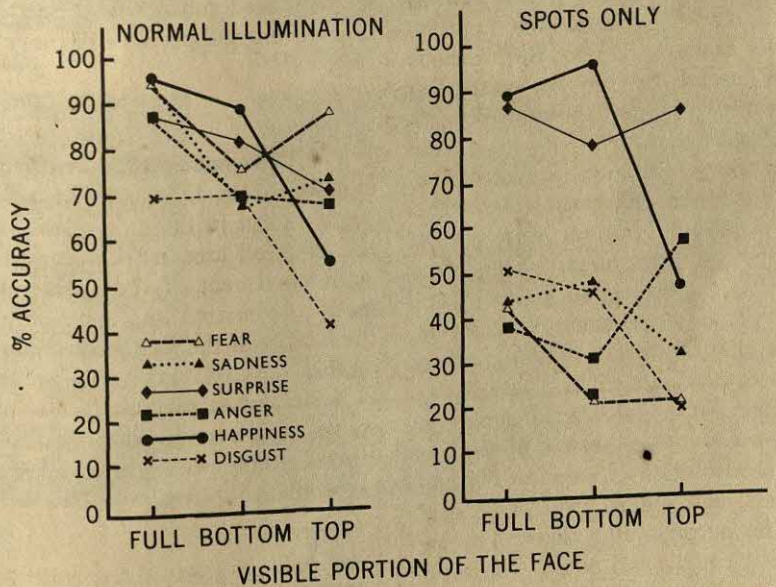


Figure 1. Recognition rates for the six emotions yielded by the three facial areas under normal illumination and spots-only conditions.

research. First, the main effect due to emotions can be compared with a number of studies using posed expressions. For example, Ekman et al. (1972) present a table of "Accuracy Studies of Posed Behavior" (p. 103). Averaging over the studies described in this table, happiness appears to be the easiest emotion to recognize. The results of the present experiment are in basic agreement with this, though surprise achieved an even higher accuracy rate. On the other side of the scale, the recognition of disgust was particularly poor in the present study. This was not generally the case in past research. The reason for the inferiority of disgust in the present study is not clear and could be due to problems in the training of actors. Another finding relevant to past research has to do with the fact that women subjects in the present study achieved significantly higher recognition rates than men did. This finding is consistent with past research (Ekman & Oster, 1979).

The results yielded by the normally illuminated displays are of particular relevance to past research on the relative importance of the upper and lower regions of the face for emotion recognition. The present results suggest that the bottom of the face is more useful than the top for the recognition of happiness, surprise, and disgust, whereas the opposite is indicated for sadness and fear. Both regions were equally useful for the recognition of anger. These results support the general proposition that the importance of upper and lower facial areas for emotion recognition depends on the emotion (Ekman et al., 1972). Given the variability and inconsistency of results yielded by past research concerned with the relative importance of various facial areas in the recognition of emotions, however, it is difficult to compare the present findings with specific results yielded by past studies.

In conditions where only facial movement serves as information, the bottom of the face seems more useful for the recognition of happiness, sadness, and disgust, whereas the reverse holds for surprise and anger. The two regions yielded equal (and poor) recognition rates for fear. The results should be approached with caution because the recognition rates with the exception of happiness and surprise were relatively low and in some cases did not differ substantially from expected chance performance. The type of facial motion information relevant to the various emotions is, however, considered further in the next section.

Movement Patterns and Confusions

This section analyzes the patterns of facial movement yielded by the expression of six emotions. The analysis is accomplished with the help of photographs depicting the paths of the white spots over the course of the expressions (see Figure 2). In addition, a descriptive analysis of major confusions is undertaken (see Table 2). For this purpose an arbitrary cutoff point of 20% was used to isolate emotions that were often confused with a target emotion.

Happiness

The movement pattern yielded by the expression of happiness consists of an upward displacement of each side of the mouth and the cheeks. This movement results from a smile. Following the 20% criterion, happiness was confused substantially only with sadness in spots-only/upper face displays (23.3%).

Sadness

The movement pattern yielded by sadness is more subtle. It consists of a slight upward displacement in the area of the chin, whereas the forehead area reflects an inward and upward movement of the eyebrows. Both normally illuminated and spots-only displays of the bottom of the face yielded confusions with disgust (23.3% and 30.0%, respectively). However, when the full face or top face was shown in spots-only displays, confusions occurred mainly with fear (22.5% and 25.8%, respectively).

Fear

Fear involves a downward and outward movement in the mouth area. The forehead

movement is similar to that of sadness (inward and upward), although the movement is more pronounced because the brows are raised higher. The similarity of movement in the forehead area in sadness and fear is reflected by the fact that in spots-only/upper face conditions fear was often confused with sadness (24.2%). In addition, fear in these displays was confused with surprise (25.8%). This latter confusion was probably caused by the strong upward movement of the brows characterizing the expression of surprise. Another common error yielded by the expression of fear involved happiness and occurred in spots-only/lower face displays (35.8%).

Surprise

Surprise was one of the easiest emotions to recognize, and was particularly so in spots-only displays. The expression involves a strong upward displacement of the brows and an equally strong downward displacement of the jaw. The rapidity of the movement (akin to a startle) is also specific to this emotion. Major confusions only occurred with fear in normally illuminated/upper face displays (28.3%).

Anger

The expression of anger involves a downward movement in the forehead area caused

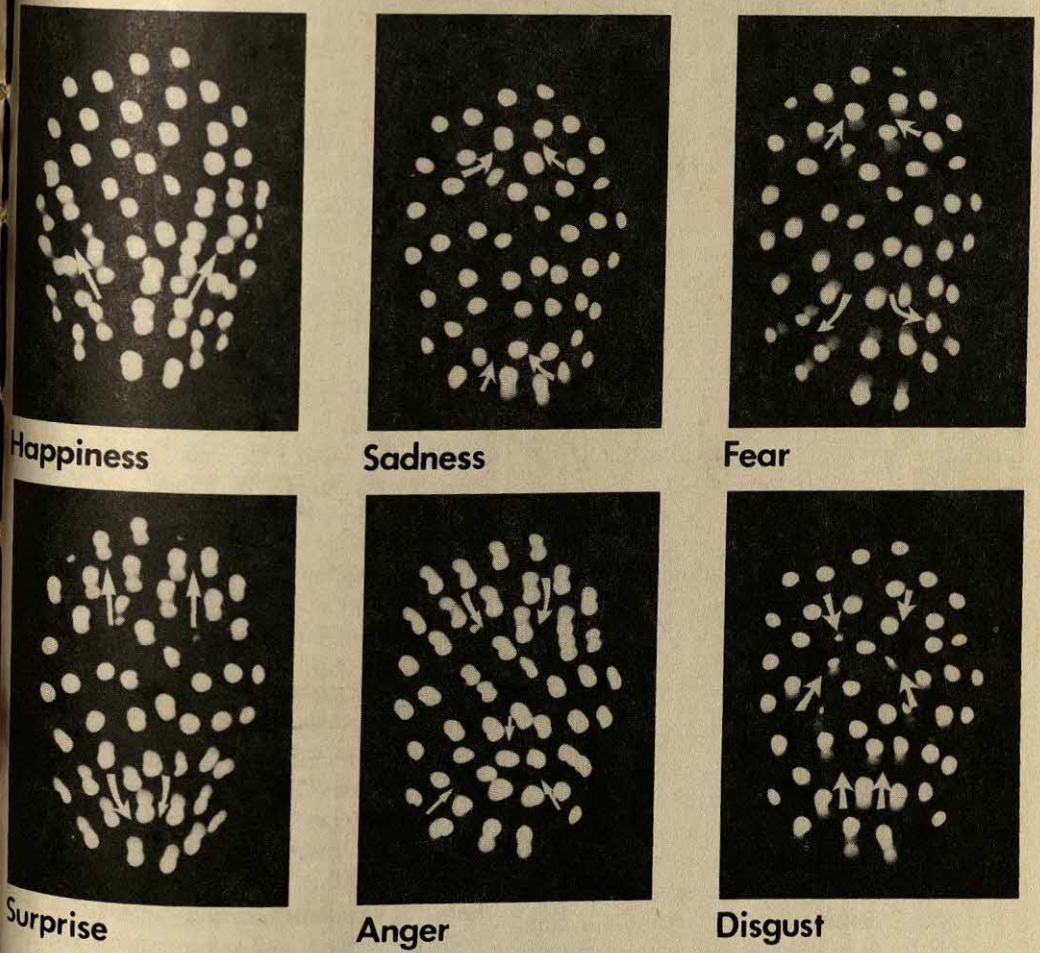


Figure 2. Movement patterns yielded by each of the six emotions investigated. (These photographs represent time exposures of actual television displays. The shutter was opened just before the onset of an expression and was closed just after the expression reached its apex. The arrows indicate the direction of movement.)

Table 2
Confusion Matrices

Stimulus	Response					
	Happiness	Sadness	Fear	Surprise	Anger	Disgust
Full face						
Happiness	97.5/90.0	0/.8	2.5/3.3	0/1.7	0/.8	0/3.3
Sadness	0/1.7	97.5/45.8	.8/22.5	0/5.0	0/12.5	1.7/12.5
Fear	0/10.0	1.7/14.2	95.8/43.3	0/16.7	.8/7.5	1.7/8.3
Surprise	1.7/1.7	0/.8	10.8/7.5	87.5/88.3	0/1.7	0/0
Anger	0/0	.8/32.5	2.5/5.8	0/4.2	87.5/39.2	9.2/18.3
Disgust	0/1.7	.8/2.5	.8/5.8	0/1.7	28.3/36.7	70.0/51.7
Bottom half						
Happiness	90.0/97.5	.8/0	8.3/0	0/1.7	0/0	.8/.8
Sadness	0/2.5	68.3/48.3	0/4.2	0/0	8.3/15.0	23.3/30.0
Fear	.8/35.8	6.7/10.8	75.8/19.2	1.7/2.5	9.2/9.2	5.8/22.5
Surprise	.8/.8	0/3.3	16.7/11.7	82.5/79.2	0/.8	0/4.2
Anger	.8/10.0	7.5/20.0	.8/7.5	0/7.5	70.0/30.0	20.8/25.0
Disgust	0/7.5	1.7/3.3	0/20.0	0/9.2	28.3/15.8	70.0/44.2
Top half						
Happiness	55.0/47.5	13.3/23.3	8.3/15.8	.8/2.5	8.3/5.0	14.2/5.8
Sadness	0/2.5	74.2/32.5	15.0/25.8	0/10.0	5.8/16.7	5.0/12.5
Fear	0/1.7	0/24.2	89.2/20.8	10.0/32.5	0/13.3	.8/7.5
Surprise	.8/0	0/.8	28.3/11.7	70.0/85.8	0/0	.8/1.7
Anger	0/0	2.5/23.3	20.0/6.7	1.7/1.7	69.2/56.7	6.7/11.7
Disgust	.8/0	3.3/18.3	2.5/10.0	0/5.8	51.7/46.7	41.7/19.2

Note. The numbers represent the percentage of choices of a column label in response to the display of an emotion described by a row label; numbers on the left of the slash are for responses to normally illuminated displays, and those on the right are for responses to spots-only displays.

by a frown, along with a compression in the mouth area, caused by the pinching of the lips. Anger was confused with disgust in lower face displays (both normally illuminated and spots-only conditions, 20.8% and 25.0% respectively). The normally illuminated/upper face displays yielded confusions with fear (20.0%). Finally, all spots-only displays yielded confusions with sadness (full face = 32.5%, lower face = 20.0%, upper face = 23.3%).

Disgust

The expression of disgust involves the wrinkling of the nose, which causes an upward movement on its sides as well as on the cheeks. Moreover, the expression can involve an upward movement in the area of the chin. Though the wrinkling of the nose is specific to disgust, this emotion leads to low recognition rates. Confusions, however, seemed to concentrate on anger. In all conditions except spots-only/bottom face displays (15.8%), anger was taken for disgust on more than 28% of the trials. This confusion was particularly pronounced in upper face displays (normal illumination = 51.7%, spots-only = 46.7%). The spots-only/bottom face displays lead to confusion with fear (20.0%).

One possible cause of confusions among emotions is that their expressions are so similar that they are difficult to discriminate. If this is the case, confusions should be reciprocal. This was seldom the case in the confusions just described. Reciprocity was only evident in two cases: fear and sadness in spots-only/upper face displays and disgust and anger in normal illumination/bottom face displays. Another possibility is that the actors failed to represent emotions adequately. This possibility, with one exception, is not very likely, since recognition rates in normally illuminated/full face displays were very high (above 87%). The exception involves expres-

ons of disgust (70%), which were often idged as anger.

The third possibility accounting for non-eciprocal confusions is that specific expres-ons may have been nondistinct or ambig-ous. In resolving the ambiguity, subjects ay have responded with an emotion sharing me components with the target emotion. he expression of that alternate choice may ave also been ambiguous and shared compo-ents with yet other expressions. In short, two mbiguous expressions need not be resolved in reciprocal manner.

Conclusion

The present experiment demonstrates learly that movement of the surface of the ace can serve as information for the recogni-ion of emotions. It is not proposed, however, hat such information is generally necessary or the purpose. Much past research using tatic stimuli has demonstrated that facial eatures, with their specific shapes and spatial rrangement, also provide very rich informa-ion for the recognition of emotions. Thus, it ppears that under natural conditions of ob-ervation the observer has access to a redun-lant system of facial information.

The issue of redundancy is complex and is manifested in a variety of ways in emotion ecognition. For example, although the pres-nt research has emphasized the distinction between feature-based static information and purely transformational movement informa-tion, each of these types of information is in turn internally redundant. Specifically, most emotions are represented in specific ways in various areas of the face. It is precisely this characteristic that has led some researchers to inquire about the relative importance of different areas of the face in emotion recogni-tion.

The findings yielded by such research, how-ever, have been complex and inconsistent. This has probably been caused by the fact that emotional expressions contain many cues that serve as information for recognition. By dividing the face into two or three regions, researchers have presented observers with selected subgroups of cues. Some more or less distinctive cues have fallen into the various groups as different emotions were considered, leading to the finding that no facial region dominates across emotions. In order to avoid this haphazard approach to the analysis of dominance in facial information, it would appear desirable for future research to de-lineate the cues relevant to expressions of specific emotions and to scale their relative information value.

Another issue relating to redundancy and more directly relevant to the distinction be-tween feature information and movement in-formation has to do with the nature of their cooperation in the specification of emotions. Specifically, is the observer in a better posi-tion to recognize emotions accurately on the basis of both kinds of information than he/she would be by using the better of the two alone? This question is particularly important in the context of the two types of information discussed here, since they are perfectly cor-related, that is, cannot vary independently. Although an empirical answer to this question will have to await future research, it is likely that the advantage of redundant information manifests itself under poor conditions of ob-servation. For example, movement may render slight or brief expressions more detectable by highlighting the contrast between the appear-ance of the face at two points in time. In a reciprocal manner, features that tend to be points of high contrast on the face may serve as particularly effective "carriers" of motion, thus rendering movement information more detectable.

Research aimed at verifying these specula-tions would have to take another issue into account. In the present research it was deemed appropriate to establish the usefulness of movement information under optimal con-ditions for the recognition of emotions. Thus, the posed "stereotypic" expressions of trained actors were used in the generation of the stimulus displays. When the concern shifts to the specific interaction of feature and move-ment under less than ideal conditions of ob-servation, the research should utilize more ecologically valid stimuli.

References

Bassili, J. N. Facial motion in the perception of faces and of emotional expression. *Journal of Experimental Psychology: Human Perception and Performance,* 1978, *4,* 373–379.

Bruner, J. S., & Taguiri, R. The perception of people. In G. Lindzey (Ed.), *Handbook of social psychology* (Vol. 2). Reading, Mass.: Addison-Wesley, 1954.

Coleman, J. C. Facial expressions of emotion. *Psychological Monographs,* 1949, *63*(1, Whole No. 296).

Darwin, C. *The expression of the emotions in men and animals.* London: John Murray, 1872.

Dunlap, K. The role of eye-muscles and mouth-muscles in the expression of the emotions. *Genetic Psychology Monographs,* 1927, *2,* 199–233.

Ekman, P., & Friesen, W. V. *Unmasking the face.* Englewood Cliffs, N.J.: Prentice-Hall, 1975.

Ekman, P., & Friesen, W. V. *The Facial Action Coding System.* Palo Alto, Calif.: Consulting Psychologists Press, 1978.

Ekman, P., Friesen, W. V., & Ellsworth, P. *Emotion in the human face.* New York: Pergamon Press, 1972.

Ekman, P., Friesen, W. V., & Tomkins, S. S. Facial Affect Scoring Technique (FAST): A first validity study. *Semiotica,* 1971, *3,* 37–58.

Ekman, P., & Oster, H. Facial expressions of emotion. *Annual Review of Psychology,* 1979, *30,* 527–554.

Frijda, N. H. Recognition of emotion. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 4). New York: Academic Press, 1968.

Frijda, N. H., & Philipszoon, E. Dimensions of recognition of emotion. *Journal of Abnormal and Social Psychology,* 1963, *66,* 45–51.

Frois-Wittmann, J. The judgment of facial expression. *Journal of Experimental Psychology,* 1930, *13,* 113–151.

Hanawalt, N. G. The role of the upper and the lower parts of the face as the basis for judging facial expressions: II. In posed expressions and "candid camera" pictures. *Journal of General Psychology,* 1944, *31,* 23–36.

Izard, C. E. *The face of emotion.* New York: Appleton-Century-Crofts, 1971.

Johansson, G. Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics,* 1973, *14,* 201–211.

Leventhal, H., & Sharp, E. Facial expressions as indicators of distress. In S. S. Tomkins & C. E. Izard (Eds.), *The communication of emotional meaning.* New York: McGraw-Hill, 1965.

Nummenmaa, T. *The language of the face. Jyvaskyla studies in education, psychology and social research.* Jyvaskyla, Finland: Jyvaskylan Yllopistoyhdistys, 1964.

Plutchik, R. *The emotions: Facts, theories, and a new model.* New York: Random House, 1962.

Ruckmick, C. A. A preliminary study of the emotions. *Psychology Monograph,* 1921, *30* (Nos. 134–139), 30–35.

Received December 5, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 11, 2059-2070

Psychological Masculinity and Femininity and Typical and Maximal Dominance Expression in Women

Helen M. Klein and Lee Willerman
University of Texas at Austin

Four groups of women differing in psychological masculinity and femininity as measured by the Personal Attributes Questionnaire, participated in two laboratory situations designed to measure their typical and maximal dominance expression. Typical dominance refers to one's average expression of the trait, whereas maximal dominance refers to one's response capability regarding the trait. As predicted, undifferentiated and feminine women were significantly less dominant than masculine and androgynous women were in both laboratory situations and as measured by the California Psychological Inventory (CPI) Dominance scale. Traditional prescriptions for sex role behavior also affected women's performance in the typical laboratory, where they were significantly less dominant with male confederates than with female confederates, but not in the maximal laboratory, where they were equally dominant with male and female confederates. However, women were more talkative with male confederates rather than female confederates in both laboratory situations. Multiple regression analysis revealed that the CPI Dominance scale was the most effective predictor of laboratory behavior.

Traditional conceptualizations of masculinity and femininity have become the target of much criticism. Many contemporary psychologists have rejected the unidimensional bipolar model in favor of a dualistic approach, that masculinity and femininity are independent dimensions and coexist to some degree in everyone (Bem, 1975; Block, 1973; Constantinople, 1973; Spence, Helmreich, & Stapp, 1974).

One major source of dispute in this literature is the diversity of definitions of "sex roles." Under this heading, investigators have included all of the behaviors, attitudes, preferences, and internal dispositions differentiating men from women. This state of affairs can be attributed in part to the widespread assumption that "biological gender, masculine and feminine sex-role behaviors, and the psychological attributes of masculinity and femininity are tightly intercorrelated" (Spence & Helmreich, 1978, p. 10).

If these qualities are only weakly related, however, it becomes imperative to distinguish between masculine and feminine sex roles (defined as overt behaviors deemed appropriate for members of each sex in certain situations) and internal dimensions of masculinity and femininity (defined as agentic and communal traits, Bakan, 1966).

A study by Megargee (1969) on dominance expression illustrates this distinction nicely. When high dominant women (as defined by the California Psychological Inventory or CPI) were paired with low dominant men on a leader–follower task, the women did not often take on the role of leader. These women, however, actively chose the leader 69% of the time, and when they did, they selected the men 90% of the time. In this way the women asserted their dominance without violating traditional sex role standards.

What if traditional sex role prescriptions

The authors thank Robert L. Helmreich, Janet T. Spence, and Robert G. Turner for advice and assistance in this study.

Requests for reprints should be sent to Lee Willerman, Department of Psychology, University of Texas at Austin, Austin, Texas 78712.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3711-2059$00.75

were relaxed? Would correlations between psychological dimensions of masculinity or femininity and behavior become more or less pronounced? One way to do this is by requiring subjects to perform under maximal conditions (Fiske & Butler, 1963; Wallace, 1966; Willerman, Turner, & Peterson, 1976).

These authors have distinguished between typical and maximal performance. The former refers to one's usual or average behavior, and the latter refers to one's maximum expression of the trait-linked behavior under conditions designed to elicit it to the greatest degree. Traditional laboratory situations, based on a typical performance format, often produce idiosyncratic differences in levels of motivation as well as differences in the interpretations of the task at hand. Laboratory situations based on a maximal performance format, however, explicitly encourage the individual to focus on one specified extreme of a target trait to determine what he or she can do. This ought to lead to greater homogeneity in motivational level and clarity about the demands of the situation. In this context, socially undesirable or inappropriate behaviors can also be made more socially desirable. Consequently, sex role stereotypes might be diminished.

Finally, in line with a view of personality that emphasizes response capability, it may be important for individuals to *generate* responses when describing themselves, rather than simply endorsing items presented to them. Turner and Gilliland (in press) found that self-generated trait descriptions significantly predicted both typical and maximal laboratory dominance behavior in college men and women. Those individuals describing themselves as dominant expressed the most dominance and those describing themselves as reserved the least, whereas those not using adjectives on the dominance dimension fell in between. No information was available on the psychological masculinity or femininity of the subjects. Since it is now clear that both dimensions can coexist within the same individual (e.g., Spence et al., 1974), it would seem appropriate to examine whether these traits interact with typical and maximal situations.

The Present Study

This study examined the effects of psychological masculinity and femininity and sex role demands on the typical and maximal expression of dominance in women. Self-generated trait descriptions and self-reports of dominance were obtained from four groups of women: undifferentiated (low masculine, low feminine), feminine (low masculine, high feminine), masculine (high masculine, low feminine), and androgynous (high masculine, high feminine). These women then participated in two laboratory paradigms designed to assess their typical and maximal dominance behavior. Within these paradigms, the women were exposed to either male or female confederates. From the studies by Bem (1975) and Bem, Martyna, and Watson (1976), it was expected that feminine and undifferentiated women would have lower scores on a self-report dominance measure and would express less dominance in the typical laboratory situation than would masculine and androgynous women. Following Megargee's (1969) findings, it was predicted that women would express less dominance with male confederates than with female confederates in the typical laboratory situation. From Turner and Gilliland's (in press) study, it was expected that self-descriptive trait lists would predict dominance expression in both laboratory situations. It was expected that women who listed adjectives on the dominance dimension as self-descriptors (e.g., dominant, reserved) would be more extreme in dominance expressions than women who did not do so.

Although no specific predictions were made for the maximal laboratory situation, several general questions were proposed for investigation: How would feminine and undifferentiated women, compared to androgynous and masculine women, respond when dominance is made explicitly salient and desirable (maximal laboratory)? Would women respond differently to male and female confederates in the maximal laboratory? Would they express more dominance with the female confederates, as hypothesized for the typical laboratory?

Method

Design and Overview

The basic design was a 4 × 2 × 2 factorial with two between-subjects variables and one within-subjects variable. One hundred and twelve women were recruited as representatives of four combinations of psychological masculinity and femininity based on their scores on the short form of the Personal Attributes Questionnaire (PAQ; Spence & Helmreich, 1978). The women completed a variety of self-report measures and then each participated with two pairs of same-sex confederates in two group problem-solving situations designed to assess her typical and maximal dominance expression. Half of the women in each of the four groups were exposed only to male confederates and half only to female confederates.

Subjects

Several weeks prior to the beginning of the experiment proper, the short form of the PAQ was administered to 635 female introductory psychology students at the University of Texas at Austin. Using norms established by Spence and Helmreich (1978), the women were classified into four groups based on median splits on their PAQ masculinity (M) and femininity (F) scores: undifferentiated (low M, low F), feminine (low M, high F), masculine (high M, low F), and androgynous (high M, high F). The total pretesting sample was comparable to Spence and Helmreich's sample in mean PAQ scores, percentage of students falling into each M and F group, and relationship among the PAQ scales.

A total of 123 women were recruited for the present study from eligibility lists for the four M and F groups. Eleven women were dropped from the experiment: 3 because of suspicion, 3 because of equipment failure, and 5 because they failed to appear for their appointments. Half of the 28 women in each M and F group were randomly assigned to one of the two sex-of-confederates conditions. The experimenter and the confederates were blind with respect to the women's classifications. Each subject received course credit for participation in the experiment.

Measures

Personal Attributes Questionnaire. The PAQ is a measure of psychological masculinity and femininity developed by Spence et al. (1974). The short version of the PAQ (Spence & Helmreich, 1978) consists of 24 bipolar items belonging to three scales: masculine (M), feminine (F), and masculine–feminine (M–F). Correlations between the scales of the short version and the original 55-item PAQ are .93 (M), .93 (F), and .91 (M–F) for college students. For each item, the women were asked to rate themselves on a 5-point scale, with the endpoints having verbal labels. Items were scored from 0 to 4, with M and M–F items keyed in a "masculine" direction and F items keyed in a "feminine" direction. The 8 item scores for each scale were summed to give M, F, and M–F scores for each person. Some examples of PAQ items are "not at all independent – very independent" (M), "not at all emotional – very emotional" (F), and "very submissive – very dominant" (M–F).

Dominance scale of the California Psychological Inventory. The CPI Dominance scale (Gough, 1956) consists of 46 items pertaining to dominance in a "true–false" format, such as "I doubt whether I would make a good leader" (false) and "I think I would enjoy having authority over other people" (true). The total score for each person was obtained by summing the number of items responded to in the significant direction.

Person Description Form. This form (adapted from Turner & Gilliland, in press) is an idiographic assessment technique calling for the generation of a series of self-descriptive adjectives. Instructions are to "list the adjectives or traits that best describe you and your personality."

Following the procedure used by Turner and Gilliland, the women were classified into four groups via their responses, based on the taxonomy developed by Layman and McDonald (*The Layman–McDonald Trait Taxonomy: Round II;* cited in Turner & Gilliland, in press). The fourth group was added by the present investigators for exploratory reasons. Disagreements on categorization were resolved by discussion.

1. Dominant group—women listing traits categorized as autocratic/domineering (Layman–McDonald Dimension #7) or self-assertive (Dimension #20).
2. Reserved group—women listing traits on the opposite poles of these dimensions (e.g., democratic, reticent).
3. Trait irrelevant group—women not listing traits on these dimensions.
4. Mixed group—women listing traits on both the dominance and reserved dimensions.

Laboratory Measures

Two methods were used to assess dominance behavior in the laboratory (Turner, 1978).

Dominance rating. The *contents* of the subject's verbalizations (i.e., forceful, assertive statements) were independently rated by the two confederates as soon as she left the room. These ratings were made according to the following rationally constructed scale (adapted from Turner, 1978):

1. No dominance statements—totally passive.
2. Statements of a general nature (comments, statements of fact, clarifying statements); a few very weak opinions voiced with little conviction of involvement.
3. A few opinions offered; no disagreements with others expressed. May agree with others' opinions.
4. Opinions offered but no attempt to lead group (or very weak one); disagreements subtle and nondirect. Might offer suggestions about organizing.

5. Opinions offered with attempts to get some agreement from others; somewhat persuasive efforts toward other members. Statements solicitous to support, such as "Don't you think. . . . ?" An organizer —democratic—asks what others think (How do you feel? How about . . . ? What do you think?).

6. Strong attempts to persuade others and open disagreement with others. Emphatic statements, such as "I think. . . ." Person may take a while to warm up but takes over eventually. Assertive about opinions; leads; not too concerned about what others think.

7. Complete usurpation of leadership of group with insistence upon influencing conclusions of group. Usurpation of leadership from the *beginning. Strong* opinions; not democratic; very persuasive; strong disagreements.

The two confederates' ratings were summed together to yield a composite dominance rating for the typical laboratory and one for the maximal laboratory for each subject. Intraclass interrater reliabilities were .96 for the typical laboratory and .95 for the maximal laboratory.

Prior to the experiment proper, all confederates (five men and five women) participated in a series of laboratory training sessions to familiarize themselves with their roles and with the dominance rating scale. They practiced in different-sex as well as same-sex pairs with 20 pilot subjects. Tapes of the sessions were reviewed and discussed to promote consistency in their ratings.

The dominance rating sheet also included a 5-point scale for *facial* attractiveness, ranging from (1) very low to (5) very high, on which each subject was independently rated by the confederates. The confederates used an idealized *Vogue* model as a standard for these ratings. The summed ratings of the four confederates (two from each laboratory) served as a facial attractiveness score for each subject. Although the agreement among the raters on this measure was rather low (intraclass reliability = .47), the reliability of the average of the four raters was substantially higher (.78), suggesting moderate reliability for the overall attractiveness rating of each subject.

Talking time. The second method of assessing typical and maximal dominance in the laboratory was to measure the *amount* of verbalization, that is, the extent to which the subject monopolized the conversation. This was obtained from tape recordings of the laboratory sessions. Two confederates independently reviewed the tapes to determine percentage of talking time during the problem-solving period in the typical laboratory and in the maximal laboratory for each subject. Confederates' times were within 12 sec of each other for all of the tapes. Since it was expected that the groups would differ in the time needed to make final decisions, the problem-solving period was defined as "the time between the exit of the experimenter and the group's agreement upon the last of the four choices required by both the typical and maximal laboratory situations" (Turner, 1978, p. 124). Talking times were unavailable for five subjects in the typical laboratory because of equipment failure. The correlation between dominance rating and talking time in the typical laboratory was .58; the corresponding correlation in the maximal laboratory was .56. These relationships were consistent for subjects with male and female confederates. Turner (1978) combined talking time and dominance rating to form composite measures of typical and maximal laboratory dominance. This was not done in the present study because of the different effect of sex of confederates on these measures, reported in the Results section.

Evaluation of Group Members Form. This measure consists of a series of 5-point scales on which each subject was asked to rate herself and the two other group members (confederates) on leadership ability and passiveness/dominance. The subject completed one of these forms after each laboratory condition, thus providing self-ratings of typical and maximal laboratory dominance behavior. For the typical laboratory, self-rating correlated .55 with confederates' dominance rating and .52 with talking time. Corresponding correlations for the maximal laboratory were .31 and .37.

Procedure

The women signed up for one of four available sessions to complete the PAQ, the Person Description Form (PDF), and the CPI Dominance scale.

After completing these tasks, the women signed up for individual laboratory sessions, held 1 to 4 weeks later.[1]

Ten senior psychology students (five men and five women) rotated through the experimenter and confederates roles in the typical and maximal laboratory situations.[2] Subjects were exposed to either a female experimenter and four female confederates or a male experimenter and four male confederates during the laboratory session.

The subject was greeted by the experimenter, who explained that the other students participating in the session were already present. The other students were two confederates. One confederate was waiting in the lounge while the other was in the laboratory in order to surreptitiously turn on a concealed tape recorder. The experimenter entered the laboratory with the subject and the confederate, introduced everyone by his or her first name and had them sit around a table. They were then read the following:

This research is concerned with problem solving. The kinds of problems that we are interested in are

---

[1] Two subjects went to the laboratory session within 3 days of the paper-and-pencil session.

[2] The authors gratefully acknowledge the laboratory assistance of Lisa Fowler, Bill Aven, Charlene Henslee, Steve Bates, Teresa Matthews, Mike Dickey, Nancy Feinberg, Adele Goldstein, Paul Burnham, and Ray Mechler.

often abstract and seem to have no clear-cut answer. However, we are interested in what people's conclusions concerning these problems are and how they arrived at them. I will read the problem to you and then give you 10 minutes to come up with your conclusions as a group. You may have seen problems similar to this before.

The school district in a small community is in financial difficulty. The school board must dismiss eight faculty members in addition to other cutbacks in order to have enough money to complete the school year. A decision has already been made that the eight faculty members would be specialists in various areas, not teachers of the basic subject areas. Which four of those listed below would you choose to remain?

1. Reading specialist
2. Counselor
3. Assistant football coach
4. Special education teacher in math
5. Assistant band director
6. Drama teacher
7. Diagnostician of learning disabilities
8. Teacher of accelerated English classes
9. Teacher of accelerated math classes
10. Baseball coach
11. Work–study supervisor
12. Art teacher

The experimenter then placed a description of the problem in the middle of the table along with a pencil and a group conclusion form, on which someone was to write down the four persons of the group's choice. The experimenter reminded the group of the 10-minute time limit, pointed out the clock on the table, and left the room.

The confederates' behavior was primarily composed of presenting both pro and con arguments as to which specialists should be dismissed. Although they interacted freely in the discussion, they were democratic in their approach and only moderately assertive. Within this general pattern, three other procedures were followed:

1. At initiation points (i.e., when the experimenter left and returned) the confederates gave the subject 3–4 seconds to begin the interaction, if she chose to do so.

2. Approximately halfway through the session, one of the confederates began to discourse on a tangential topic. This lasted for 1 minute unless it was terminated by the subject.

3. At some point during the 10 minutes, one of the confederates openly disagreed with one of the subject's statements to determine her reaction to this challenge.

After 10 minutes the experimenter returned, and if the group had not yet finished the task, they were given an additional minute. All of the groups finished within 12 minutes' total time. The experimenter talked briefly about the problem and collected the group conclusion form. The experimenter then told the group that it was necessary to get their individual reactions to the group process and asked the subject to come to an adjacent room, saying that he (she) would return to get the others' reactions after talking to the subject. The subject then completed an evaluation of group members form. During this time, the confederates independently rated the subject on dominance and facial attractiveness.

The subject was then read the following statement:

> For the second part of this research, you will be with another group of people solving a problem similar to the one you worked on here. In this group, I want you to try to be the leader. I want you to be as dominant and assertive as you can be. This is an ability test. We will tape-record this session so that we can see how well you do. Remember, this is a dominance ability test. You are to see that your group comes up with satisfactory answers to the problem.

The experimenter then stated that the other students (confederates) in this part of the experiment were ready to begin and led the subject to the second laboratory while stressing the ability nature of the assessment.

The maximal laboratory situation was similar to the typical situation. Again, one confederate was waiting in the lounge while the other was in the laboratory to turn on the tape recorder. After following the same introductory procedure as enacted in the typical situation, the experimenter read the following:

> This research is concerned with problem solving. The kinds of problems that we are interested in are often abstract and seem to have no clear-cut answer. However, we are interested in what people's conclusions concerning these problems are and how they arrived at them. I will read the problem to you and then give you 10 minutes to come up with your conclusions as a group. You may have seen problems similar to this before.
>
> Imagine that our country is under threat of imminent nuclear attack. A man approaches you and asks you to make an independent decision: There is a fallout shelter nearby that can accommodate 4 people but there are 12 people vying to get in. Which 4 do you choose to go into the shelter? Here is all the information we have about the 12 people:
>
> 1. A 40-year-old male violinist who is a suspected narcotics pusher.
> 2. A 34-year-old male architect who is thought to be a homosexual.
> 3. A 26-year-old lawyer.
> 4. The lawyer's 24-year-old wife who has just gotten out of a mental institution.
> 5. A 75-year-old priest.
> 6. A 34-year-old retired prostitute who was so successful that she's been living off her annuities for 5 years.
> 7. A 20-year old black militant.
> 8. A 23-year-old female graduate student who speaks publicly on the virtues of chastity.

9. A 28-year-old male who will only come into the shelter if he can bring his gun with him.
10. A 12-year-old girl who has a low IQ.
11. A 30-year-old MD who is an avowed bigot.
12. A high school student.

The experimenter followed the same procedure as before and left the room. The confederates' behavior during the problem-solving period was the same as that of the confederates in the typical laboratory.

After 10 minutes, the experimenter returned and the same procedure as outlined above was followed, with the subject filling out another evaluation of group members form. The subject was then debriefed and was asked not to discuss the experiment with anyone until the study was completed.

The school board and bomb shelter paradigms were counterbalanced for the typical and maximal laboratory conditions. Since preliminary analyses did not reveal a significant order effect or a significant difference between the two paradigms across laboratory conditions, it was assumed that the two paradigms were equivalent.

Results

As expected, mean scores on the CPI Dominance scale differed significantly by M and F group, $F(3, 108) = 6.41$, $p < .001$. A planned comparisons [3] revealed that, as predicted, the means of the undifferentiated and feminine women (27.3 and 26.9) were significantly lower than the means of the masculine and androgynous women (32.1 and 32.7), $F(1, 108) = 19.04$, $p < .01$. This difference is entirely due to the correlation of .49 ($p < .01$) between the PAQ M scale and the CPI Dominance scale; the F scale has no relationship to CPI dominance ($r = .00$).

Table 1 presents the means and standard deviations of the laboratory measures by M and F group and sex of confederates. Results were analyzed by the three-way analysis of variance (M and F Groups × Sex of Confederates × Laboratory Condition) presented in Table 2. The difference between typical and maximal measures was highly significant for dominance rating ($p < .001$) and talking time ($p < .001$), showing that the maximal instructions were highly effective in eliciting more dominance behavior in the second laboratory situation.

As anticipated, M and F group had a significant effect on dominance rating ($p < .002$) and talking time ($p < .01$). For dominance rating, planned comparisons showed a significant difference between the means of the undifferentiated and feminine women and those of the masculine and androgynous women for the typical laboratory, $F(1, 208) = 9.82$, $p < .01$, and for the maximal laboratory, $F(1, 208) = 7.75$, $p < .01$.

For talking time, results of planned comparisons were parallel: significant for the typical laboratory, $F(1, 198) = 9.86$, $p < .01$, and marginally significant for the maximal laboratory, $F(1, 198) = 3.47$, $p < .07$.

Closer examination of Table 1 showed that the means of the four M and F groups tended to follow a consistent pattern: masculine, androgynous, undifferentiated, and feminine, with the masculine women having the highest dominance ratings and the feminine women the lowest.

Sex Role Behavior

The effect of sex of confederates on women's dominance behavior is complex. The prediction that women would express less dominance with male confederates than with female confederates in the typical laboratory was supported for dominance rating, but contradictory results were obtained for talking time.

Although for dominance rating the main effect for sex of confederates was nonsignificant (Table 2), the interaction between sex of confederates and laboratory was significant ($p < .03$). Planned comparisons showed that for the typical laboratory, the means of women with male and female confederates were significantly different, $F(1, 208) = 4.38$, $p < .05$, whereas for the maximal laboratory, the means were practically identical, $F(1, 208) = .02$, *ns*. Thus, for dominance rating, women's behavior was not consistent in the two laboratory situations. The women were significantly less dominant with male confederates in the typical laboratory, as predicted, but equally dominant with male and female confederates in the maximal laboratory.

[3] Planned comparisons among cell means were performed according to procedures described by Keppel (1973; pp. 445–446).

Table 1
Means and Standard Deviations of the Laboratory Measures by M and F Group, Sex of Confederates, and Combined

M and F group	n[a]	Laboratory dominance rating: Typical *M*	Typical *SD*	Maximal *M*	Maximal *SD*	Laboratory talking time (%): Typical *M*	Typical *SD*	Maximal *M*	Maximal *SD*
Male confederates									
Undifferentiated	14 (14)	6.79	2.81	9.50	2.21	21.0	14.0	36.5	13.4
Feminine	14 (13)	6.14	2.77	9.43	1.34	21.9	10.7	33.9	12.7
Masculine	14 (13)	9.00	2.35	10.86	1.29	28.5	14.1	41.3	11.6
Androgynous	14 (14)	7.43	2.31	9.79	1.48	25.4	14.6	37.7	13.1
Combined	56 (54)	7.34	2.69	9.89	1.67	24.2	13.3	37.4	12.5
Female confederates									
Undifferentiated	14 (13)	7.93	2.95	9.93	2.09	16.4	7.8	29.4	10.5
Feminine	14 (14)	7.57	2.95	8.14	2.03	13.4	7.4	27.5	15.5
Masculine	14 (12)	9.29	1.82	10.71	2.02	27.2	11.0	36.4	10.0
Androgynous	14 (14)	8.29	3.38	10.57	2.53	21.3	11.4	31.7	13.0
Combined	56 (53)	8.27	2.81	9.84	2.34	19.4	10.5	31.2	12.4
Combined									
Undifferentiated	28 (27)	7.36	2.83	9.71	2.08	18.8	11.3	33.0	12.1
Feminine	28 (27)	6.86	2.85	8.79	1.78	17.5	9.7	30.7	14.0
Masculine	28 (25)	9.14	2.03	10.79	1.63	27.9	12.2	38.8	10.7
Androgynous	28 (28)	7.86	2.82	10.18	2.04	23.3	12.8	34.7	12.9
Total	112 (107)	7.80	2.80	9.87	2.04	21.8	12.3	34.3	12.9

Note. Means are derived from the summed scores of two raters. M = masculinity. F = femininity.
[a] *n*s in parentheses are for talking time in the typical laboratory.

Findings for talking time were the reverse of what was expected. Women talked *more* with male confederates than with female confederates ($p < .01$). Planned comparisons were marginally significant for the typical laboratory, $F(1, 198) = 3.81$, $p < .06$, and

Table 2
Analysis of Variance Summary for M and F Group × Sex of Confederates × Laboratory Condition for the Laboratory Measures

Source	*df*	Dominance rating: *MS*	*F*	Talking time: *MS*	*F*
A	3	45.28	5.47*	796.56	3.67*
B	1	10.72	1.30	1,541.44	7.09*
C	1	238.22	86.87**	8,125.32	100.57**
A × B	3	2.50	.30	30.63	.14
A × C	3	1.62	.59	50.69	.63
B × C	1	13.50	4.92*	30.64	.38
A × B × C	3	4.91	1.79	28.17	.35
Error (between)	104	8.27	—	217.29	—
Error (within)	104	2.74	—	80.79	—

Note. For talking time, $df = 99$. M = masculinity. F = femininity. A = M and F group. B = sex of confederates. C = laboratory condition.
* $p < .05$. ** $p < .001$.

significant for the maximal laboratory, $F(1, 198) = 8.19$, $p < .01$.

Self-Rating of Laboratory Behavior

Unlike findings for dominance rating and talking time, self-ratings of the M and F groups do not follow a consistent pattern. The main effects for M and F group and sex of confederates were not significant, $F(3, 104) = 1.49$, *ns*, and $F(1, 104) = 1.39$, *ns*, respectively. Overall, the women did not perceive themselves as less dominant with male than female confederates.

Facial Attractiveness

For the facial attractiveness measure, the main effect for sex of confederates was nonsignificant, $F(1, 103) = 1.63$, *ns*, indicating that the male and female confederates perceived the women in their respective groups as equally attractive. Differences among the M and F groups were unremarkable, and the effect of M and F group on facial attractiveness was nonsignificant, $F(3, 103) = .21$, *ns*. Facial attractiveness, however, correlated significantly with both typical and maximal talking time ($r = .32$; $r = .21$) and typical and maximal dominance ratings ($r = .30$; $r = .22$); sex of confederates did not affect these correlations.

Self-Generated Trait Descriptions

As expected, trait group significantly predicted women's self-reports on the CPI Dominance scale, $F(3, 108) = 10.96$, $p < .001$, and their behavior in the laboratory situations; dominance rating, $F(3, 104) = 5.67$, $p < .001$; talking time, $F(3, 99) = 5.86$, $p < .001$. For all of these measures, the means of the trait-irrelevant group fell between those of the dominant and reserved groups. The mixed group behaved similarly to the trait-irrelevant group in the typical laboratory and the reserved group in the maximal laboratory. Results for self-rating of laboratory behavior were consistent with those for dominance rating and talking time. Trait group had a significant effect on this measure, $F(3, 104) = 4.34$, $p < .007$. For both laboratory situations, women in the dominant group perceived themselves as the most dominant, followed by women in the trait-irrelevant, mixed, and reserved groups, respectively.

Trait group had a borderline significant effect on the facial attractiveness measure, $F(3, 107) = 4.36$, $p < .06$. The facial attractiveness of the trait-irrelevant and dominant groups were similar and both were significantly greater than that of the reserved group. Post hoc comparisons showed that the differences between these groups were significant: trait irrelevant versus reserved, $F(1, 107) = 6.41$, $p < .05$; dominant versus reserved, $F(1, 107) = 11.54$, $p < .01$.

Multiple Regression Analysis

To get an idea of the relative contributions of the independent variables to the laboratory dominance scores, five predictors were entered into a multiple regression in a hierarchical sequence, CPI Dominance, M and F group (coded as three dummy variables), trait group classification (also coded as three dummies), sex of confederates, and facial attractiveness. The outcomes to be predicted were typical and maximal laboratory dominance and talking times.

CPI Dominance was entered first to see if M and F group contributed unique variance to the laboratory criteria beyond that predicted by CPI Dominance and because the laboratory outcomes could reasonably be regarded as validity criteria for the CPI scale itself.

Results of the multiple regression analysis are shown in Table 3. It is apparent the CPI Dominance accounts for the greatest amount of variance in the laboratory outcomes, and in only one instance (maximal dominance rating) did the M and F classification significantly contribute to variance beyond that contributed by the CPI scale. Facial attractiveness and sex of confederates both made significant contributions to the outcomes in the typical laboratory, and sex of confederates also contributed significantly to talking time in the maximal laboratory.

Although the trait group classification pro-

Table 3
Proportion of Variance (R^2) *in Laboratory Dominance Behavior Accounted for by the Predictors*

Predictor	Dominance rating		Talking time	
	Typical	Maximal	Typical	Maximal
CPI Dominance scale	.193*	.103*	.200*	.162*
M and F group (PAQ)	.040	.075*	.013	.020
Trait group (PDF)	.037	.033	.049	.017
Sex of confederates	.033*	.000	.026*	.044*
Facial attractiveness	.042*	.011	.051*	.003
Total variance accounted for	.345	.222	.339	.246

Note. Significance levels take into account the degrees of freedom for the predictors. CPI = California Psychological Inventory. PAQ = Personal Attributes Questionnaire. PDF = Person Description Form.
*$p < .05$.

vided no additional information in the multiple regression, it should be noted that the correlation between CPI Dominance and the dominant trait group dummy was $r = .42$, ($p < .01$), with the reserved group, $r = -.39$ ($p < .10$), and with the trait-irrelevant groups, $r = .00$. This is important because it suggests a moderate amount of multicolinearity for two of the dummy variables, which reduces the potential predictiveness of trait group classification in this analysis.

In another hierarchical analysis, M and F classification was entered first and CPI Dominance entered second. In every instance, M and F group accounted for a significant proportion of the variance (ranging from 5% to 13%, average = 8.6%). However, the contribution of CPI Dominance was also always significant (range from 5% to 14%, average = 11.5%).

The results of these analyses indicate that the fourfold M and F classification did not regularly provide additional information beyond that given by CPI Dominance scores. This interpretation cannot be readily generalized, however, since the outcome criteria in this particular study could be regarded as exemplars of the CPI scale itself. In studies with other behavior outcomes, the pattern of results might be different.

Despite these findings from the multiple regression, useful information on women's dominance expression would not have come to light without the perspective of psychological masculinity and femininity, which conceptually guided this study. The consistent pattern observed for the four M and F groups in both laboratory situations for both dependent variables (masculine > androgynous > undifferentiated > feminine) could not be explained solely by the relationship between CPI Dominance and the PAQ M and F scales. For example, since the CPI scale is significantly related to the M scale ($r = .49$, $p < .01$) and unrelated to the F scale ($r = .00$), why did the masculine women consistently express more dominance than the androgynous women when the latter actually had slightly higher CPI Dominance scores (32.7 vs. 32.1) and M scores (24.1 vs. 23.6)? With male confederates, masculine women obtained significantly higher dominance ratings than the androgynous women in the maximal situation ($p < .01$) and marginally significantly higher ratings in the typical situation ($p < .10$). Apparently, psychological "type" had an effect on women's dominance behavior beyond that predicted from CPI Dominance and PAQ scale scores.

Discussion

Sex role demands were an inhibitory force in the typical laboratory, where women either consciously or unconsciously suppressed their dominance behavior with men, but not in the maximal laboratory, where women were equally dominant with men and women. The latter gave us the opportunity to observe how women would behave in a situation in which

dominance, a stereotypically "masculine" behavior, was made salient and socially desirable. Apparently, women have the capacity to express relatively more dominance than they generally do under typical or everyday circumstances calling for appropriate sex role behavior. The explicit instructions to be maximally dominant seemed to serve as a "green light" for the women to ignore established norms. The women were actually capable of behaving in a dominant fashion with men but typically inhibited this behavior. These results demonstrate the importance of distinguishing between what individuals do and what they can do. Despite reserved or submissive behavior under typical conditions, many individuals are capable of displaying dominant behavior.

Findings for laboratory talking time were the reverse of expectations. Women were significantly more talkative with male confederates than with female confederates in both laboratory conditions. In retrospect, this finding is hardly surprising. Women may have been more "turned on" by participating in an experiment with men. Although talking time showed substantial relationships with other dominance measures (e.g., CPI Dominance scale; dominance rating), it is by no means a "pure" measure of dominance. Verbal behavior undoubtedly contains components of sociability, and just as dominance behavior has a negative valence in terms of traditional sex role standards for women, sociable behavior, particularly with the opposite sex, has a positive valence. Thus, sex role expectations were actually conducive to verbal interaction with men in the present study.

We believe that mean dominance rating differences as a function of the rater's sex are valid and not due to rating artifacts. Male and female confederates were trained together in pilot work, and raters' sex was significant only under typical instructions. Also, masculine women rated by males in the typical laboratory obtained higher dominance ratings than did the other three M and F groups rated by female confederates. In addition, correlations between talking time and laboratory dominance ratings were virtually identical regardless of whether males or females did the dominance ratings. And finally, correlations of CPI Dominance with laboratory dominance did not differ significantly as a function of sex of confederate ($r = .30$ with female confederates; $r = .49$ with male confederates).

These findings have several important implications. First, traditional sex role prescriptions have an impact on women's behavior. They can serve as facilitators or inhibitors depending on situational context. Second, women's capabilities for opposite sex role related behaviors may be underestimated. Women tend to suppress certain dispositional inclinations when sex role demands are operating but can exhibit them when these demands are no longer salient. Finally, it is possible to remove inhibitory forces in the environment so that women can express dominance more freely.

On the other hand, consistent individual differences were observed in the enactment of sex roles. Psychological masculinity and femininity, as measured by the PAQ, had a significant impact on women's self-reports and behavior even when they were given explicit instructions to be as dominant as possible. These findings are consistent with those reported by Bem and her coworkers (Bem, 1975; Bem et al., 1976).

Additionally, these psychological variables seemed to influence the degree to which women responded to the situational sex role demands, that is, masculine women showed the least difference in behavior as a function of confederates' sex in the typical laboratory. It is noteworthy that feminine women showed little increase in dominance when moving to the maximal laboratory with female confederates but showed a large increase with male confederates. Whether this is due to the feminine women being more influenced by the sex of the experimenter giving the maximal instructions or to less motivation in these women to express their dominance capabilities in the context of other women is unclear (see Eagly, 1978). That feminine and undifferentiated women showed less dominance than masculine women in the maximal laboratory suggests, contrary to Bem's (1975) conclusions, that they may be somewhat lacking

in basic competence and not simply inhibiting stereotypically masculine behavior. To study the nature of masculinity and femininity, therefore, researchers need to distinguish conceptually between sex role behaviors and psychological dimensions of masculinity and femininity as advocated by Spence & Helmreich (1978), especially if they are attempting to predict sex role related behaviors.

Another finding consistent with Spence and Helmreich's (1978) formulations concerns facial attractiveness. Undifferentiated, feminine, masculine, and androgynous women were rated as *equally* attractive by the confederates. These results contradict popular notions that masculine women are physically unattractive. Interestingly, although facial attractiveness was independent of psychological masculinity and femininity, this variable was significantly related to dominance behavior in the laboratory. In the typical laboratory, facial attractiveness accounted for 14% of the explained variance.

Many of our commonly held notions about psychological differences between the sexes are based on observations of behavior in everyday situations eliciting sex role behavior. The notion that women are less dominant than men may reflect adherence to traditional sex role standards and expectations more than strong psychological differences. Women may be more similar to men in psychological makeup than is generally assumed. Perhaps research paradigms utilizing a response capability approach will shed new light on psychological differences between the sexes. Unfortunately, this investigation did not have a male sample for purposes of comparison.

Results for the self-generated trait descriptions were as anticipated and similar to those reported by Turner and Gilliland (in press). These descriptions are predictively valid measures and refer to relatively consistent dispositional and behavioral variables. Unlike traditional personality measures, which require people to respond to a specific set of stimuli chosen by the experimenter, self-generated descriptions enable people to report what *they* perceive as relevant to their personalities. Bem and Allen (1974) argue that a more idiographic approach to assessment would improve the predictive validity of personality tests. Further, they point out that "there is no inherent conflict between an idiographic approach to assessment and a nomothetic science of personality" (p. 511). Results of the present study and Turner and Gilliland's study show that an idiographic assessment technique can significantly predict dominance behavior.

The behavior of the dominant, reserved, and trait-irrelevant groups is of particular interest in light of Bem and Allen's (1974) formulations. Women who perceived the dominance dimension as relevant to their personalities (i.e., dominant and reserved groups) were extreme in their dominance expression in accordance with their self-descriptions, whereas women who perceived this dimension as unimportant (i.e., trait-irrelevant group) were moderate in their dominance expression. The mixed group was inconsistent in the pattern of responses. This trait, then, was not of equal importance to all of the women, and this had a significant impact on their behavior. The predictive validity of personality measures may well be improved by focusing on those individuals who perceive the trait under consideration as relevant to their psychological makeup.

Finally, one must come away impressed with the strength of the CPI Dominance scale as a potent predictor of dominance behavior in the laboratory. The multiple regression analysis showed that this scale explained an average of 75% of the total variance accounted for by all predictors. The lesson to be learned from this finding is that in studies of psychological masculinity and femininity employing traditional aspects of dominance behavior as a criterion, one needs to be concerned with the *unique* contributions that these concepts make to outcomes beyond that given by traditional psychological measures. Although the concepts of psychological masculinity and femininity can be regarded as theoretically more valuable than the traditional concept of dominance, recognition of this close alliance allows one to make contact with an already established literature of research findings whose origins lie in a different conceptual framework.

References

Bakan, D. *The duality of human existence.* Chicago: Rand McNally, 1966.

Bem, D. J., & Allen, A. On predicting some of the people some of the time: The search for cross-situational consistencies in behavior. *Psychological Review,* 1974, *81,* 506–520.

Bem, S. L. Sex role adaptability: One consequence of psychological androgyny. *Journal of Personality and Social Psychology,* 1975, *31,* 634–643.

Bem, S. L., Martyna, W., & Watson, C. Sex-typing and androgyny: Future explorations of the expressive domain. *Journal of Personality and Social Psychology,* 1976, *34,* 1016–1023.

Block, J. H. Conceptions of sex role: Some cross-cultural and longitudinal perspectives. *American Psychologist,* 1973, *28,* 512–526.

Constantinople, A. Masculinity–Femininity: An exception to a famous dictum? *Psychological Bulletin,* 1973, *80,* 389–407.

Eagly, A. H. Sex differences in influenceability. *Psychological Bulletin,* 1978, *85,* 86–116.

Fiske, D. W., & Butler, J. M. The experimental conditions for measuring individual differences. *Educational and Psychological Measurement,* 1963, *23,* 249–266.

Gough, H. *California Psychological Inventory Manual.* Palo Alto, Calif.: Consulting Psychologists Press, 1956.

Keppel, G. *Design and analysis: A researcher's handbook.* Englewood Cliffs, N.J.: Prentice-Hall, 1973.

Megargee, E. I. Influence of sex roles on the manifestation of leadership. *Journal of Applied Psychology,* 1969, *53,* 377–382.

Spence, J. T., & Helmreich, R. L. *The psychological dimensions of masculinity and femininity: Their correlates and antecedents.* Austin: University of Texas Press, 1978.

Spence, J. T., Helmreich, R. L., & Stapp, J. The Personal Attributes Questionnaire: A measure of sex role stereotypes and masculinity–femininity. JSAS *Catalog of Selected Documents in Psychology,* 1974, *4,* 127.

Turner, R. G. Consistency, self-consciousness, and the predictive validity of typical and maximal measures. *Journal of Research in Personality,* 1978, *12,* 117–132.

Turner, R. G., & Gilliland, L. The comparative relevance and predictive validity of subject generated trait descriptions. *Journal of Research in Personality,* in press.

Wallace, J. An abilities conception of personality: Some implications for personality measurement. *American Psychologist,* 1966, *21,* 132–138.

Willerman, L., Turner, R. G., & Peterson, M. A comparison of the predictive validity of typical and maximal personality measures. *Journal of Research in Personality,* 1976, *10,* 482–492.

Received December 11, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 11, 2071–2080

Effects of Person Salience Versus Role Salience on Reward Allocation in a Dyad

Elena M. Carles and Charles S. Carver
University of Miami

Previous research shows that resource allocation in a dyad sometimes follows the principle of equity (proportional reward) and sometimes that of parity (equal reward). Existing evidence does not clarify the conditions under which each of these rules is invoked, however. A number of theorists have suggested that salience of the other as a person should lead to parity-based allocation, whereas salience of the other as a functionary filling a role should lead to equity-based allocation. The present study tested these possibilities. Subjects were led to perceive their own inputs to group performance as being either substantially lower or substantially higher than a partner's inputs. The partner had been portrayed to the subject in terms that made salient the partner's personal characteristics, the partner's role assignment, or neither of these. Among females, subsequent reward allocation followed the predicted pattern in both high- and low-input conditions. Among males, contrary to expectation, person salience led to heightened feelings of competitiveness and to increased allocations to the self. Discussion centers on the theoretical implications of these findings.

When group members interact, each member gives something to the interaction and each takes something in return. These are the costs and the rewards of group interaction (Thibaut & Kelley, 1959). Although rules for apportioning a group's rewards are sometimes determined by some form of open bargaining (e.g., Morgan & Sawyer, 1967), members of a group often arrive at implicit allocation rules without direct feedback from each other.

When persons are required to allocate group resources between themselves and their partners in a dyad, the division can be accomplished in many different ways. One may act solely in self-interest, keeping all the resources for oneself; one may give all the resources to one's partner; or one may divide the resources, giving some percentage to oneself and some to one's partner. Research in this area, which is often termed "distributive justice," usually finds that the allocator chooses the latter, that is, some division between self and partner. The question, then, that is basic to that area of investigation is this: What characteristics of the group members, the context, or the ongoing interaction affect the allocator's decision as to how the resources will be divided?

Studies in this area typically formalize the interaction between group members as follows. The subject (future allocator) and the other member are given some cooperative task. The subject subsequently receives feedback as to the relative productivity (inputs) of the two members. The group then is supplied with some reward, usually monetary, and the subject is asked to allocate this reward (or in some cases, reallocate it after an initial distribution by the partner).

Allocation Rules

Equity. One rule that could be used to allocate the group's resources is the principle

This research was conceived and conducted as a doctoral dissertation by the first author, under the supervision of the second author.

Requests for reprints should be directed to Elena M. Carles, who is now at Miami Mental Health Center, 3750 South Dixie Highway, Miami, Florida 33133.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3711-2071$00.75

of equity; that is, the notion that one should be rewarded in proportion to one's inputs. There is considerable evidence that this rule is often used in resource allocation. For example, Adams (1963) showed in a series of experiments that people tend to minimize the psychological differences between their inputs and their outcomes, in relation to the other members of a dyad.

Because many factors (e.g., psychological, social, physical) influence people's perceptions of inputs and outcomes, however, equity theory has been much modified since it was originally proposed. For example, Walster, Berscheid, and Walster (1973) postulated that there could be any number of systems of equity within a given social structure and that powerful individuals within that structure will attempt to induce others to accept systems beneficial to the group in power. According to this reasoning, group members can be influenced to believe that any method of apportionment is "equitable." Leventhal (1976) has also noted limitations of equity theory and has pointed out a number of other allocation rules that could be used instead of the equity rule.

Equity and parity. In spite of the multiple possibilities, however, a review of the literature indicates that available resources are usually distributed in one of two ways. In some cases the equity rule operates, and resources are distributed to each member of the dyad according to merit or observed performance in the prior task. In other cases, the resources are distributed *equally* between the members. This is called the rule of parity.

It is generally accepted that people choose one of these two rules as being appropriate in a given situation. Furthermore, it is clear that which rule appears appropriate in one set of circumstances does not necessarily dictate which is seen as appropriate in other circumstances (cf. Walster & Walster, 1975). Studies involving unequal input[1] have yielded some support for the equity solution (Leventhal & Anderson, 1970; Leventhal & Lane, 1970; Leventhal & Michaels, 1969) and some for the parity solution (Morgan & Sawyer, 1967; Wiggins, 1966). Apparently there is some additional factor that must account for this discrepancy between findings.

Degree of interaction. Shapiro (1975), in reviewing these studies, found that when equity predominated, there had been minimal social interaction between the allocator and the partner. In contrast, where parity occurred, there was either an acquaintance between the two or the implication of future interaction. Shapiro attempted to explore the possibility that the critical influence on resource allocation was the expectation of future interaction. He found, in a subsequent experiment, that when such an expectation existed, subjects were more likely to allocate resources according to parity; when no expectation existed, equity prevailed.

Shapiro's method, however, leaves serious questions as to exactly what variable determined his subjects' choice of allocation rule. In the condition involving no expectation of future interaction, so-called partners ostensibly participated in the study at completely different times. On the other hand, in the future-interaction condition the partners not only anticipated future interaction, but they worked in adjoining rooms with the door left open and could hear each other respond to questions. This confounding of the expectation of future interaction with a number of other variables prevents the ruling out of other potential causes as having determined subjects' choices of allocation rule.

Salience as person. Other theory suggests the possibility that Shapiro's effect actually depended on a different variable: the degree to which the partner was salient as a person. For example, Lerner (1977; see also Lerner, Miller, & Holmes, 1976) has theorized that the perceived relationship between partners should affect which rule of allocation is pre-

[1] Several researchers (Leventhal, Allen, & Kemelgor, 1969; Leventhal & Bergman, 1969; Leventhal, Weiss, & Long, 1969) have reported that when inputs are equal, subjects react to an unequal division of resources by reallocation to restore equity (equal pay for equal work). However, in the case of equal input (e.g., equal work), the equity solution (proportional allocation) happens also to be the parity solution (equal allocation). Thus comparative tests require that there be unequal inputs.

ferred. Lerner suggested that when one's partner is seen as similar to oneself, parity prevails. On the other hand, when "people relate to one another on the basis of their *position or function* in the cooperative endeavor . . . the justice of equity would be applied" (Lerner, 1977, p. 47, emphasis added). Thus as others are seen in terms of their functions (e.g., "co-worker") rather than as individuals, the rule for allocation of group resources should change from parity to equity. Deutsch (1975) has similarly suggested that seeing people in terms of their usefulness (function) leads to equity, whereas salience of social relations causes parity to dominate.

Low-input situations. Subjects in what might be regarded as personalizing circumstances often do allocate resources according to parity (Leventhal, Michaels, & Sanford, 1972; Shapiro, 1975). But there are special circumstances in which equity-type allocation has been found regardless of an implied relationship (Shapiro, 1975). Specifically, this has occurred when the subject's input was *less* than the partner's input. For the subject to invoke a parity rule in this case would mean, in effect, taking some of the other's earning for himself. This is not something that one would do to another "person" without his permission; nor is it the way one would be inclined to treat a "partner." It thus seems that the parity–equity decision discussed above may be most relevant to circumstances in which one's own contribution to the task had outweighed the other's input.

Present Study

It was suggested above that personalization of one's partner in a dyad may bias the choice of allocation rule toward parity, though this possibility has not yet been tested explicitly. Moreover, a review of the literature suggests that equity has prevailed when subjects were implicitly led to view their partners in terms of their roles. It seems plausible that an even greater bias toward an equity solution may occur if the other person is explicitly made salient as a functionary in a role. The present study examined these possibilities more closely.

In virtually all previous research, the relationship between the partners has been, either explicitly or implicitly, competitive. The "other person" in the present research similarly occupied what was implicitly a competitor role.[2] The present subjects experienced outcomes into which they themselves ostensibly had greater or lesser input than their "partners." The partner was portrayed to subjects such that (a) the partner's identity as a person was made salient, (b) the partner's role in the task was made salient, or (c) neither of these was made salient. The subject then was asked to allocate a group reward. It was expected that role salience would lead to equity-based allocation and that person salience would lead to parity-based allocation, both relative to the condition in which neither person nor role was salient. It was also expected that these tendencies would be maximized when the subject's perceived input was greater than the partner's.

Method

Subjects and Procedures

Subjects were 120 undergraduates (60 males and 60 females) from the University of Miami. One other subject was deleted because questioning revealed that he had not understood the experimental manipulation. Subjects were tested one at a time. Each was told, however, that he or she was participating as part of a (same-sexed) pair. The partner was ostensibly in a different experimental cubicle.

The experimenter explained that the study was intended to test the effects of group concentration on telepathic ability. The subject and his or her partner would be trying to receive telepathic communications

---

[2] There seems to be only one *non*competitive context in which roles are salient: when the roles are defined strictly in terms of cooperative partnership. If resource allocation becomes an issue in such a relationship, presumably both of the influences under consideration here—personalization and role salience—would influence allocation in the direction of parity, because sharing is a component of this role. It would thus be difficult to separate the two influences from each other. In part for this reason, the present research focuses on a relationship whose role characteristics should exert an influence opposite to that expected for personalization. It is likely that this implicitly competitive relationship corresponds relatively well in this respect to most relationships in which resource allocation is an issue.

from a sender.[3] The experimenter emphasized that success depended on concentration exerted by both subjects, and not on their specific abilities. Subjects were told that they would get feedback on the accuracy of both group members later on. The experimenter also stressed that partners would *not* meet either during or after the experiment. Ostensibly this was being arranged so that the partners would not be distracted by the anticipation of interacting. In reality, this information was being provided explicitly to remove anticipation of interaction as a possible cause of the behavioral effect under investigation. Thus the test of our hypothesis was not contaminated by this additional variable (cf. Shapiro, 1975).

Salience manipulations. In the person-salient condition, after the above task description had been given, the subject was told that distraction during the task would be minimized if the subject thought of the other as a unique individual. The experimenter ostensibly had been able to gather some personality information on potential participants from certain class sections. Subjects were told that although such information had not been obtained from them themselves, the information was available for their partner.[4] The subject then was given a brief description of the hypothetical partner, including the following specific information:

> Your corecipient's name is Pete (Linda), and he (she) is 19 years old. S/He lives in south Florida, and it says here that s/he's a Methodist.
>
> The personality tests show that Pete is mildly extraverted and somewhat impulsive. On a forced-choice questionnaire, Pete endorsed the item 'I often leap before I look, and that gets me into trouble sometimes.' The report indicates that Pete rarely feels depressed or anxious. He said 'I really enjoy life, and I don't let the bad things that happen get me down.'
>
> Pete prefers an active to a passive life. He participates in many sports, and says that he would rather play tennis on a Sunday afternoon than watch it on TV. Pete is highly sociable and loves to talk to people.
>
> Since Pete uses so much energy doing sports and socializing, it is not surprising that he reported himself to be somewhat slack in getting his school-work done, and leaves everything till the last minute.

The emphasis of all the information was to "personalize" the partner as much as possible, but to do so in such a way that the partner would be fairly representative of students at the University of Miami.

In the role-salient condition, the subject was told that thinking about who his or her counterpart might be would be disruptive. If the subject must think about the other at all, he or she should think of the role that the other plays as a receiver of telepathic communications.

Control subjects were told that the study was to test the effects of a specific vitamin complex on concentration and success in ESP tasks. Subjects were told that they were in a control group that would receive no vitamin. Although subjects were told that they would be working with a partner who was in another room and who would be concentrating on the same task, the partner was not emphasized either as an individual or as a role. The intent behind these instructions was to give the subject an orientation to the task in which neither person nor role would be salient.

In all experimental conditions, subjects were told to do their best and were reminded that the other person would also be trying to do his or her best at the task.

Input manipulation. Each subject then was asked to work on the ESP task. On each trial, the subject attempted to guess (when a signal light went on) which ESP card an experimenter was holding, from among four possibilities. After the task was completed, the subject was given false feedback as to the number of correct guesses made by each individual of the alleged pair. The subject was told that he or she had performed either substantially better than the partner (6 correct for the subject vs. 2 correct for the partner) or substantially worse than the partner (2 correct for the subject vs. 6 correct for the partner).

Dependent measure. The subject was then told that a small grant had enabled the experimenter to be able to reward the group's effort. Each group ostensibly was being rewarded according to how well the group had performed. After checking the group's total of "correct" responses, the experimenter indicated that the subject's group would be awarded four dollars. The subjects were told that they had been chosen by lot to divide the earnings between their partner and themselves. The experimenter said that how to divide the money was entirely up to the subject. The partner implicitly did not know what the group's total monetary reward would be,

---

[3] This cover story was chosen because it allowed easy manipulation of perceived input and because it allowed the roles held by the subject and "co-subject" to be equivalent to each other in terms of nature of function performed, status, and so forth. Doubts as to the credibility of the cover story among subjects proved to be unfounded, as assessed by pilot testing. Indeed, many subjects expressed some disappointment upon learning that the study had not actually concerned telepathy.

[4] This statement was included to guard against the possibility that subjects' behavior might be influenced by being "known" by the other. The effects of the study thus depended solely on the subjects' perceptions of the cosubject, rather than concern over the cosubjects' knowledge of the subjects.

and the subject was told that the division would be completely anonymous, even to the experimenter. Thus the subject should have perceived complete freedom in allocating resources.

The subject was left alone in the cubicle to fill out a form indicating what amount should go to himself or herself and what amount to the partner. The subject sealed this form (coded only by a special number) in a envelope and placed it in a depository, from which a departmental secretary ostensibly would take it later to type checks to be mailed out. As soon as this secret division of earnings was completed, the subject was given a postexperimental questionnaire and was debriefed. No subject actually received money. All were debriefed immediately in that regard, and were provided with an opportunity to vent any negative feelings that the deception had caused. In general, subjects had little negative reaction, feeling that the small amount of money involved did not warrant being upset at its loss.

Results

In the interest of clarity, we will present the behavioral results of the study first, and then the results of the various manipulation checks.

Allocation of Resources

A 3 (Role-Salient vs. Person-Salient vs. Control) × 2 (High vs. Low Input) × 2 (Male vs. Female) analysis of variance of subjects' dollar allocations to themselves revealed three significant effects. First, there was a large effect for input level, $F(1, 108) = 34.23$, $p < .001$, indicating that subjects who were told that they had done better in the task than their cosubject (high input) allocated more money to themselves than those who were told they did more poorly than their cosubject (low input). There was also a significant main effect for sex, $F(1, 108) = 11.71$, $p < .001$, such that males allocated more earnings to themselves than did females. Of greatest interest was a highly significant interaction between relationship and sex of subject, $F(2, 108) = 9.38$, $p < .001$ (see Figure 1). This interaction indicated that subjects' allocations were indeed influenced by their relationships to their partners, but the nature of that influence differed dramatically between males and females. As is clear from Figure 1, each gender's behavior was consistent across the two levels of perceived input.

A Duncan multiple range test indicated that females (combined across input levels) allocated significantly more to themselves in the role-salient than in the person-salient condition ($p < .05$). The control mean did not differ reliably from either the person mean or the role mean but fell approximately

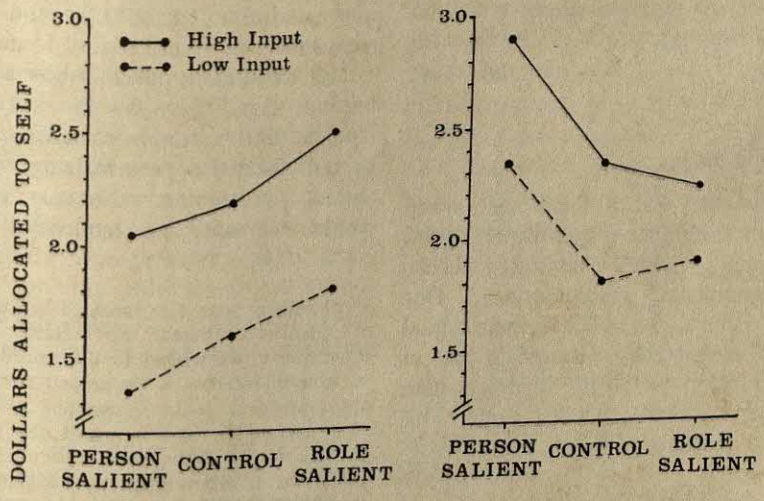


Figure 1. Mean allocation to self, in dollars, by (a) females and (b) males in each of three relationship conditions.

Table 1
Mean Ratings of Perceived Uniqueness of Cosubject, Knowledge of Cosubject, Degree to Which Cosubject Was Like Oneself, and Felt Competitiveness Toward Cosubject

Item	Relationship		
	Person	Control	Role
Uniqueness of cosubject			
Males	3.25	2.60	2.20
Females	3.80	2.75	2.00
Knowledge of cosubject			
Males	2.65	1.95	1.50
Females	3.30	1.50	1.20
Similarity to self			
Males	4.80	3.90	4.90
Females	5.35	5.05	3.80
Felt competitiveness			
Males	3.25	2.00	1.85
Females	1.60	1.90	2.75

Note. Group $n = 20$.

midway between them. These findings, among females, were precisely as had been predicted.

The findings for males, on the other hand, took a form opposite to the prediction. That is, when combined across input levels, males allocated to themselves more of the group's earnings in the person-salient condition than they did in either the role-salient condition or the control condition ($ps < .05$ by Duncan test). The latter two means did not differ reliably from each other.

Manipulation Checks

Degree of input. Subjects had been asked to respond to a number of postexperimental questions, some of which constituted checks on the experimental manipulations.[5] One item asked subjects to indicate their actual performance in the task compared to that of the ostensible cosubject, with three response options: higher, lower, and the same. This item served as a check on the input manipulation. All subjects retained for data analysis answered correctly for their input levels.[6]

Nature of relationship. Four questions indirectly assessed the way subjects perceived their cosubjects. These can be construed as checks on the relationship manipulation. Separate analyses of variance compared subjects' responses (on 7-point scales with suitable anchoring phrases) on each of these postexperimental questions.

The first item asked subjects to what degree they perceived the other as a unique individual. It was expected that this tendency would be increased by the person manipulation and decreased by the role manipulation. Consistent with this expectation, analysis of responses to this item revealed a significant main effect for relationship, $F(2, 108) = 6.72$, $p < .005$ (see Table 1). A Duncan multiple range test showed that when data were combined across both input and gender variables, subjects in the person condition perceived the other as a unique individual more consistently than did subjects in either the role or the control condition ($ps < .05$).

A second item asked subjects how much they knew about the other. It was anticipated that the personalizing manipulation should increase this tendency. As expected, an analysis of variance showed only a highly significant main effect for relationship, $F(2, 108) = 18.89$, $p < .001$ (see Table 1). A Duncan multiple range test showed that, overall, subjects in the person condition indicated significantly more knowledge about the cosubject than did subjects in either the control or the role condition ($ps < .05$); the latter two means did not differ reliably from each other.

The third item asked subjects to indicate to what degree they saw the other as another student like themselves. Again, it was expected that the personalizing manipulation would increase and the role manipulation would decrease this tendency. Analysis of

[5] Questions were also included to assess a number of potential confounds. For example, as was indicated above, we wished to distinguish explicitly the variable of interest in the present research from the anticipation of future interaction (Shapiro, 1975). It is noteworthy in this regard that subject groups proved not to differ from each other in their expectations of future interaction, expressing uniformly low likelihoods that that would occur. Nor did the groups differ regarding other variables assessed.

[6] As noted above, one subject was deleted because he chose the opposite response from the manipulated input value.

responses to this item, however, revealed a significant Sex × Relationship interaction, $F(2, 108) = 3.09$, $p < .05$ (see Table 1). A Duncan multiple range test revealed that when combined across input levels, females perceived the other as a student like themselves to a greater degree in the person condition than in the role condition ($p < .05$). For males, there was no reliable difference among the means.

The fourth item asked subjects to indicate to what degree they saw the other as a competitor. It was anticipated that the personalizing manipulation would tend to decrease perception of the other as a competitor, and that the role manipulation would tend to increase perception of the other as a competitor. Analysis of subjects' responses to this item revealed a highly significant Relationship × Sex interaction, $F(2, 108) = 7.30$, $p < .001$. This interaction (see Table 1) followed the same pattern as the interaction on the dollar-allocation variable. That is, subjects reported varying feelings of competitiveness according to nature of relationship, but the effects of the relationship variable differed between the genders.

As was anticipated, females perceived the other more as a competitor in the role condition than in the person condition. A Duncan multiple range test indicated that this difference was reliable between the person and role conditions when the input levels were combined, and for the high input condition taken separately ($ps < .05$). Male subjects, on the other hand, tended to perceive the other more as a competitor in the *person* condition than in either the control or role conditions. A Duncan multiple range test indicated that the person condition mean differed reliably from the control and the role condition means ($ps < .05$) when input levels were combined and also differed for the high-input condition taken separately.

As might be expected, based on the similarity between the results on this item and dollar-allocation variable, the degree to which subjects reported perceiving the other as a competitor was reliably related to the dollar allocations they made to themselves, overall $r(118) = .27$, $p < .01$.

Correlations. These postexperimental data provide mixed support for the contention that the relationship manipulations did what they were intended to do. Ratings of uniqueness of the other, knowledge about the other, felt competitiveness (among females), and seeing the other as a student like themselves (among females) were consistent with expectations. But there was a clear reversal of the expected relationship in felt competitiveness among males.

This gender difference was also revealed by correlational analyses. Rated knowledge of the other, which had been influenced by the personalizing information consistently across sex (Table 1), proved to be consistently correlated with perceived uniqueness of the other among the entire subject sample, $r(118) = .38$, $p < .001$; among males taken separately, $r(58) = .26$, $p < .03$; and among females taken separately, $r(58) = .47$, $p < .001$. Among females, neither knowledge of the other nor perceived uniqueness was related to feelings of competitiveness. But both these perceptions correlated positively with felt competition among men, $r(58) = .44$, $p < .001$, and $r(58) = .23$, $p < .04$, respectively.

This pattern of correlations suggests that one unanticipated perceptual–cognitive transformation was responsible for the unpredicted behavior of men in this study. Apparently personalizing information, felt knowledge of the other, and perceptions of the other's uniqueness were translated among men into feelings of competitiveness toward the other. Felt competitiveness, then, influenced even male subjects' allocations in a perfectly intelligible fashion.

Discussion

The primary focus of this experiment was upon the effect of the nature of a relationship on resource allocation. The nature of the relationship, as perceived by the allocating member of the dyad, clearly did influence the way in which that member divided the group earnings. The nature of that influence, however, depended on the gender of the subject.

Allocation Among Females

It was expected that role salience would lead to an equity-based allocation, and person salience would lead to a parity-based allocation, both relative to a control condition in which neither was made salient. Females behaved precisely in this fashion. In the high-input condition, female subjects who were encouraged to view their partners as persons allocated themselves almost exactly 50% of the group's earnings (a parity solution). Female subjects encouraged to view their partners in their competitive role allocated the earnings more in line with the relative perceived input of each member (an equity solution). Although prediction was more ambiguous among subjects led to perceive their own inputs as low, dollar allocations to self in those groups proved to follow the same pattern as that found among high-input subjects. In both cases, women were more willing to concede earnings when the other was perceived as a person than when the other was perceived as a functionary.

Allocation Among Males

Males, on the other hand, responded to the relationship manipulation in a very different way. Among males, allocation did not differ between role and control conditions, hovering only slightly above parity in both cases. Allocation among men was reliably influenced only by the personalizing information, and it was an influence that was opposite to what had been expected. In the person-salient condition, high-input males allocated themselves almost exactly 75% of the group's earnings (an equity solution). Although willing to concede some of their "equitable" earnings to the other in control and role conditions, they were completely *un*willing to do so in the person-salient condition. As was true among females, the same pattern of dollar allocations was retained in low-input groups, although the overall level of self-allocation was reduced.

Allocation and Perceptions of the Other

This pattern among males was not anticipated. But the study yielded additional information that allows some insight into the dynamics behind males' surprising behavior. In order to gain some perspective on the issue, let us reexamine the rationale behind the study. As was noted earlier, both Lerner (1977; Lerner et al., 1976) and Deutsch (1975) have theorized that if one person views the other in a personal way, as a unique individual, as being "like oneself," a parity-based division should occur. Women's ratings and allocations were fully consistent with this reasoning. Among males, in contrast, the effects of the personalizing manipulation fit with some of this characterization (enhanced knowledge of the other, perceived uniqueness), but not all of it. For some reason, the personalizing information induced a highly competitive orientation among men. It seems likely that this difference was the basis for the gender difference in allocation patterns.[7]

Why should personalizing lead to feelings of competitiveness among males? One possible explanation rests on the notion that males are generally more attuned to competitive aspects of the performance/allocation situation than are females, and take such competition more or less for granted. Thus, in same-sex dyads at least, men may be *less capable of extricating themselves from the competition role* in inherently competitive situations than are women. This would fit with the common characterization of males in our culture as "task oriented," whereas females are seen as more "social interaction" oriented (cf. Bakan, 1966; Maccoby & Jacklin, 1974; Parsons & Bales, 1955). When the other person was not particularly salient to them, as was the case in both the role-salient and control conditions, the situation may have been relatively comfortable for males, by virtue of its familiarity. That is, among males this sort of "ritualized" role-based competition would be normal and relatively nonthreatening. But providing specific informa-

[7] The possibility should be acknowledged that males' self-reports of competitive feelings occurred as a function of having made the allocation decision as they did, inasmuch as the self-report occurred after the allocation. Although this possibility cannot be discounted entirely, it appears to suggest very little basis for the pattern of allocations themselves.

tion about the other person made him more salient as an individual. Perhaps this enhanced awareness of the other's individuality made him appear more threatening *as a competitor*. In effect, enhanced salience of the other *person* might have been perceived by male subjects as enhanced salience of their *competitors*. This would account for the fact that males most strongly asserted their position of power in this condition.[8] This interpretation also suggests an important limitation on the theorizing of Lerner (1977) and Deutsch (1975). Specifically, personalizing the other may have the consequences that those theorists predicted only when the personalization *replaces* the competitive role orientation, rather than being *added* to that role orientation.

Implications

It is evident that the nature of the relationship between the members of a dyad has an important impact on reward allocation. At least two factors—the degree to which the other is personalized, and the perceived competitiveness of the relationship—appear to mediate this impact. Moreover, the clear and reliable sex difference in mediation of this effect in the present study makes it evident that future research concerning relationships and reward allocation must also take sex of subject into consideration.

Indeed, this gender difference appears to open fertile ground for future research opportunities. For example, if (as we argued above) the unexpected findings among males reflect a stereotypic "masculine" orientation toward competitive situations, one might expect this effect to be influenced by differences in subjects' sex role orientations (e.g., Bem, 1974, 1977). Another important possibility concerns the type of person who is portrayed to subjects. That is, all personalizing information is not alike, and we examined only one type of personalization in this experiment. Perhaps other portrayals might lead males to perceive their partner as deserving egalitarian rather than competitive treatment. Similarly, there may also be portrayals that would induce an especially competitive orientation among females.

These are only two among many possibilities. There can be little doubt, however, that men and women differ in important ways in how they perceive personal relationships and in what cues they use to define those relationships. These differences between the sexes remain an important area of scientific investigation.

---

[8] Inspection of the raw data shows that 5 subjects out of 120 allocated all the money to themselves. All were males in the person-salient condition, 4 in the high-input group, and 1 in the low-input group.

References

Adams, J. S. Toward an understanding of inequity. *Journal of Abnormal and Social Psychology*, 1963, *67*, 422–436.

Bakan, D. *The duality of human existence*. Chicago: Rand McNally, 1966.

Bem, S. L. The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology*, 1974, *42*, 155–162.

Bem, S. L. On the utility of alternative procedures for assessing psychological androgyny. *Journal of Consulting and Clinical Psychology*, 1977, *45*, 196–205.

Deutsch, M. Equity, equality, and need: What determines which value will be used as the basis of distributive justice? *Journal of Social Issues*, 1975, *31*, 137–149.

Lerner, M. J. The justice motive: Some hypotheses as to its origins and forms. *Journal of Personality*, 1977, *45*, 1–52.

Lerner, M. J., Miller, D. T., & Holmes, J. G. Deserving and the emergence of forms of justice. In L. Berkowitz & E. Walster (Eds.), *Advances in experimental social psychology* (Vol. 9). New York: Academic Press, 1976.

Leventhal, G. S. The distribution of rewards and resources in groups and organizations. In L. Berkowitz & E. Walter (Ed.), *Advances in experimental social psychology* (Vol. 9). New York: Academic Press, 1976.

Leventhal, G. S., Allen, J., & Kemelgor, B. Reducing inequity by reallocating rewards. *Psychonomic Sciences*, 1969, *14*, 295–296.

Leventhal, G. S., & Anderson, D. Self-interest and the maintenance of equity. *Journal of Personality and Social Psychology*, 1970, *15*, 57–62.

Leventhal, G. S., & Bergman, J. Self-depriving behavior as a response to an unprofitable inequity. *Journal of Experimental and Social Psychology*, 1969, *5*, 153–171.

Leventhal, G. S., & Lane, D. W. Sex, age, and equity behavior. *Journal of Personality and Social Psychology*, 1970, *15*, 312–316.

Leventhal, G. S., & Michaels, J. Extending the equity model: Perception of inputs and allocation of rewards as a function of duration and quantity of performance. *Journal of Personality and Social Psychology*, 1969, *12*, 303–309.

Leventhal, G. S., Michaels, J. W., & Sanford, C. Inequity and interpersonal conflict: Reward allocation and secrecy about reward as methods of preventing conflict. *Journal of Personality and Social Psychology*, 1972, *23*, 88–102.

Leventhal, G. S., Weiss, T., & Long, G. Equity, reciprocity, and reallocating the rewards in a dyad. *Journal of Personality and Social Psychology*, 1969, *13*, 300–305.

Maccoby, E. E., & Jacklin, C. N. *The psychology of sex differences*. Stanford, Calif.: Stanford University Press, 1974.

Morgan, W., & Sawyer, J. Bargaining expectations and the preference for equality over equity. *Journal of Personality and Social Psychology*, 1967, *6*, 139–149.

Parsons, T., & Bales, R. F. *Family, socialization, and interaction processes*. New York: Free Press, 1955.

Shapiro, E. G. Effect of expectations of future interaction on reward allocations in dyads: Equity or equality. *Journal of Personality and Social Psychology*, 1975, *31*, 873–880.

Thibaut, J. W., & Kelley, H. H. *The social psychology of groups*. New York: Wiley, 1959.

Walster, E., Berscheid, E., & Walster, G. W. New directions in equity research. *Journal of Personality and Social Psychology*, 1973, *25*, 151–176.

Walster, E., & Walster, G. W. Equity and social justice: An essay. *Journal of Social Issues*, 1975, *31*, 21–44.

Wiggins, J. Status differentiation, external consequences, and alternate reward distributions. *Sociometry*, 1966, *28*, 89–103.

Received January 9, 1979 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 11, 2081–2090

Allocation of Attention and the Type A Coronary-Prone Behavior Pattern

Karen A. Matthews
Department of Psychiatry
University of Pittsburgh

Bradford I. Brunson
Kansas State University

Three studies were conducted to assess the attentional style of individuals with the Type A coronary-prone behavior pattern. Experiment 1, which made use of a dual-task paradigm, revealed that Type A's focus their attention on central tasks; thus, they attend less to peripheral tasks than do Type B's. Experiments 2 and 3, which used a single task performed in the presence of a distracting stimulus, indicated that Type A's actively inhibit or suppress their attention to task-irrelevant peripheral events that might distract them from task performance. These findings validated anecdotal observations that Type A's appear hyperalert (focused in their attention) but neglect task-irrelevant cues. Previous research has demonstrated that Type A's fail to report fatigue as well as a variety of other physical symptoms of illness during task performance. To the extent that symptoms are analogous to peripheral events that distract from task performance, the data suggest that Type A's suppress their attention to symptoms. Implications of the attentional style of Type A's for the pathogenesis of coronary artery and heart disease are discussed.

A behavior pattern called *Type A* has become widely known for its negative effects on cardiovascular health. Both retrospective and prospective studies have demonstrated that Pattern A is associated with at least twice the occurrence of heart disease as an opposing Pattern B, which is defined as the absence of Pattern A (Friedman & Rosenman, 1974; Jenkins, Rosenman, & Zyzanski, 1974). This relationship remains when statistical controls are introduced to partial out the effects of traditional risk factors for coronary heart disease (CHD), such as serum cholesterol level and systolic blood pressure. Moreover, Pattern A is also associated with the degree of atherosclerosis in the coronary arteries (Blumenthal, Williams, Kong, Schanberg, & Thompson, 1978; Zyzanski, Jenkins, Ryan, Flessas, & Everist, 1976).

Behavioral Characteristics of the Type A

The Type A behavior pattern is defined as "an action–emotion complex that can be observed in any person who is aggressively involved in a chronic, incessant struggle to achieve more and more in less and less time, and if required to do so, against the opposing efforts of other things or persons" (Friedman & Rosenman, 1974, p. 67). Systematic investigations by Glass and his colleagues have verified the existence of a number of behavioral manifestations of this behavior pattern (cf. Glass, 1977). For example, Type A's signal the passage of 1 minute sooner than Type B's do, apparently sensing time passing rapidly (Burnam, Pennebaker, & Glass, 1975). When frustrated, Type A's become aggressive and hostile (Carver & Glass, 1978; Glass, Snyder, & Hollis, 1974). Type A's work at near their maximal rate even when there is no explicit time deadline (Burnam et al., 1975; Carver, Coleman, & Glass, 1976).

The authors thank Chuck Carver, Kent Houston, Michael Scheier, and Richard Jennings for their helpful comments on an earlier draft of this manuscript, and David Hufford and Kathy Menna, who performed ably as experimenters.

Requests for reprints should be sent to Karen A. Matthews, Department of Psychiatry, University of Pittsburgh, 3811 O'Hara Street, Pittsburgh, Pennsylvania 15261.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3711-2081$00.75

Simultaneously they report less fatigue and other physical symptoms than B's do (Carver et al., 1976; Weidner & Matthews, 1978).

One behavioral attribute of Type A's that has not been systematically investigated is their policy for allocating attention to environmental events. This is indeed surprising, considering that policies for allocation of attention play a critical role in such diverse behaviors as experiences of affect and bodily states (Gibbons, Carver, Scheier, & Hormuth, 1979; Scheier & Carver, 1977), human judgments (Taylor & Fiske, 1978), and attainment of performance goals (Carver & Scheier, Note 1). We suggested elsewhere that Type A's and Type B's might differ in how they allocate their attention to the environment and that this difference might explain why Type A's report less severe physical symptoms than Type B's do while performing a stressful task (Weidner & Matthews, 1978). Specifically, we reasoned that because of Type A persons' chronic efforts to achieve, they devote their full attention to tasks at hand. Consequently, they do not notice peripheral events that are not immediately relevant to task performance, such as physical symptoms or subtle cues of another's distress. In other words, Type A's focus their attention on environmental stimuli that will aid them in their efforts to excel. This notion is consistent with descriptions of Type A individuals as being extraordinarily alert but inattentive to extraneous cues (Bortner & Rosenman, 1967; Friedman & Rosenman, 1959; Rosenman & Friedman, 1961) and with the argument advanced by Glass (1977) that only *highly salient* tasks elicit controlling behavior from Type A's.

Do A's differ from B's in the policies by which they allocate attention to their environment? The purpose of the present research was to assess this possibility. Our major hypothesis was that Type A's focus their attention on central aspects of their environment; consequently they attend less to peripheral aspects than Type B's do. To test this notion, the first experiment used a dual-task paradigm (e.g., Bahrick, Fitts, & Rankin, 1952; Hockey, 1970; Kahneman, 1973) in which one task was defined as central or primary and one task was defined as peripheral or secondary through physical arrangements of the tasks and experimenter instructions. Assumptions of this paradigm are that humans have a limited capacity of attention and that performance on the two tasks varies with the amount of attention allocated to them, provided that the capacity of the individual is fully used. Focusing of attention is said to occur when performance on the secondary task is poor and performance on the central task is the same or good, compared to a referent.

In Experiment 1, Type A and B undergraduates worked for 6½ minutes on two tasks simultaneously. The primary task was the Stroop Color Naming Task (Stroop, 1935), and the secondary task was to depress a telegraph key upon the onset of a light, located at the right periphery of the field of vision. We expected that Type A's would allocate less attention to the secondary task than Type B's would; they would allocate as much and perhaps more attention to the primary task than Type B's would. Consequently, Type A's should perform poorer on the reaction time task and the same or better on the Stroop Color Naming Task than B's should.

Experiment 1

Method

Subjects. Subjects were 23 undergraduates enrolled in introductory psychology courses at Kansas State University who gained course credit by participating in research. Subjects were classified as Type A or B on the basis of their scores on the Jenkins Activity Survey (JAS) (Jenkins, Zyzanski, & Rosenman, 1971) revised for students by Glass (1977). In large samples of undergraduates the median of the JAS falls between 7 and 8 (Glass, 1977). Type A's were those who scored above the median of this sample of subjects (8.0 and above), whereas Type B's were those who scored below the median (7.0 and below). Two female Type A's were deleted because of equipment failure, one female Type B for failure to follow instructions. The final sample contained 10 Type A's and 10 Type B's; half of each group was male, half female. None were color blind, and all were native English speakers. Nonnative English speakers did not participate because they answer the JAS questions from a different cultural referent (see Cohen, 1978, for a discussion of this point).

Materials and apparatus. Directly in front of the subjects on a table were the materials for the primary task, the Stroop Color Naming Task

(Stroop, 1935) modified by Hartley and Adams (1974). The materials were a stack of sheets that contained 30 lines of six color names; each color name was printed in a color of ink other than its own color name. Each line contained one stimulus color name on the left and five possible choices of color names on the right. Subjects were to circle with a pen in their dominant hand the color name of the stimulus word's ink among the five choices.

The apparatus for the secondary task, a simple reaction time task, was located in the subject's right field of vision.[1] At approximately two o'clock in the visual field was a 12 inch × 12 inch (30.5 × 30.5 cm) black vertical panel on which was mounted a 1-inch (2.54 cm) square red light. At the base of the panel was a telegraph key secured to the horizontal platform adjoining the panel. During the 6½ minutes of the experiment, the light was activated 12 times; the intertrial interval averaged 20 sec, but varied from 10 to 30 sec. The duration of the light was 10 sec unless the subject terminated the light by depressing the telegraph key. The time to depress the key was measured in milliseconds by a Lafayette 54517 clock in conjunction with a digital printer.

Procedure. Subjects were tested individually. A female undergraduate experimenter greeted the subjects and explained that they would participate in two studies. After giving a brief overview of the two studies, she obtained their consent to participate in the research. In the first study, subjects were asked to complete the JAS because the experimenter wished to collect normative data on the life styles of college students. Following completion of the JAS, subjects filled out a second questionnaire while the experimenter scored the JAS. Then subjects were escorted to another testing room for the second study and were greeted by a male experimenter, a graduate student, who remained blind to their Type A–B categorization.

The second experimenter explained that they would be involved in an experiment designed "to test the effects of extraneous stimuli on task performance." He instructed the subjects on how to do the Stroop Color Naming Task and allowed them to practice until they solved several correctly in a row. Then he informed them that they would have about 6 minutes for this task and that they should work as fast and accurately as they could. A time deadline was imposed because Type A's and B's tend to perform at a similar maximal rate when there is such a time deadline (Burnam et al., 1975). Finally, almost as an afterthought, he said, "In addition to working on this Stroop task, I want you to press this telegraph key every time you notice this light come on." After several practice trials turning off the light, the instructions for the Stroop task were reviewed briefly, subjects began working, and the experimenter left the room. After 6 minutes had passed, the experimenter turned off the reaction time apparatus and returned to the testing room. (As a result, subjects worked on the Stroop task for approximately 6½ minutes.) Then he administered a postexperimental questionnaire and fully debriefed them.

Table 1
Performance by Type A's and Type B's on Secondary and Primary Tasks

Dependent measure	Type A	Type B
Secondary task		
M reaction time to each light[a]	4.14	1.97
Number of noticed lights[b]	10.3	11.8
M reaction time to each noticed light only[a]	3.18	1.83
Primary task		
Number of Stroop problems completed correctly[b]	165.0	106.7

[a] The *lower* the number, the better the performance on this measure.
[b] The *higher* the number, the better the performance on this measure.

Results and Discussion

Performance on the secondary and primary tasks and items on the postexperimental questionnaire were analyzed by 2 (Type A–B) × 2 (Sex of Subject) analyses of variance. Because there were no significant effects for sex of subject in any of the analyses, only analyses by type are reported. Table 1 presents the mean Type A and B scores for the following dependent measures: (a) mean reaction time to each of the 12 lights, (b) number of lights noticed (as indicated by depression of the telegraph key prior to automatic turning off of the light), (c) mean reaction time to each of the noticed lights (all measures from the secondary task), and (d) the number of problems completed on the Stroop Color Naming Task (the primary task).

Secondary task performance: Reaction time to lights. We expected that Type A's would attend less to the secondary task than would Type B's and consequently would react more slowly to the onset of the stimulus

[1] For the one left-hander in the study, the apparatus was switched to the subject's left field of vision, and he was required to press the reaction time key and circle the correct response on the Stroop task with his left hand.

light than Type B's would. As row 1 of Table 1 reveals, the mean reaction time to each of the 12 lights was significantly greater for Type A's than for Type B's, $F(1, 16) = 31.98$, $p < .001$. In addition, Type A's failed to notice as many lights as did Type B's (see Table 1, row 2), $F(1, 16) = 12.16$, $p < .003$. Because it was possible that the mean reaction time to the lights was greater for A's than for B's simply because A's did not notice as many lights as B's did, we also examined the mean reaction time to each light that had been noticed. As row 3 of Table 1 shows, Type A's were significantly slower than B's to react to these lights, $F(1, 16) = 17.88$, $p < .001$.[2]

Primary task performance: Stroop Color Naming Task. If Type A's focus their attention on the primary task, they should perform the same or better on the Stroop Color Naming Task than should B's. As row 4 of Table 1 reveals, Type A's did correctly complete significantly more Stroop problems than Type B's did, $F(1, 16) = 27.92$, $p < .001$.[3]

Postexperimental questionnaire. In response to items on a postexperimental questionnaire, Type A's believed that they performed better on the Stroop task than Type B's believed they performed, $F(1, 16) = 5.65$, $p < .03$. There were no other significant effects for sex or type on ratings of difficulty of the Stroop task, quality of performance on the reaction time task, disruptive effect of responding to the light on the Stroop task performance, or the importance of the research.

In summary, the results of Experiment 1 indicate that, relative to Type B's, A's perform more poorly on the secondary task and better on the primary task. It is worthwhile noting that the precise rewards for excelling on the primary and secondary task were left ambiguous in this study. We chose to do this to maximize the possibility that individual differences in allocation of attention would emerge. In fact, we suspect that should precise rewards be established, both A's and B's would be capable of allocating their attention to receive maximum reward. In other words, both types might behave similarly in explicit situations, for example, when performance on secondary tasks is highly rewarded. Nonetheless, the results support the notion that A's focus their attention on central environmental events. Thus, they allocate less attention to peripheral tasks than do B's in relatively unstructured settings.

Because A's attend less to peripheral tasks than B's do, it is reasonable to extrapolate that they also attend less to task-irrelevant peripheral events. Experiment 1 did not, however, address the issue of how active this inattentiveness of A's to peripheral events might be. That is, the presence or absence of a task-irrelevant peripheral event may simply be inconsequential for A's, who are attending to task performance. Alternatively, however, it may be that Type A's actively inhibit their attention to such peripheral events.

Experiment 2 was conducted to compare these two possibilities. In Experiment 2 we began by assuming that task-irrelevant sounds are peripheral events that can distract from task performance if attention is directed to them. Previous research has shown that the presence of extraneous noise improves Stroop color naming performance (Hartley & Adams, 1974). Apparently this occurs because subjects presented with a noise actively inhibit their attention to the noise; in doing so, they also inhibit their attention to other task-irrelevant cues—including the name of the color of the stimulus word (Houston, 1968, 1969; Houston & Jones, 1967). Thus, performance is better as subjects are less distracted by even the distractor that is inherent in the task. This fact allows us to ascertain whether Type A's actively inhibit their attention to the extraneous sounds or whether this peripheral event is simply inconsequential in the following way. If Type A's attend less to task-irrelevant peripheral events than

---

[2] On this measure only, there was a marginally significant interaction term, Sex × Type A–B, $F(1, 16) = 4.21$, $p = .06$. Although both male and female Type A's were slower than male and female Type B's, the difference between the types was larger for males than for females.

[3] Analyses of the number of *attempted* Stroop problems for Experiment 1 also revealed the same pattern of results: Type A–B, $F(1, 16) = 27.85$, $p < .001$. There were no significant effects in percentage of errors.

B's do simply because they are attending to the central task, then the Stroop performance of A's should not be affected by the presence or absence of a distractor. On the other hand, if Type A's actively inhibit their attention to distractors, the performance of A's should be facilitated by the presence of a distractor. To compare these possibilities, Type A and B undergraduates performed the Stroop Color Naming Task either in the presence or in the absence of a tape of distracting sounds. Each subject was given an explicit time deadline of 5 minutes to insure that both A's and B's worked at the same high rate (cf. Burnam et al., 1975).

Experiment 2

Method

Subjects. Subjects were 22 male and 21 female undergraduates enrolled in introductory psychology courses at Kansas State University who obtained course credit for their participation. Subjects were classified as Type A (9.0 and above on JAS) or Type B (6.0 and below) prior to experimental procedures.[4] One Type B woman was deleted from the analyses because of her failure to follow instructions. In the final sample, there were 11 Type A's and 10 Type B's in the distractor group and 10 Type A's and 11 Type B's in the no distractor group. None were color blind; all were native English speakers.

Procedure. Subjects were tested individually. A female undergraduate greeted the subjects, and as in Experiment 1, she explained that they would participate in two studies. After giving a brief overview of the studies, she obtained their consent to participate in the research. Then, during the first study, the JAS was administered. After subjects completed the JAS, they completed a second personality quiz while the experimenter scored the JAS and assigned subjects to either the distractor or the no distractor group. She then escorted subjects to an adjoining testing room where she introduced them to a male experimenter and informed the experimenter of the subjects' stimulus group assignment but did not inform him of the subjects' Type A–B score.

The second experimenter explained to subjects in the distractor group that the study concerned the effects of noise on task performance or to subjects in the no distractor group that the study concerned the effects of visual interference on task performance. Subjects were then given instructions identical to those in Experiment 1 as to how to perform the Stroop Color Naming Task and were told that they had a time deadline of 5 minutes and to work as accurately and quickly as possible. In addition, subjects in the distractor group were told that they would be exposed to a tape of sounds through headphones while they were working on the Stroop task. The tape contained continuous electronic music interspersed with such common sounds as a sports broadcast, brief verbal remarks, and a ticking clock. The average intensity of the sounds was 78 dB(A), 3000 Hz measured at the headphones with the aid of a coupler device.[5] The range was 72 to 85 dB(A). The no distractor group were also asked to wear headphones, but of course the tape was not played for them. Subjects then worked on the Stroop task until the experimenter stopped them after 5 minutes. Subjects completed a questionnaire concerning their perceptions of the experiment; questions concerning perceptions of the sounds were not administered to no distractor subjects. All subjects were questioned regarding their suspicions of experimental procedures and were then fully debriefed.

Results and Discussion

Except for ratings of the sounds, the data were analyzed by 2 (Type A–B) × 2 (No Distractor – Distractor Groups) × 2 (Sex of Subject) analyses of variance.

Performance on the Stroop Color Naming Task. We expected that if relative to Type B's, Type A's attend less to peripheral events simply because they attend more to central events, they should not perform better in the presence of a distractor. However, if Type A's actively inhibit their attention to the distractor, the presence of the distractor should facilitate their performance. In either case, there should be no differences in the types' performance in the no distractor condition because the types work at a similar rate with a time deadline (Burnam et al., 1975), and in general, performance should improve in the presence of sounds. As row 1 of Table 2 indicates, A's did perform better with a distractor than without. There were no differences in the performance of A's and B's without a distractor. The analysis of variance confirmed this pattern of results with a statistically significant interaction term, Type A–B × Distractor – No Distractor Group, $F(1, 34) =$

[4] There were 10 additional subjects who scored at 7.0 or 8.0 on the JAS who were a priori excluded from the study.

[5] This tape was designed to be similar to the Houston (e.g., 1968) tape of sounds in dB(A) level as well as in type of random sounds. The tape was distracting but not aversive.

Table 2
Number of Stroop Problems Completed Correctly by Type A's and Type B's

	Distractor				No distractor			
	Type A		Type B		Type A		Type B	
Experiment	No.	*n*	No.	*n*	No.	*n*	No.	*n*
2	113.0	11	81.7	10	93.5	10	97.5	11
3	125.1	9	104.2	10	101.5	8	95.2	9

5.90, $p < .02$, and a significant main effect for Type A–B, $F(1, 34) = 4.16$, $p < .05$.[6] Internal contrasts revealed that Type A's performed significantly better in the presence of a distractor than without, $t(19) = 2.14$, $p < .05$, and outperformed B's in the distractor condition, $t(19) = 3.44$, $p < .01$. Unexpectedly, however, Type B performance tended to worsen with distraction, $t(19) = 1.73$, $p < .10$. There were no other significant effects, including those for sex of subject.

Postexperimental questionnaire. Type A's rated the tape of sounds as more unpleasant on a 7-point scale than Type B's did (Ms = 3.3 vs. $M = 4.7$, respectively), $F(1, 17) = 4.80$, $p < .05$. Women rated the tapes as significantly louder and more stressful than men did, Fs$(1, 17) > 4.63$, ps $< .05$. There was a significant interaction term for ratings of task difficulty, Distractor – No Distractor Group × Sex of Subject, $F(1, 34) = 4.05$, $p = .05$. Although female subjects rated task difficulty similarly in the two groups, men in the distractor group thought the Stroop task was less difficult than those in the no distractor group did. There were no other significant differences across types, distractor conditions, or sex of subject for ratings of quality of performance, importance of research, and distractibility of the distractor.

In summary, the results of Experiment 2 are consistent with the notion that Type A's inhibit their attention to task-irrelevant peripheral events. Unexpectedly, Type B performance tended to deteriorate with distraction. Thus, we did not replicate previous research that indicates distracting sounds lead to improved performance overall on the Stroop test. Therefore, in order to examine the reliability of the Type B responses to distraction, we decided to replicate Experiment 2.

Experiment 3

Method

Subjects. Subjects were 18 male and 19 female undergraduates enrolled in introductory psychology courses at Kansas State University who obtained course credit for their participation. Subjects were classified as Type A (8.0 and above) or Type B (7.0 and below) prior to experimental procedures and assigned to cells so that an approximately equal number of Type A men and women and Type B men and women were in each group. One Type A man was deleted because of his failure to follow instructions. In the final sample, there were 9 Type A's and 10 Type B's in the distractor group and 8 Type A's and 9 Type B's in the no distractor group. None were color blind; all were native English speakers.

Procedure. The procedure was identical to that used in Experiment 2 with one exception. Subjects were administered the JAS during one test session. At that time, the female experimenter scheduled them for a second session one or two weeks later during which they were administered the Stroop Color Naming Task. As before, the experimenter for the second session was blind to the subjects' Type A–B score.

Results and Discussion

Performance on the Stroop Color Naming Task. As in Experiment 2, Type A's performed better with a distractor than without (Table 2, row 2). There were no differences in the performance of A's and B's without a distractor. In contrast to the results of Ex-

[6] Analyses of the number of *attempted* Stroop problems for Experiment 2 also revealed the same pattern of results: Type A–B × Distractor – No Distractor Group, $F(1, 34) = 5.60$, $p < .03$. There were no significant effects in percentage of errors.

periment 2, B's did not deteriorate in their performance in the presence of a distractor. The analyses of variance confirmed this pattern of results with a significant main effect for distractor – no distractor group, $F(1, 28) = 4.17$, $p = .05$, and no other significant results. Internal contrasts revealed that the main effect was largely due to Type A individuals' behavior. Type A's improved in their performance in the presence of a distractor, $t(15) = 2.09$, $p < .06$, whereas B's did not, $t < 1$. Type A's tended to perform better than B's did in the distractor group, $t(17) = 1.95$, $p = .07$. There were no other significant effects, including those due to sex.[7]

Postexperimental questionnaire. Women rated the tape of sounds as more distracting on a 7-point scale than did men, $F(1, 15) = 7.03$, $p < .02$. There were no other differences across types, distractor conditions, or sex of subject for ratings of task difficulty, quality of performance, and loudness, stressfulness, and pleasantness of the distractor.

The results of Experiment 3 replicate previous research that distracting sounds aid performance on the Stroop Color Naming Task. Taken together, the results of Experiments 2 and 3 reveal that this effect is primarily due to the Type A's. Apparently Type A's are able to inhibit their attention to the sounds. By doing so, they can actively ignore task-irrelevant cues and their performance improves.

General Discussion

The results of the three experiments are consistent with the hypothesized differences in the types' policies for allocating attention to the environment. It might be argued, however, that the results were due to an unfortunate selection of tasks such that Type A's simply are better on the Stroop Color Naming Task and poorer on simple reaction time tasks than Type B's are. This explanation is rendered much less plausible, though, by the similar Stroop performances of A's and B's without distraction in Experiments 2 and 3, and the lack of consistent differences in the simple reaction times of A's and B's in other studies (Abrahams & Birren, 1973; Krantz & Matheson, Note 2; Spieth, 1965). Therefore, performance differences of A's and B's on the tasks probably reflect true differences in allocation of Type A and Type B individuals' attention to the environment in these laboratory investigations.

In summary, then, the present findings provide the first systematic documentation of Type As' policy for allocating attention to the environment. They focus their attention on central environmental information; consequently, they attend less to peripheral aspects than Type B's do. Moreover, Type A's can actively inhibit their attention to peripheral events that might distract them from performing well on a task. These findings verify anecdotal observations that Type A's appear hyperalert but are less able than B's to report task-irrelevant (peripheral) cues (e.g., Bortner & Rosenman, 1967).

Arousal and Type A Individuals' Allocation of Attention

It is documented that arousal leads to a focusing of attention and to attention to the most task-relevant or salient cues (Easterbrook, 1959). Also, accumulating evidence shows that Type A's have greater increases in indices of sympathetic nervous system arousal than Type B's do while performing challenging tasks, although no differences between the types exist in resting levels (e.g., Dembroski, MacDougall, & Shields, 1977; Friedman, Byers, Diamant, & Rosenman, 1975; Manuck, Craft, & Gold, 1978). It could be argued that if A's in our studies were more aroused than B's, they attended more to the salient aspects of the Stroop task—in this case, to the color of the stimulus word's print—and ignored relatively irrelevant aspects, like the name of the Stroop stimulus word, the secondary task in Experiment 1, or the tape of sounds in Experiments 2 and 3. Thus, it is feasible that differences in arousal underlie the differences in the types' policies of attention allocation.

---

[7] Analyses of the number of *attempted* Stroop problems for Experiment 3 also revealed the same pattern of results: distractor – no distractor group, $F(1, 28) = 4.48$, $p < .05$. For error rates there was a significant interaction term, Distractor – No Distractor Group × Sex of Subject, $F(1, 28) = 5.11$, $p < .05$. Males had a larger percentage of errors in the no distractor than in the distractor group, whereas females showed the reverse pattern.

While this possibility is attractive in some respects and may be true for a variety of situations, it has a number of negative features as an explanation of the findings from Experiments 2 and 3. First, it is inconsistent with physiological evidence of cardiac deceleration during Stroop performance, which is associated with deliberate pacing used by subjects to reduce errors and confusion (see Kahneman, 1973). Second, it cannot explain why Type A's (presumably aroused) do not outperform Type B's in the no distractor group in the latter two experiments unless one assumes that the Stroop task alone is insufficiently challenging to create arousal. For the arousal hypothesis to fit the existing data, we must assume that Type A's were aroused while executing the Stroop task *only* in the presence of distracting sounds. But this assumption is inconsistent with the absence of arousal (measured by pulse rate) in subjects working on the Stroop task while listening to 78 dB(A) sounds (Houston & Jones, 1967). Presumably some of those subjects were Type A's. Finally, discernible arousal generated in ways other than noise (e.g., threat of shock) has a detrimental effect on performance of the digit symbol test, a subtest of the Wechsler Adult Intelligence Test (Carlson & Lazarus, 1953; Goldfarb, 1961), in contrast to distracting sounds, which have a facilitating effect (Houston, 1968). Type A's but not Type B's show significant improvement in their performance of the digit symbol test in the presence of distracting sounds (Matthews, Note 3), just as was true of the present Stroop performance. For these reasons, we do not favor an arousal interpretation of Type As' allocation of attention in Experiments 2 and 3, although we do expect that Type A individuals' sympathetic nervous system arousal affects their attentional style in other situations. In future research, therefore, we plan to examine the physiological correlates of Type As' attentional style in a variety of settings.

Implications for Previous Research

The differences in the types' allocation of attention to the environment provide a plausible interpretation of Type As' reactions to the salience of uncontrollable events (Glass, 1977). Type A's respond to highly salient losses of environmental control by making greater initial efforts than B's do to reassert control, followed by more extreme giving up relative to B's. When the uncontrollable event is moderately salient, Type A's actually make less effort than Type B's do to reassert control and do not give up, whereas Type B's do. Glass (1977) suggests that because of Type A persons' need to avoid losses of environmental control, they distort uncontrollability cues if they are not highly salient, or at least they do not effectively encode them. However, contrary to this suggestion is the absence of differences between A's and B's in perceptions of uncontrollability (Glass, 1977). On the other hand, the present data suggest that Type A's attend to the most salient aspects of their environment and inhibit their attention to any peripheral cue that might deter them from an excellent performance. This attentional interpretation of the salience findings can explain Type As' lack of responsiveness to uncontrollable events of low salience. It does not explain why Type B's do not give up following highly salient uncontrollable events, given that they do give up following moderately salient ones.

As stated earlier, our initial speculations about the attentional styles of Type A's were derived from our attempt to explain their failure to report symptoms during task performance. Indeed, the results of the present research are consistent with an attentional explanation of this phenomenon. That is, to the extent that symptoms are analogous to distractors from task performance, the data suggest that Type A's can inhibit their attention to symptoms. Tesser (e.g., Sadler & Tesser, 1973) has shown that the greater the amount of attention to an event or experience, the more extreme the evaluation of it. During task performance, then, Type A's may rate their symptoms as less extreme than B's do because they attend to them less than B's do.

We have stated elsewhere (Weidner & Matthews, 1978) that the term "suppression" was inappropriate to describe Type As' failure to report symptoms because it implied that we knew the mechanism underlying this phenomenon. At that time, we had no evidence why Type A's might fail to report symptoms. The present data, however, suggest that the

term *suppression* is appropriate to describe Type As' failure to report symptoms in the limited sense that Type A's *suppress their attention* to symptoms. It is important to emphasize that existing data do not suggest that A's have less sensory input of symptoms than do B's. That is, physiological indices of effort and sympathetic arousal are often greater or the same for A's relative to B's during task performance (Carver et al., 1976; Weidner & Matthews, 1978). It is also important to note that Type As' suppression of attention to symptoms is not the same as denying or distorting the truth in order to avoid admitting the disruptive effects of symptoms on task performance. Type A's report as much disruption of task performance from the distractor in Experiments 2 and 3 and from symptoms (Weidner & Matthews, 1978) as Type B's do. Rather, suppression appears to be a useful strategy for maintaining and even accelerating the efforts of Type A's, certainly a strategy that Type A's might seek out given their needs to achieve (see Carver et al., 1976).

Relevance to Heart Disease

We have observed elsewhere that failure of Type A's to report symptoms may contribute to Pattern A risk of coronary artery and heart disease in two ways (Weidner & Matthews, 1978). First, Type A's may delay seeking medical attention when experiencing early heart attack symptoms (Carver et al., 1976). With increasing delays, prodromal symptoms are more likely to result in a full-fledged heart attack and an ongoing infarction becomes more severe. In fact, delay is in itself a risk factor for heart disease (Insull, 1973). Second, Type A individuals' failure to report symptoms does not allow them to use symptoms elicited by stress as cues to alter their behavior to less stressful forms. To the degree that the attentional style of Type A's can account for their failure to report symptoms, this style may contribute to Pattern As' risk for heart disease. Furthermore, the attentional style of Type A's may also play a third, more critical role in the etiology of heart disease. Individuals who are described as hyperalert and who exhibit a high level of ability to focus their attention also show correlated increase in sympathetic nervous system activation (Williams, Note 4). If such activation were to occur repeatedly, it would subject the cardiovascular system to substantial physical loads on a daily basis, which might result in eventual heart disease (Williams, 1975). Indeed Williams (1975) has speculated that the perceptual and attentional aspects of Pattern A might provide the intervening link between Pattern A and coronary disease pathogenesis. The present research offers evidence that the above "coronary-prone" attentional style is consistent with Type A persons' style of attending to their environment. Thus, the attentional style of Type A's may serve them well in the sense of allowing them to focus their attention on work and to ignore annoying distractions, but it may also place them in triple jeopardy for coronary artery and heart disease in their later years.

Reference Notes

1. Carver, C. S., & Scheier, M. F. *The self-attention-induced feedback loop and human motivation: A control-systems analysis of social facilitation.* Unpublished manuscript, 1979. (Available from C. S. Carver, University of Miami).
2. Krantz, D. S., & Matheson, L. N. *Performance on simple and complex reaction times in asymptomatic Type A and B middle-aged men.* Unpublished manuscript, University of Southern California, July 1978.
3. Matthews, K. A. *A study of Type As' ability to ignore a distractor.* Unpublished data, Kansas State University, 1977.
4. Williams, R. B. *Peripheral sympathetic nerve activity and sensory intake behavior: A newly recognized association.* Paper presented at the meeting of the American Psychosomatic Society, Philadelphia, 1974.

References

Abrahams, J. P., & Birren, J. E. Reaction time as a function of age and behavioral predisposition to coronary heart disease. *Journal of Gerontology,* 1973, *28,* 471–478.

Bahrick, H. P., Fitts, P. M., & Rankin, R. E. Effects of incentives upon reactions to peripheral stimuli. *Journal of Experimental Psychology,* 1952, *44,* 400–406.

Blumenthal, J. A., Williams, R. B., Kong, Y., Schanberg, S. M., & Thompson, L. W. Type A behavior pattern and coronary atherosclerosis. *Circulation,* 1978, *58,* 634–639.

Bortner, R. W., & Rosenman, R. H. The measurement of Pattern A behavior. *Journal of Chronic Diseases,* 1967, *20,* 525–533.

Burnam, M. A., Pennebaker, J. W., & Glass, D. C. Time consciousness, achievement striving and the Type A coronary-prone behavior pattern. *Journal of Abnormal Psychology,* 1975, *84,* 76–79.

Carlson, V. R., & Lazarus, R. S. A repetition of Meyer Williams' study of intellectual control under stress and associated Rorschach factors. *Journal of Consulting Psychology,* 1953, *17,* 247–253.

Carver, C. S., Coleman, A. E., & Glass, D. C. The coronary-prone behavior pattern and the suppression of fatigue on a treadmill test. *Journal of Personality and Social Psychology,* 1976, *33,* 460–466.

Carver, C. S., & Glass, D. C. Coronary-prone behavior pattern and interpersonal aggression. *Journal of Personality and Social Psychology,* 1978, *36,* 361–366.

Cohen, J. B. The influence of culture on coronary-prone behavior pattern. In T. M. Dembroski, S. M. Weiss, J. L. Shields, S. G. Haynes, & M. Feinleib (Eds.), *Coronary-prone behavior.* New York: Springer, 1978.

Dembroski, T. M., MacDougall, J. M., & Shields, J. L. Physiologic reactions to social challenge in persons evidencing the Type A coronary-prone behavior pattern. *Journal of Human Stress,* 1977, *3,* 2–9.

Easterbrook, J. A. The effect of emotion on cue utilization and organization of behavior. *Psychological Review,* 1959, *66,* 183–201.

Friedman, M., Byers, S. O., Diamant, J., & Rosenman, R. H. Plasma catecholamine response of coronary-prone subjects (Type A) to a specific challenge. *Metabolism,* 1975, *24,* 205–210.

Friedman, M., & Rosenman, R. H. Association of a specific overt behavior pattern with blood and cardiovascular findings. *Journal of the American Medical Association,* 1959, *169,* 1286–1296.

Friedman, M., & Rosenman, R. H. *Type A behavior and your heart.* New York: Knopf, 1974.

Gibbons, F. X., Carver, C. S., Scheier, M. F., & Hormuth, S. Self-focused attention and the placebo effect: Fooling some of the people some of the time. *Journal of Experimental Social Psychology,* 1979, *15,* 263–274.

Glass, D. C. *Behavior patterns, stress, and coronary disease.* Hillsdale, N.J.: Erlbaum, 1977.

Glass, D. C., Snyder, M. L., & Hollis, J. F. Time urgency and the Type A coronary-prone behavior pattern. *Journal of Applied Social Psychology,* 1974, *4,* 125–140.

Goldfarb, A. Performance under stress in relation to intellectual control and self-acceptance. *Journal of Consulting Psychology,* 1961, *25,* 7–12.

Hartley, L. R., & Adams, R. G. Effect of noise on the Stroop test. *Journal of Experimental Psychology,* 1974, *102,* 62–66.

Hockey, G. R. J. Effect of loud noise on attentional selectivity. *Journal of Human Experimental Psychology,* 1970, *22,* 28–36.

Houston, B. K. Inhibition and the facilitating effect of noise on interference tasks. *Perceptual and Motor Skills,* 1968, *27,* 947–950.

Houston, B. K. Noise, task difficulty and Stroop color-word performance. *Journal of Experimental Psychology,* 1969, *28,* 403–404.

Houston, B. K., & Jones, T. M. Distraction and Stroop color-word performance. *Journal of Experimental Psychology,* 1967, *74,* 54–56.

Insull, W. (Ed.). *Coronary risk handbook.* New York: American Heart Association, 1973.

Jenkins, C. D., Rosenman, R. H., & Zyzanski, S. J. Prediction of clinical coronary heart disease by a test for the coronary-prone behavior pattern. *New England Journal of Medicine,* 1974, *290,* 1271–1275.

Jenkins, C. D., Zyzanski, S. J., & Rosenman, R. H. Progress toward validation of a computer scored test for the Type A coronary-prone behavior pattern. *Psychosomatic Medicine,* 1971, *33,* 193–202.

Kahneman, D. *Attention and effort.* Englewood Cliffs, N.J.: Prentice-Hall, 1973.

Manuck, S., Craft, S., & Gold, K. Coronary-prone behavior pattern and cardiovascular response. *Psychophysiology,* 1978, *15,* 403–411.

Rosenman, R. H., & Friedman, M. Association of specific behavior pattern in women with blood and cardiovascular findings. *Journal of the American Medical Association,* 1961, *24,* 1173–1184.

Sadler, O., & Tesser, A. Some effects of salience and time upon interpersonal hostility and attraction during social isolation. *Sociometry,* 1973, *36,* 99–112.

Scheier, M., & Carver, C. S. Self-focused attention and the experience of emotion: Attraction, repulsion, elation, and depression. *Journal of Personality and Social Psychology,* 1977, *35,* 625–636.

Spieth, W. Slowness of task performance and cardiovascular diseases. In A. T. Welford & J. E. Birren (Eds.), *Behavior, aging, and the nervous system.* Springfield, Ill.: Charles C Thomas, 1965.

Stroop, J. R. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology,* 1935, *18,* 643–662.

Taylor, S. E., & Fiske, S. T. Salience, attention, and attribution: Top of the head phenomena. In L. Berkowitz (Ed.), *Advances in experimental social psychology.* New York: Academic Press, 1978.

Weidner, G., & Matthews, K. A. Reported physical symptoms elicited by unpredictable events and the Type A coronary-prone behavior pattern. *Journal of Personality and Social Psychology,* 1978, *36,* 1213–1220.

Williams, R. B. Physiological mechanisms underlying the association between psychosocial factors and coronary disease. In W. D. Gentry & R. B. Williams, Jr. (Eds.), *Psychological aspects of myocardial infarction and coronary care.* St. Louis, Mo.: Mosby, 1975.

Zyzanski, S. J., Jenkins, C. D., Ryan, T. J., Flessas, A., & Everist, M. Psychological correlates of coronary angiographic findings. *Archives of Internal Medicine,* 1976, *136,* 1234–1237.

Received February 12, 1979

Journal of Personality and Social Psychology
1979, Vol. 37, No. 11, 2091-2097

The Power Motive as an Influence on Use of Power

Eugene M. Fodor
School of Management
Clarkson College of Technology

Dana L. Farrow
Division of Management
Florida International University

An industrial simulation experiment was conducted with students earning a master's degree in business administration to determine the influence of the power motive on their use of power. Need for power was assessed by means of the Thematic Apperception Test measure devised by Winter. Students scoring among the top and bottom third of those taking the test then proceeded to act as "supervisor" in the industrial simulation by directing the labors of a work crew in the next room. One member of the crew, Man C, behaved as an ingratiator. Supervisors high and low on *n* Power evaluated Man C about the same when he was neutral and noningratiating in his manner, but when Man C was an ingratiator, supervisors high on the power motive evaluated his performance more favorably than did low-scoring supervisors. High *n* Power supervisors also perceived themselves as exerting greater influence on the work group. These findings are seen as extending the heuristic value of the *n* Power construct by demonstrating its relationships to use of power in an industrylike situation.

Organizational supervisors typically have at their disposal a variety of powers over subordinates. Depending on the nature of the organization, these powers may include the power to grant raises, recommend promotion, recommend transfer to a more fulfilling job assignment, and even discharge subordinates for inadequate performance. It is surely axiomatic that supervisory use of power can substantially influence employee morale and need fulfillment. One would expect that a supervisor would promote need fulfillment among subordinates if he allocated benefits in a manner that was closely contingent on their quality of performance (Porter & Lawler, 1968). Conversely, need fulfillment would probably diminish to the extent that the supervisor distributed rewards on the basis of arbitrary criteria that were extraneous to recognized organizational objectives.

The purpose of the present investigation was to examine the relation between the power motive and use of power by persons acting as supervisor in an industrial simulation experiment. Need for power was assessed by means of the Thematic Apperception Test (TAT) measure devised by David Winter (1973). The experimental design represented a variation of the paradigm developed in studies reported by Goodstadt and Kipnis (1970) and Kipnis and Vanderveer (1971).

Studies that examined only situational influences on use of power have yielded inconsistent findings. Kipnis and Vanderveer (1971) found that persons acting as supervisor in an industrial simulation gave more pay increases and higher performance evaluations to an ingratiator than to a worker of similar performance who was not an ingratiator. A later experiment (Fodor, 1974) was unable to replicate these findings. Some studies have suggested that the presence of a problem worker increases the likelihood that a supervisor will regard with favor a compliant member of the work crew (Fodor, 1974; Goodstadt & Kipnis, 1970; Kipnis & Vanderveer, 1971). Another investigation obtained opposite findings (Fodor, 1976).

Grateful acknowledgment is expressed to Robert Conforti, Clifton King, and Robert Zogg, who served as experimenters.

Requests for reprints should be sent to Eugene M. Fodor, Area of Psychology and Organizational Behavior, School of Management, Clarkson College of Technology, Potsdam, New York 13676.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3711-2091$00.75

An influence unexplored in these studies was the variable of individual differences in motivation or in some dimension of personality. A likely candidate for consideration emerged with Winter's publication of *The Power Motive* (1973). By need for power, Winter means a need to influence, persuade, or control others and to gain recognition or acclaim through these forms of behavior. He developed a measure of need for power that is based on the TAT much in the same manner as McClelland's concept of achievement motivation (McClelland, 1958; McClelland & Winter, 1969). The Winter scoring system builds upon TAT criteria of need for power previously put forth by Veroff (1957) and Uleman (1966) and adds numerous refinements of its own. One method Winter used for selecting scoring categories was to administer the TAT both before and after exposure to an event that was replete with power imagery (John Kennedy's inaugural address) and then to note what categories of TAT imagery increased in frequency of usage. Winter completed a number of investigations in an attempt to establish the validity of his scoring system. Students who were or had been officers in student organizations scored higher on need for power than those who were not. Resident advisers obtained high scores. These students, it should be recognized, occupy positions of considerable influence and respect in the university community. They serve as academic and personal counselors for students, give advice, and act as the liaison between students and the administration. Athletes who engaged in competitive sports likewise scored high.

A major problem associated with experimentally derived TAT measures of motives has been low test–retest reliability (Atkinson, 1958). Winter and Stewart (1977) have presented evidence, however, that low temporal stability may derive from an implicit demand for novel stories during the second test administration. The problem substantially diminished when they instructed subjects at the retest either (a) not to worry about whether their stories were similar or different or (b) to try to write the same stories. Subjects instructed to write different stories produced reliability only at chance levels.

Winter's conceptualization implies that in their quest for recognition and acclaim, persons high in power motivation would probably respond favorably to persons supplying them with unswerving support and loyalty in the form of ingratiation. The principal hypothesis of the present experiment was that persons scoring high on need for power would dispense greater benefits on an ingratiating subordinate than would individuals low in power motivation. Persons scoring high were also expected to differentiate more broadly between two compliant but noningratiating workers of comparable performance in the pay increases and performance evaluations that they awarded. The assumption guiding this prediction was that those who are strongly dominated by power concerns take satisfaction in the exercise of power for its own sake and are apt to allocate benefits to subordinates in an arbitrary and whimsical fashion. The two compliant but noningratiating workers were equal in performance, appearing to deserve equal pay increases and performance appraisals. The expectation, however, was that supervisors high on the power motive would allow personal preference based on a subordinate's voice quality or some casual remark to take precedence over more objective performance criteria, causing them to make sharp distinctions between these two seemingly equal subordinates. A further prediction was made that high scoring persons would "read" members of the work crew as having been highly susceptible to their supervisory influence. They would report that they had exerted greater influence on the work group than would low-scoring supervisors.

Method

Subjects

Male students enrolled in the master's in business administration (MBA) program at Clarkson College served as experimental subjects. MBA students are preparing for careers as business managers and as a consequence of exposure to organizational behavior courses, they have begun to reflect on what style of leadership they should employ in their ultimate career role. The requirement that they act as supervisor in an industrial simulation it was therefore presumed, should for them have greater "experimental realism" (Aronson & Carlsmith, 1969) than for the typical undergraduate

Students were contacted by phone and were offered $2.50 to complete the TAT. They were told that they would be paid a like sum should they later be asked to participate in the industrial simulation. One hundred and twenty-two MBA students completed the TAT, and those scoring in the top and bottom third of the group on need for power were invited to return for the industrial simulation. This process was continued until a total of 80 students had successfully performed in the experiment. Subjects were assigned randomly to the two conditions.

The TAT Measure of Need for Power

There is no standard set of pictures for use in measuring power motivation, so eight pictures were selected that connoted some degree of power imagery. One was a scene from a football game, another was the photograph of a young man speaking to a crowd through a bullhorn, and a third depicted a well-dressed man passing a less well-dressed younger man standing against a wooden fence. Three of the pictures were taken from the set compiled by McClelland, Davis, Kalin, and Wanner (1972).

The Winter scoring manual sets forth a procedure that enables the researcher to attain high interrater reliability in the scoring of TAT stories for power imagery. The researcher first attempts to score four sets of 30 stories, checking the accuracy of his or her scoring against the standard after completion of each set. Three final sets are then scored for the purpose of obtaining an estimate of scoring proficiency. Winter describes two indices of interrater reliability. One is "category agreement," which measures level of agreement with the standard for presence or absence of power imagery, and the second is a rho coefficient that assesses correlation of ranks between one's own scoring and the standard. Winter recommends that the investigator regard a coefficient of .85 on both reliability measures as a prerequisite to research application of the Winter scoring system. The first author succeeded in meeting these criteria by obtaining a category agreement of .94 and a rho coefficient of .88.

To obtain an estimate of the stability of Winter's measure, two different sets of four pictures were presented to a class of 20 undergraduate students on each of two occasions separated by a span of 2 weeks. Scores for each story range from 0 to 11, and a total score can be computed by simply averaging scores across stories; so an overall score can be obtained on the basis of only four of the eight pictures. The Spearman–Brown formula yielded a stability coefficient of .46 for this split-half administration. Seventy-five male MBA students took the complete test at one sitting, and a division of scores placed them into a top, medium, and bottom third. Viewed against these points of division, only 1 of the 20 undergraduates changed position by more than an adjacent category from one administration to the next. Ten students remained within their third of the distribution across administrations. The test would appear to have research value if interest were confined to the behavior of persons scoring at the top and bottom third of the distribution, that is, those distinctly high or low on need for power.

Procedure

The experimenter informed subjects upon their arrival at the laboratory that the purpose of the industrial simulation was to study their managerial ability. They would, he told them, assume the role of supervisor and direct the work activities of three college men in the next room (who, unknown to the supervisor, would actually be nonexistent). The supervisor was further given to believe that individual workers had been assigned the task of assembling Tinker Toy models from pictured diagrams. These workers were to assemble as many models per trial as they were able, and which model they were to put together would change on each of five successive 5-minute trials. Each trial was followed by a 5-minute rest period. Successive trials featured models that were progressively more difficult to construct: a pair of eyeglasses, a tire swing, a teeter-totter, a hand winch for cranking up rope, and a windmill. The supervisor was expected to use every means at his disposal to foster high productivity in his men. These methods, the experimenter explained to him, might include words of encouragement, reprimands, promises of pay increase, and general advice on construction. He would speak to his workers by means of a one-way communication device comprised of a microphone and light signal for contacting individual workers. Communications allegedly coming from the various workers consisted of preprogrammed, play-acted tape recordings. The explanation presented to supervisors for not being permitted to see the workers was that the impressions they would acquire by seeing individual workers might bias them in their later evaluations. At the end of a trial the supervisor heard voiced comments from individual men, and then the experimenter brought in a count of each worker's output. Before and during every trial the supervisor had in front of him a fully assembled illustration of the model on which the men were currently working and also a fabricated set of norms (based on actual pretesting). These norms were expressed as performance percentile scores for college students. The student supervisor understood that workers had been promised a base pay of $2 for their participation but that he as supervisor could administer a 25¢ pay increase or decrease at the end of a given trial.

There were two conditions in the experiment, one with Man C acting as an ingratiator and the other with Man C assuming a neutral, noningratiating role. All three workers maintained roughly an average performance throughout both conditions. Number of words spoken per man was also equalized across workers and experimental treatments.

Table 1
Mean Performance Evaluations, Pay Increases, and Liking Ratings Given the Ingratiating Worker

Level of *n* Power in supervisor	Performance		Pay		Liking	
	Ingratiator	Neutral worker	Ingratiator	Neutral worker	Ingratiator	Neutral worker
High						
M	29.3_a	22.1_b	50.0	35.0	6.8	5.7
SD	4.2	4.8	19.9	27.4	1.7	1.8
Low						
M	24.2_b	23.0_b	37.5	31.3	6.2	5.9
SD	4.9	3.9	26.3	24.2	1.9	2.0

Note. Means with differing subscripts differ significantly from each other, $p < .01$.

Twenty of the student supervisors assigned to the ingratiator condition had scored in the top third of the need for power distribution, and 20 had scored in the bottom third. Of people assigned to the noningratiator condition, 20 had likewise obtained high scores on the power motive, and 20 were low. The ingratiator, Man C, established his style of relating to the supervisor at the end of Trial 2 by saying, "You know, I really like your approach. I can see you're gonna be a good supervisor." Trial 4 produced the remark, "Personally, I think you're doin' a great job." Man C completed Trial 5 on an equally complimentary note by commenting, "You've done pretty good." All of these utterances were illustrative of the form of ingratiation that Jones (1964) has designated as "complimentary other-enhancement."

Other communications by Man C and the remaining workers were neutral both in content and in emotional tone. At the conclusion of Trial 2, for example, Man A blandly proclaimed, "Not much to say this time, I guess." Man B said, "Well, so much for this job. What's next?"

When the supervisor had finished directing the work activities of his crew, the experimenter led him to another room and requested that he evaluate the performance of each worker on four 9-point scales: (a) worker's ability, (b) worker's overall worth to the company, (c) willingness to rehire worker for a second experiment, and (d) recommendation that the worker be promoted to supervisor in a future experiment. For purposes of analysis, these four scale scores were summated to obtain a single overall score. The student supervisor also completed ratings on how much he would like the worker as a personal friend (liking score) and how much he thought he had influenced the performance of each (influence score).

The experimenter then closely interrogated the supervisor to determine whether he believed there were no workers in the next room. Six supervisors so stated and were dropped from further data analysis. Eighty subjects remained, 20 in each of four groups as previously indicated. Another function of the postexperimental interview was to explain to subjects the details and underlying rationale of the study and then to swear them to secrecy.

Results

The first and major hypothesis of the study predicted that student supervisors scoring high on the power motive would bestow disproportionately high rewards upon the ingratiator. The relevant data appear in Table 1. Supervisory reward could take three different forms—performance ratings, pay increase, and liking. By all three forms of reward, persons scoring high on the power motive were found to confer higher rewards upon the ingratiator than were those scoring low. Moreover, persons high on need for power favored the ingratiator more than their counterparts in the control condition benefited Man C who was not an ingratiator. Since reward could take any of these three forms, the decision was made to perform a multivariate analysis of variance initially and then, as recommended by Hummel and Sligo (1971), to follow with univariate analyses of variance of the three variables taken separately to trace the source of multivariate effects. Overall, considering the three dependent variables together, there was a main effect for condition, $F(1, 76) = 6.20$, $p < .001$, but not for *n* Power, $F(1, 76) = 1.88$, *ns*. Supervisors distributed more of various forms of reward to Man C who was an ingratiator than their counterparts in the control group awarded to Man C who was not an ingratiator. Supervisors high in power motivation did not generally give greater reward to

Man C than did those low in power motivation. A significant interaction effect was obtained, however, $F(1, 76) = 3.23$, $p < .05$, indicating that high-scoring supervisors dispensed greater reward to Man C, who was an ingratiator, than did low-scoring supervisors, whereas high- and low-scoring supervisors treated Man C of the control condition, who was not an ingratiator, about the same. Analysis of the three dependent variables individually revealed that the performance ratings alone accounted for most of the multivariate effect just described. There was a main effect for *n* Power, $F(1, 76) = 4.32$, $p < .05$, for condition, $F(1, 76) = 17.89$, $p < .001$, and for the interaction as well, $F(1, 76) = 8.87$, $p < .01$. Specifically, there was a pronounced tendency for supervisors high in need for power to evaluate very favorably Man C who was an ingratiator. These same supervisors did not evaluate Man C as favorably in the control condition when he was not an ingratiator. Supervisors low in power motivation completed roughly similar performance appraisals on Man C regardless of whether or not he was an ingratiator, and in neither case did these appraisals approach those that high-scoring subjects gave Man C when he was an ingratiator. There was in addition a marginally significant main effect for condition in relation to pay increase, $F(1, 76) = 3.74$, $p < .06$, indicating a tendency for supervisors to award greater pay increase to Man C when he was an ingratiator. None of the remaining univariate analyses done in relation to pay increase or liking proved to be significant.

Post hoc comparisons were made by means of Tukey's honestly significant differences (HSD) test, and Table 1 portrays those results. Student supervisors high on *n* Power clearly gave higher evaluations to the ingratiator than were granted by supervisors to Man C in other conditions of the experiment.

The interaction effect for performance appraisal is illustrated in Figure 1. Supervisors high and low on *n* Power evaluated Man C about the same when he was neutral and noningratiating in his manner, but when Man C assumed the role of ingratiator, supervisors high in power motivation rated his performance substantially higher than did those low on *n* Power.

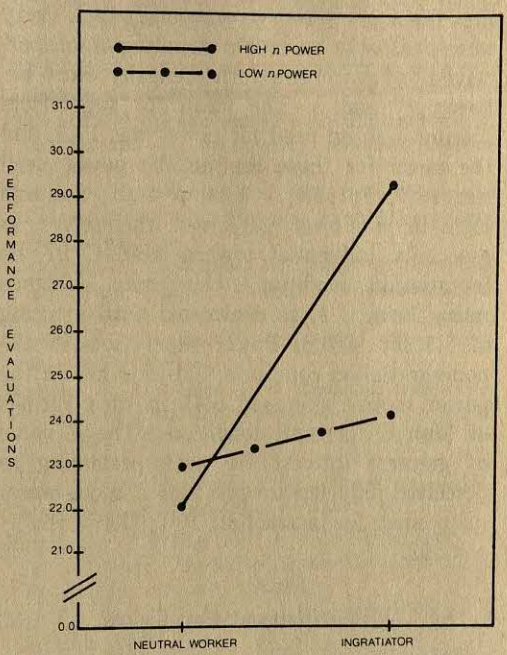


Figure 1. Effects of *n* Power and ingratiation on performance evaluations by student supervisors.

The second hypothesis was that high *n* Power supervisors would make greater distinctions between Man A and Man B, the two neutral workers, in performance appraisal and total amounts of pay increase they awarded. Data were collapsed across the ingratiator and noningratiator conditions so that all 40 supervisors high on the power motive were compared against the 40 who were low. Not only did the performance appraisal data fail to confirm the hypothesis, but they actually indicated a slight but nonsignificant trend in the direction contrary to prediction. Supervisors high in *n* Power gave performance evaluations to Man A and Man B that differed by a mean of 4.1, whereas the corresponding figure for supervisors low on *n* Power was 5.7. For pay increases, the figures were 13.1 and 12.5, respectively, again a nonsignificant difference.

Perceived supervisory influence was the focus of the last hypothesis. A "total influence score" was computed by tallying across workers figures circled to represent how responsive supervisors believed the various workers were to the supervisor's efforts to influence their performance. Individual rat-

ings ran from 9 (very responsive) to 1 (very unresponsive); so scores denoting total perceived influence could theoretically vary between 27 and 3. The mean for supervisors scoring high on need for power was 17.3, and the mean for those low on the power need was 15.9, $F(1, 78) = 4.44$, $p < .05$. Subjects high in *n* Power perceived themselves as especially influential toward Man C in the ingratiation condition, their mean influence rating being 7.3, as contrasted with a rating of 5.5 for high *n* Power supervisors in the noningratiation condition. Subjects low in the power motive averaged 6.45 in their ratings of Man C for both conditions. The *F* value of greatest interest for data pertaining to perceived influence toward Man C alone was a main effect for interaction, $F(1, 78) = 16.72$; $p < .02$.

Discussion and Conclusions

The major findings of the present experiment accord well with Winter's conceptualization of need for power. Ingratiation, even though it is not sincere, may provide the target person with the overt appearance of recognition for exercise of power-oriented behaviors, the quest for which, Winter believes, is integral to the personality of persons high in power motivation. These features were unmistakably present in the personality of Woodrow Wilson, one of the presidents whose inaugural address Winter scored as reflecting high power imagery. Consider the observations of Wilson's biographers, George and George (1956):

> All of Wilson's close friends—the men, the women, the professors, the politicians, the socialites—shared one characteristic: they were, or at least had to seem to him to be, *uncritical* admirers of the man and of everything he did. Intellectual disagreement or the feeling that a friend disapproved of some project he had in hand aroused intolerable anxieties. . . . To . . . Wilson, identity of opinion and love were inseparably linked. He never learned that he and a friend could disagree and still retain a mutual affection. To him, if a friend disagreed with him about a matter of importance, it signified that the friend no longer cared for him. (pp. 31–32)

Maccoby (1976) makes a similar point when he observes that power-seeking managers "must surround themselves with admirers, and admiration becomes a drug for them. . ." (p. 199).

Horney (1937) has argued that the quest for power arises from thwarted childhood needs for parental affection. As viewed in this perspective, ingratiation may have brought reassurance to power-motivated student supervisors that they were worthy of respect or possibly even affection.

Belief by the supervisor high in *n* Power that he influenced the behavior of subordinates may have served a function similar to that provided by ingratiation from subordinates. Both events create a cognitive awareness that one has had an impact on others, that one has elicited a response from them that is contingent upon one's own actions. It is that very awareness, according to our foregoing theoretical analysis, which power-oriented persons especially crave.

Student supervisors high on the power motive, it will be recalled, did not reveal any significant tendency to grant high pay increases to an ingratiating subordinate. The pay increase measure probably was not highly sensitive to differences in response from workers, because it was not until the end of the session that Man C had thoroughly established himself as an ingratiator in the critical condition of the experiment. By that time, much of the opportunity for dispensing differential reward in the form of pay increase had passed.

McClelland and Burnham (1976) found that managers high in power motivation tended to foster organizational climates that were maximally conducive to high work performance. It should be noted, however, that McClelland and Burnham regarded as critical to their analysis high *n* Power scores accompanied by low scores on *n* Affiliation; so their measure was not strictly comparable to the one devised by Winter. Winter (1973) did find, however, that persons aspiring to careers in management scored high in *n* Power.

The McClelland and Burnham finding that managers high in power motivation perform well in the sense of creating an optimal organizational climate lends plausibility to the argument that perhaps high *n* Power people may be susceptible to flattery because, on the basis of past experience with their own per-

formance, they have discovered that they actually *do* make good decisions and perform well at tasks they undertake. Thus the ingratiator may more easily be believed, since the information he transmits accords well with earlier evidence the high *n* Power person has about himself.

An interesting parallel occurs between the present conclusions concerning *n* Power and previous findings on self-confidence. Goodstadt and Kipnis (1970) reported that persons high in self-confidence, as contrasted with persons low in self-confidence, exercised more influence over subordinates, whereas Kipnis and Vanderveer (1971) found persons high in self-confidence to be more susceptible to the ingratiator's flattery. What correlation exists between *n* Power and self-confidence is not known, although it is apparent that McClelland and Burnham (1976) see a heightening of self-confidence as a consequence likely to result from their training program for promoting greater power motivation among managers.

These findings can be viewed as an illustration of the joint interaction of personality and situation as codeterminants of social behavior, especially when contrasted with previous inquiries into use of power. One is reminded of Kurt Lewin's classic formulation, $B = f(P,E)$, which stated that behavior is some interactive function of both person and environment.

The central message of the present investigation for those interested in managerial training is clear. Persons best equipped by personality to influence and control others may not be well suited to evaluate human performance. Training programs should perhaps stress the discontinuity between the control and judgmental functions of management, emphasizing greatly the need for objective systems of performance appraisal.

References

Aronson, E., & Carlsmith, J. M. Experimentation in social psychology. In Lindzey, G., & Aronson, E. (Eds.). *Handbook of social psychology* (2nd ed.; Vol. 2). Reading, Mass.: Addison-Wesley, 1969.

Atkinson, J. W. (Ed.). *Motives in fantasy, action, and society*. Princeton, N.J.: Van Nostrand, 1958.

Fodor, E. M. Disparagement by a subordinate as an influence on use of power. *Journal of Applied Psychology*, 1974, *59*, 652–655.

Fodor, E. M. Group stress, authoritarian style of control, and use of power. *Journal of Applied Psychology*, 1976, *61*, 313–318.

George, A. L., & George, J. L. *Woodrow Wilson and Colonel House: A personality study*. New York: Dover, 1956.

Goodstadt, B., & Kipnis, D. Situational influences on the use of power. *Journal of Applied Psychology*, 1970, *54*, 201–209.

Horney, K. *The neurotic personality of our time*. New York: Norton, 1937.

Hummel, T. J., & Sligo, J. R. Empirical comparison of univariate and multivariate analysis of variance procedures. *Psychological Bulletin*, 1971, *76*, 49–57.

Jones, E. E. *Ingratiation*. New York: Appleton-Century-Crofts, 1964.

Kipnis, D., & Vanderveer, R. Ingratiation and the use of power. *Journal of Personality and Social Psychology*, 1971, *17*, 280–286.

Maccoby, M. *The gamesman*. New York: Simon & Schuster, 1976.

McClelland, D. C. Methods of measuring human motivation. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society*. Princeton, N.J.: Van Nostrand, 1958.

McClelland, D. C., & Burnham, D. H. Power is the great motivator. *Harvard Business Review*, 1976, *54*, March–April, 100–110.

McClelland, D. C., Davis, W. N., Kalin, R., & Wanner, E. *The drinking man*. New York: Free Press, 1972.

McClelland, D. C., & Winter, D. G. *Motivating economic achievement*. New York: Free Press, 1969.

Porter, L. W., & Lawler, E. E. III. *Managerial attitudes and performance*. Georgetown, Ontario: Irwin-Dorsey, 1968.

Uleman, J. S. *A new TAT measure of the need for power*. Unpublished doctoral dissertation, Harvard University, 1966.

Veroff, J. Development and validation of a projective measure of power motivation. *Journal of Abnormal and Social Psychology*, 1957, *54*, 1–8.

Winter, D. G. *The power motive*. New York: Free Press, 1973.

Winter, D. G., & Stewart, A. J. Power motive reliability as a function of retest instructions. *Journal of Consulting and Clinical Psychology*, 1977, *45*, 436–440.

Received February 12, 1979 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 11, 2098-2109

Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence

Charles G. Lord, Lee Ross, and Mark R. Lepper
Stanford University

People who hold strong opinions on complex social issues are likely to examine relevant empirical evidence in a biased manner. They are apt to accept "confirming" evidence at face value while subjecting "disconfirming" evidence to critical evaluation, and as a result to draw undue support for their initial positions from mixed or random empirical findings. Thus, the result of exposing contending factions in a social dispute to an identical body of relevant empirical evidence may be not a narrowing of disagreement but rather an increase in polarization. To test these assumptions and predictions, subjects supporting and opposing capital punishment were exposed to two purported studies, one seemingly confirming and one seemingly disconfirming their existing beliefs about the deterrent efficacy of the death penalty. As predicted, both proponents and opponents of capital punishment rated those results and procedures that confirmed their own beliefs to be the more convincing and probative ones, and they reported corresponding shifts in their beliefs as the various results and procedures were presented. The net effect of such evaluations and opinion shifts was the postulated increase in attitude polarization.

The human understanding when it has once adopted an opinion draws all things else to support and agree with it. And though there be a greater number and weight of instances to be found on the other side, yet these it either neglects and despises, or else by some distinction sets aside and rejects, in order that by this great and pernicious predetermination the authority of its former conclusion may remain inviolate. (Bacon, 1620/1960)

Often, more often than we care to admit, our attitudes on important social issues reflect only our preconceptions, vague impressions, and untested assumptions. We respond to social policies concerning compensatory education, water fluoridation, or energy conservation in terms of the symbols or metaphors they evoke (Abelson, 1976; Kinder & Kiewiet, Note 1) or in conformity with views expressed by opinion leaders we like or respect (Katz, 1957). When "evidence" is brought to bear it is apt to be incomplete, biased, and of marginal probative value—typically, no more than a couple of vivid, concrete, but dubiously representative instances or cases (cf. Abelson, 1972; Nisbett & Ross, in press). It is unsurprising, therefore, that important social issues and policies generally prompt sharp disagreements, even among highly concerned and intelligent citizens, and that such disagreements often survive strenous attempts at resolution through discussion and persuasion.

An interesting question, and one that prompts the present research, involves the consequences of introducing the opposing factions to relevant and objective data. This question seems particularly pertinent for contemporary social scientists, who have frequently called for "more empirically based" social decision making (e.g., Campbell, 1969). Very likely, data providing consistent and

This research was supported in part by Research Grants MH-26736 from the National Institute of Mental Health and BNS-78-01211 from the National Science Foundation to Mark R. Lepper and Lee Ross. The authors would like to express their appreciation to Daryl Bem, Richard Nisbett, Robert Zajonc, and Mark Zanna for their thoughtful comments and suggestions on earlier drafts of this article, and to Lisa Burns for acting as the experimenter.

Requests for reprints should be sent to any of the authors, Department of Psychology, Stanford University, Stanford, California 94305.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3711-2098$00.75

unequivocal support for one or another position on a given issue can influence decision makers and, with sufficiently energetic dissemination, public opinion at large. But what effects can be expected for more mixed or inconclusive evidence of the sort that is bound to arise for most complex social issues, especially where full-fledged experiments yielding decisive and easy-to-generalize results are a rarity? Logically, one might expect mixed evidence to produce some moderation in the views expressed by opposing factions. At worst, one might expect such inconclusive evidence to be ignored.

The present study examines a rather different thesis—one born in an analysis of the layperson's general shortcomings as an intuitive scientist (cf. Nisbett & Ross, in press; Ross, 1977) and his more specific shortcomings in adjusting unwarranted beliefs in the light of empirical challenges (cf. Ross, Lepper, & Hubbard, 1975). Our thesis is that belief polarization will *increase*, rather than decrease or remain unchanged, when mixed or inconclusive findings are assimilated by proponents of opposite viewpoints. This "polarization hypothesis" can be derived from the simple assumption that data relevant to a belief are not processed impartially. Instead, judgments about the validity, reliability, relevance, and sometimes even the meaning of proffered evidence are biased by the apparent consistency of that evidence with the perceiver's theories and expectations. Thus individuals will dismiss and discount empirical evidence that contradicts their initial views but will derive support from evidence, of no greater probativeness, that seems consistent with their views. Through such biased assimilation even a random set of outcomes or events can appear to lend support for an entrenched position, and both sides in a given debate can have their positions bolstered by the same set of data.

As the introductory quotation suggests, the notions of biased assimilation and resulting belief perseverance have a long history. Beyond philosophical speculations and a wealth of anecdotal evidence, considerable research attests to the capacity of preconceptions and initial theories to bias the consideration of subsequent evidence, including work on classic Einstellung effects (Luchins, 1942, 1957), social influence processes (Asch, 1946), impression formation (e.g., Jones & Goethals, 1971), recognition of degraded stimuli (Bruner & Potter, 1964), resistance to change of social attitudes and stereotypes (Abelson, 1959; Allport, 1954), self-fulfilling prophecies (Merton, 1948; Rosenhan, 1973; Snyder, Tanke, & Berscheid, 1977), and the persistence of "illusory correlations" (Chapman & Chapman, 1967, 1969). In a particularly relevant recent demonstration, Mahoney (1977) has shown that trained social scientists are not immune to theory-based evaluations. In this study, professional reviewers' judgments about experimental procedures and resultant publication recommendations varied dramatically with the degree to which the findings of a study under review agreed or disagreed with the reviewers' own theoretical predilections.

Thus, there is considerable evidence that people tend to interpret subsequent evidence so as to maintain their initial beliefs. The biased assimilation processes underlying this effect may include a propensity to remember the strengths of confirming evidence but the weaknesses of disconfirming evidence, to judge confirming evidence as relevant and reliable but disconfirming evidence as irrelevant and unreliable, and to accept confirming evidence at face value while scrutinizing disconfirming evidence hypercritically. With confirming evidence, we suspect that both lay and professional scientists rapidly reduce the complexity of the information and remember only a few well-chosen supportive impressions. With disconfirming evidence, they continue to reflect upon any information that suggests less damaging "alternative interpretations." Indeed, they may even come to regard the ambiguities and conceptual flaws in the data *opposing* their hypotheses as somehow suggestive of the fundamental *correctness* of those hypotheses. Thus, completely inconsistent or even *random* data—when "processed" in a suitably biased fashion—can maintain or even reinforce one's preconceptions.

The present study was designed to examine both the biased assimilation processes that

may occur when subjects with strong initial attitudes are confronted with empirical data concerning a controversial social issue and the consequent polarization of attitudes hypothesized to result when subjects with differing initial attitudes are exposed to a common set of "mixed" experimental results. The social controversy chosen for our investigation was the issue of capital punishment and its effectiveness as a deterrent to murder. This choice was made primarily because the issue is the subject of strongly held views that frequently do become the target of public education and media persuasion attempts, and has been the focus of considerable social science research in the last twenty years. Indeed, as our basic hypothesis suggests, contending factions in this debate often cite and derive encouragement from the same body of inconclusive correlational research (Furman v. Georgia, 1972; Sarat & Vidmar, 1976; Sellin, 1967).

In the present experiment, we presented both proponents and opponents of capital punishment first with the results and then with procedural details, critiques, and rebuttals for two studies dealing with the deterrent efficacy of the death penalty—one study confirming their initial beliefs and one study disconfirming their initial beliefs. We anticipated biased assimilation at every stage of this procedure. First, we expected subjects to rate the quality and probative value of studies confirming their beliefs on deterrent efficacy more highly than studies challenging their beliefs. Second, we anticipated corresponding effects on subjects' attitudes and beliefs such that studies confirming subjects' views would exert a greater impact than studies disconfirming those views. Finally, as a function of these assimilative biases, we hypothesized that the net result of exposure to the conflicting results of these two studies would be an increased polarization of subjects' beliefs on deterrent efficacy and attitudes towards capital punishment.

Method

Subjects

A total of 151 undergraduates completed an in-class questionnaire that included three items on capital punishment. Two to four weeks later, 48 of these students were recruited to participate in a related experiment as partial fulfillment of a course requirement. Twenty-four were "proponents" who favored capital punishment, believed it to have a deterrent effect, and thought most of the relevant research supported their own beliefs. Twenty-four were "opponents" who opposed capital punishment, doubted its deterrent effect, and thought that the relevant research supported *their* views.

Procedure

Upon entering the experiment, mixed groups of proponents and opponents were seated at a large table. The experimenter, blind to subjects' attitudes, told them that they would each be asked to read 2 of 20 randomly selected studies on the deterrent efficacy of the death penalty and asked them to use their own "evaluative powers" in thinking about what the author(s) of the study did, what the critics had to say, and whether the research provided support for one side or the other of this issue.

The experimenter next showed subjects a set of 10 index cards, each containing a brief statement of the results of a single study. Each subject was asked to choose one card and read it silently. In reality, all 10 cards in any one session were identical, providing either prodeterrent information, for example:

> Kroner and Phillips (1977) compared murder rates for the year before and the year after adoption of capital punishment in 14 states. In 11 of the 14 states, murder rates were *lower after* adoption of the death penalty. This research supports the deterrent effect of the death penalty.

or antideterrent information, for example:

> Palmer and Crandall (1977) compared murder rates in 10 pairs of neighboring states with different capital punishment laws. In 8 of the 10 pairs, murder rates were *higher* in the state *with* capital punishment. This research opposes the deterrent effect of the death penalty.

To control for order effects, half of the proponents and half of the opponents saw a "prodeterrence" result first, and half saw an "antideterrence" result first. The studies cited, although invented specifically for the present study, were characteristic of research found in the current literature cited in judicial decisions.

After reading one of these "result cards," subjects answered two sets of questions, on 16-point scales, about changes in their attitudes toward capital punishment (from −8 = more opposed, to 8 = more in favor) and their beliefs about the deterrent efficacy of the death penalty (from −8 = *less* belief that capital punishment has a deterrent effect, to 8 = *more* belief in the deterrent effect). One set of questions examined change occasioned by the single piece of information they had just finished reading; a second set of questions assessed the cumulative change

produced by all of the materials read since the start of the experiment.[1]

Next the experimenter distributed detailed research descriptions bearing code letters corresponding to those on the result cards. The descriptions gave details of the researchers' procedure, reiterated the results, mentioned several prominent criticisms of the study "in the literature," listed the authors' rebuttals to some of the criticisms and depicted the data both in table form and graphically. After reading this more detailed description and critique of the first study, subjects were asked to judge how well or poorly the study had been conducted (from $-8 =$ very poorly done, to $8 =$ very well done), and how convincing the study seemed as evidence on the deterrent efficacy of capital punishment (from $-8 =$ completely unconvincing, to $8 =$ completely convincing).[2] Following this evaluation, subjects were asked to write why they thought the study they had just read did or did not support the argument that capital punishment is a deterrent to murder, and then to answer a second set of attitude and belief change questions on the effects of the description alone and the effects of all experimental materials (i.e., the results and subsequent description and critique) up to that point in time.

Following completion of these questions, the entire procedure was repeated, with a second fictitious study reporting results opposite to those of the first. Again, subjects initially received only a brief description of the results of this second study but were then provided with a detailed presentation of the procedure, results, and critiques. As before, subjects were asked to evaluate both the impact of each single piece of evidence and the impact of all experimental materials up to that point in the experiment on their attitudes toward capital punishment and their beliefs concerning its deterrent efficacy.

To control for possible differences in the inherent plausibility of the two studies, two sets of materials were employed that interchanged the ostensible results of the two invented experiments. The overall design was thus completely counterbalanced with respect to subjects' initial attitudes, order of confirming vs. disconfirming evidence, and the association of the "before–after" vs. "adjacent states" designs with positive or negative results.[3] At the end of the procedure, subjects were carefully debriefed concerning the fictitious nature of the studies and were asked not to reveal this deception to others. In addition to insuring that subjects understood the fictional nature of the experimental materials, this debriefing included discussion of the processes underlying the assimilation of evidence to previous theories and reassurance that a skeptical reaction to poorly designed research is often a praiseworthy cognitive response.

Results

Evaluations of the Two Studies

Our first hypothesis was that subjects holding different initial positions would differentially evaluate the quality and "convincingness" of the same empirical studies and findings. The relevant evaluations, presented in Table 1, revealed strong support for the hypothesized bias in favor of the study that confirmed subjects' initial attitudes.

A two-way analysis of variance (Initial Attitude × Order of Presentation) on the differences between ratings of convincingness of the prodeterrence and antideterrence studies yielded only a main effect of initial attitude, $F(1, 44) = 32.07$, $p < .001$. Proponents

---

[1] Since most of our subjects had reported initial positions at, or very close to, the ends of the attitude and belief scales used for selection purposes, our initial plan to assess attitude polarization—in terms of difference scores assessing changes in subjects' attitudes and beliefs on these same scales from these initial measures to the completion of the experiment—proved impossible. As a substitute, we employed three sorts of measures to assess attitude change. First, we asked subjects, after each new piece of information, to indicate any changes in their attitudes and beliefs occasioned by that single piece of information. Second, we asked subjects, at these same points, to report on "cumulative" changes in their attitudes and beliefs since the start of the experiment. Third, subjects were asked to keep "running records" of their attitudes and beliefs on enlarged versions of the scales initially used for selection purposes. Although all of these measures individually raise some problems, the congruence of data across these different measurement devices gives us some confidence concerning the results reported. Indeed, because the results obtained on the "running record" measure so completely parallel the findings obtained on the cumulative change question depicted in Figures 1 and 2, in terms of both the array of means and the obtained significance levels, the data from this measure will not be reported separately.

[2] Subjects were also asked, at this point, whether they thought the researchers had favored or opposed the death penalty and whether they thought an unbiased consideration should lead one to treat the study as evidence for or against capital punishment. Analyses on the first question showed only that subjects believed the researchers' attitudes to coincide with their stated results. Analyses on the second question proved wholly redundant with those presented for the "convincingness" and "well done" questions.

[3] Preliminary analyses were conducted to see if the particular association of positive versus negative results with either the before–after or adjacent-states designs would affect the results obtained. There were no significant effects or interactions involving this variation in stimulus materials; hence, the data were collapsed across this factor.

Table 1
Evaluations of Prodeterrence and Antideterrence Studies by Proponents and Opponents of Capital Punishment

Study	Proponents	Opponents
Mean ratings of how well the two studies had been conducted		
Prodeterrence	1.5	−2.1
Antideterrence	−1.6	− .3
Difference	3.1	−1.8
Mean ratings of how convincing the two studies were as evidence on the deterrent efficacy of capital punishment		
Prodeterrence	1.4	−2.1
Antideterrence	−1.8	.1
Difference	3.2	−2.2

Note. Positive numbers indicate a positive evaluation of the study's convincingness or procedure. Negative numbers indicate a negative evaluation of the study's convincingness or procedure.

regarded the prodeterrence study as significantly more convincing than the antideterrence study, $t(23) = 5.18$, $p < .001$,[4] regardless of whether it was the "before–after" design that suggested the efficacy of capital punishment and the "adjacent states" design that refuted it, or vice versa. Opponents, by contrast, regarded the prodeterrence study as significantly less convincing than the antideterrence study, $t(23) = -3.02$, $p < .01$, again irrespective of which research design was purported to have produced which type of results. The same was true of the difference between ratings of how well done the two studies had been, $F(1, 44) = 33.52$, $p < .001$.[5] As above, proponents found the prodeterrence study to have been better conducted than the antideterrence study, $t(23) = 5.37$, $p < .001$, whereas opponents found the prodeterrence study to have been less well conducted, $t(23) = -2.80$, $p < .05$. As one might expect, the correlation between the "convincingness" and "well done" questions was substantial, $r = .67$, $p < .001$.

These differing opinions of the quality of the two studies were also reflected in subjects' written comments. At the risk of opening ourselves to a charge of "biased assimilation," we present a set of subjects' comments—selected for dramatic effect but not unrepresentative in content—in Table 2. As these comments make clear, the same study can elicit entirely opposite evaluations from people who hold different initial beliefs about a complex social issue. This evidence of bias in subjects' evaluations of the quality and convincingness of the two studies is consistent with the biased assimilation hypothesis and sets the stage for testing our further predictions concerning attitude and belief polarization.

Overall Attitude Polarization

Given such biased evaluations, our primary hypothesis was that exposure to the "mixed"[6] data set comprised by the two studies would result in a further polarization of subjects' attitudes and beliefs rather than the convergence that an impartial consideration of these inconclusive data might warrant. To test this hypothesis requires a consideration

[4] All p values reported in this article are based on two-tailed tests of significance.

[5] In order to examine possible main effects of either study direction or initial attitude on subjects' ratings of how convincing and how well done the studies were—findings that would not be portrayed in the difference score analysis reported—a three-way analysis of variance (Initial Attitude × Order of Presentation × Direction of Study) was also performed. There were no main effects of study direction on either measure. A main effect of initial attitude—indicating that opponents evaluated the total set of evidence more negatively than did proponents—proved significant for the "well done" question, $F(1, 44) = 4.69$, $p < .05$, but not for the "convincing" question, $F(1, 44) = 1.53$, *ns*.

[6] The term *mixed*, we should emphasize, refers to the fact that one study yielded evidence confirming the deterrent efficacy of the death penalty, whereas the other study yielded evidence disconfirming such efficacy (with appropriate counterbalancing of purported procedures and purported results). Subjects, regardless of initial position, clearly recognized this discrepancy between results, as will be apparent in our analyses of their responses to the simple statements of the study's main findings. We do not mean to imply that the subjects "phenomenologically" judged the two studies to be of equal probative value; indeed, as indicated in the preceding discussion, identical procedures were clearly judged to differ in their probativeness depending on the congruity between the study's outcomes and the subject's initial beliefs.

Table 2
Selected Comments on Prodeterrence and Antideterrence Studies by Proponents and Opponents of Capital Punishment

Subject	Comments on Prodeterrence study	Comments on Antideterrence study
	Set 1 materials	
S8 Proponent	"It does support capital punishment in that it presents facts showing that there is a deterrent effect and seems to have gathered data properly."	"The evidence given is relatively meaningless without data about how the overall crime rate went up in those years."
S24 Proponent	"The experiment was well thought out, the data collected was valid, and they were able to come up with responses to all criticisms."	"There were too many flaws in the picking of the states and too many variables involved in the experiment as a whole to change my opinion."
S35 Opponent	"The study was taken only 1 year before and 1 year after capital punishment was reinstated. To be a more effective study they should have taken data from at least 10 years before and as many years as possible after."	"The states were chosen at random, so the results show the average effect capital punishment has across the nation. The fact that 8 out of 10 states show a rise in murders stands as good evidence."
S36 Opponent	"I don't feel such a straightforward conclusion can be made from the data collected."	"There aren't as many uncontrolled variables in this experiment as in the other one, so I'm still willing to believe the conclusion made."
	Set 2 materials	
S14 Proponent	"It shows a good direct comparison between contrasting death penalty effectiveness. Using neighboring states helps to make the experiment more accurate by using similar locations."	"I don't think they have complete enough collection of data. Also, as suggested, the murder rates should be expressed as percentages, not as straight figures."
S15 Proponent	"It seems that the researchers studied a carefully selected group of states and that they were careful in interpreting their results."	"The research didn't cover a long enough period of time to prove that capital punishment is not a deterrent to murder."
S25 Opponent	"The data presented are a randomly drawn set of 10. This fact seems to be the study's biggest problem. Also many other factors are not accounted for which are very important to the nature of the results."	"The murder rates climbed in all but two of the states after new laws were passed and no strong evidence to contradict the researchers has been presented."
S38 Opponent	"There might be very different circumstances between the sets of two states, even though they were sharing a border."	"These tests were comparing the same state to itself, so I feel it could be a fairly good measure."

of subjects' final attitudes, after exposure to both studies and related critiques and rebuttals, relative to the start of the experiment.

The relevant data provide strong support for the polarization hypothesis. Asked for their final attitudes relative to the experi-

Table 3
Mean Attitude and Belief Changes for a Single Piece of Information

Issue and study	Initial attitudes: Proponents	Initial attitudes: Opponents
Results only		
Capital punishment		
Prodeterrence	1.3	0.4
Antideterrence	−0.7	−0.9
Combined	0.6	−0.5
Deterrent efficacy		
Prodeterrence	1.9	0.7
Antideterrence	−0.9	−1.6
Combined	1.0	−0.9
Details, data, critiques, rebuttals		
Capital punishment		
Prodeterrence	0.8	−0.9
Antideterrence	0.7	−0.8
Combined	1.5	−1.7
Deterrent efficacy		
Prodeterrence	0.7	−1.0
Antideterrence	0.7	−0.8
Combined	1.4	−1.8

Note. Positive numbers indicate a more positive attitude or belief about capital punishment and its deterrent effect. Negative numbers indicate a more negative attitude or belief about capital punishment and its deterrent effect.

ment's start, proponents reported that they were *more* in favor of capital punishment, $t(23) = 5.07$, $p < .001$, whereas opponents reported that they were *less* in favor of capital punishment, $t(23) = -3.34$, $p < .01$. In a two-way analysis of variance (Initial Attitude × Order of Presentation), the effect of initial attitude was highly significant, $F(1, 44) = 30.06$, $p < .001$, and neither the order effect nor the interaction approached significance. Similar results characterized subjects' beliefs about deterrent efficacy. Proponents reported greater belief in the deterrent effect of capital punishment, $t(23) = 4.26$, $p < .001$, whereas opponents reported less belief in this deterrent effect, $t(23) = -3.79$, $p < .001$. Final attitudes toward capital punishment and beliefs concerning deterrent efficacy were highly correlated, $r = .88$, $p < .001$.

Such results provide strong support for the main experimental hypothesis that inconclusive or mixed data will lead to increased polarization rather than to uncertainty and moderation. Moreover, the degree of polarization shown by individual subjects was predicted by differences in subjects' willingness to be less critical of procedures yielding supportive evidence than of procedures yielding nonsupportive evidence. Significant correlations were found between overall attitude change regarding capital punishment and differences in ratings of both how convincing, $r = .56$, $p < .001$, and how well done, $r = .56$, $p < .001$, the studies were. Overall changes in beliefs in deterrent efficacy produced comparable correlations of .53 and .57, both $ps < .001$.

Components of Attitude Polarization

In view of this strong evidence of overall attitude polarization, it is worth examining the course of attitude polarization as subjects' opinions were successively assessed after exposure to the first study, the details and critiques of the first study, the results of the second study, and the details and critiques of the second study. At each stage, it will be recalled, subjects were asked about the impact of the single piece of information they had just considered and the cumulative impact of all information presented to that point. Let us first examine the reported effects of single segments of evidence and then the effects of accumulated evidence over time.

Effect of Exposure to the Results of Each Study

Considering the result cards as single pieces of evidence, both proponents and opponents reported shifting their attitudes in the direction of the stated results for both the pro-deterrence, $t(47) = 4.67$, $p < .001$, and anti-deterrence, $t(47) = -5.15$, $p < .001$, studies. As shown in the top half of Table 3, however, subjects' responses to the two studies also varied with initial attitude. Proponents tended to be influenced more by the prodeterrence study and opponents more by the anti-deterrence study. Thus a two-way analysis of variance (Initial Attitude × Order of Pre-

sentation) on combined change from the two result cards considered individually yielded only a main effect of initial attitude for both attitudes toward the death penalty, $F(1, 44) = 6.35$, $p < .02$,[7] and beliefs about its deterrent effect, $F(1, 44) = 10.37$, $p < .01$. Interestingly, the analysis of beliefs regarding deterrent efficacy also showed an unanticipated interaction effect, $F(1, 44) = 7.48$, $p < .01$, with proponents showing a differential response to results alone regardless of order of presentation but opponents showing a differential response to results alone only when the confirming study was presented first.

Effect of Exposure to Procedures and Data, Critiques and Rebuttals

When provided with a more detailed description of the procedures and data, together with relevant critiques and authors' rebuttals, subjects seemed to ignore the stated results of the study. As shown in the bottom half of Table 3, both proponents and opponents interpreted the additional information, relative to the results alone, as strongly supporting their own initial attitudes. Detailed descriptions of either the prodeterrence or the antideterrence study, with accompanying critiques, caused proponents to favor capital punishment more and believe in its deterrent efficacy more, but caused opponents to oppose capital punishment more and believe in its deterrent efficacy less. A two-way analysis of variance (Initial Attitude × Order of Presentation) on attitude change for the two descriptions combined yielded only a significant main effect of initial attitude for both the capital punishment issue, $F(1, 44) = 28.10$, $p < .001$, and the deterrent efficacy question, $F(1, 44) = 26.93$, $p < .001$.

Changes in Attitudes Across Time

Subjects' reported changes in attitudes and beliefs, relative to the start of the experiment, following exposure to each of the four separate pieces of information are depicted in Figure 1 for attitudes concerning capital punishment and in Figure 2 for beliefs concerning deterrent efficacy. These data, portrayed separately for subjects who received first either the prodeterrence study or the antideterrence study, provide a more detailed view of the attitude polarization process. They allow, as well, an examination of the hypothesized "rebound effect," that the provision of any plausible reason for discounting data that contradict one's preconceptions will eliminate the effects that mere knowledge of those data may have produced.

The existence of such a "rebound effect" is obvious from examination of these figures. Whether they encountered the disconfirming result first or second, both proponents and opponents seemed to be swayed momentarily by this evidence, only to revert to their former attitudes and beliefs (and in 23% of the individual cases, to even more extreme positions) after inspecting the procedural details and data, and the critiques and rebuttals found in the literature. Across all subjects, this rebound in opinions proved significant for both the capital punishment, $t(47) = 4.43$, $p < .001$, and deterrent efficacy, $t(47) = 4.58$, $p < .001$, issues. By contrast, no compensating rebound effects resulted from reading the descriptions and critiques of studies supporting subjects' initial attitudes, for either capital punishment, $t(47) = .60$, *ns*, or deterrent efficacy, $t(47) = .23$, *ns*.

Discussion

The results of the present experiment provide strong and consistent support for the attitude polarization hypothesis and for the biased assimilation mechanisms postulated to underlie such polarization. The net effect of exposing proponents and opponents of capital punishment to identical evidence—studies ostensibly offering equivalent levels of support and disconfirmation—was to increase further the gap between their views. The mechanisms responsible for this polarization of subjects' attitudes and beliefs were clearly

---

[7] In order to rule out the possibility that direction of study interacted with initial attitude, a three-way analysis of variance (Initial Attitude × Order of Presentation × Direction of Study) was also performed on these data. The relevant interaction term did not approach significance, $F(1, 44) = 1.62$, *ns*.

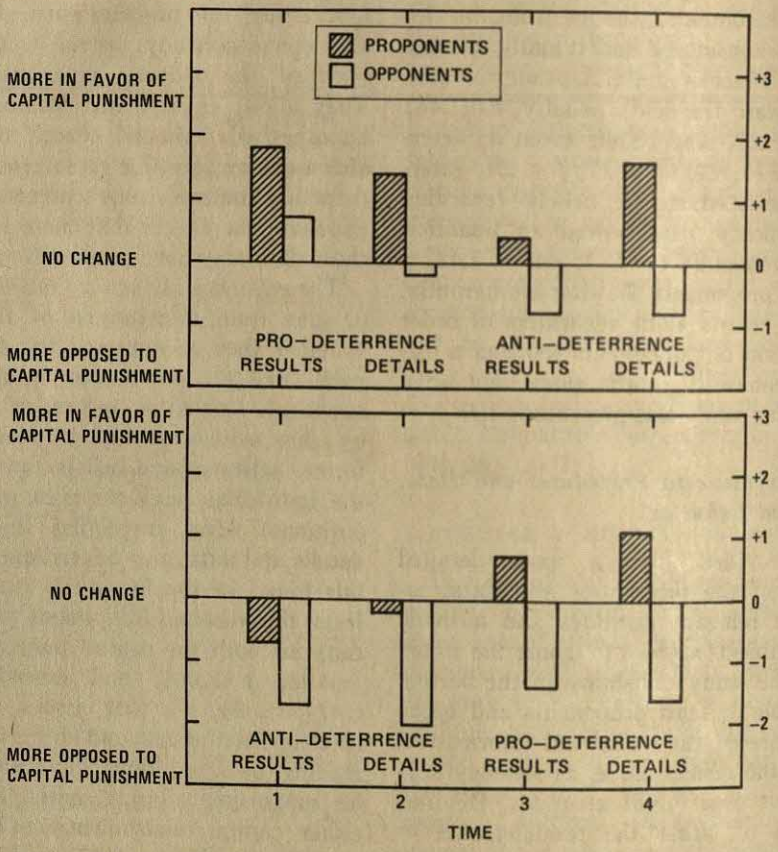


Figure 1. Top panel: Attitude changes on capital punishment relative to start of experiment as reported across time by subjects who received prodeterrence study first. Bottom panel: Attitude changes on capital punishment relative to start of experiment as reported across time by subjects who received antideterrence study first.

suggested by correlational analyses. Subjects' decisions about whether to accept a study's findings at face value or to search for flaws and entertain alternative interpretations seemed to depend far less on the particular procedure employed than on whether the study's results coincided with their existing beliefs.

The Normative Issue

It is worth commenting explicitly about the normative status of our subjects' apparent biases. First, there can be no real quarrel with a willingness to infer that studies supporting one's theory-based expectations are more probative than, or methodologically superior to, studies that contradict one's expectations. When an "objective truth" is known or strongly assumed, then studies whose outcomes reflect that truth may reasonably be given greater credence than studies whose outcomes fail to reflect that truth. Hence the physicist would be "biased," but appropriately so, if a new procedure for evaluating the speed of light were accepted if it gave the "right answer" but rejected if it gave the "wrong answer." The same bias leads most of us to be skeptical about reports of miraculous virgin births or herbal cures for cancer, and despite the risk that such theory-based and experience-based skepticism may render us unable to recognize a miraculous event when it occurs, overall we are surely well served by our bias. Our subjects' willingness to impugn or defend findings as a function of their conformity to expectations can, in part, be similarly defended. Only the strength of

their initial convictions in the face of the existing inconclusive social data and arguments can be regarded as "suspect."

Our subjects' main inferential shortcoming, in other words, did not lie in their inclination to process evidence in a biased manner. Willingness to interpret new evidence in the light of past knowledge and experience is essential for any organism to make sense of, and respond adaptively to, its environment. Rather, their sin lay in their readiness to use evidence already processed in a biased manner to bolster the very theory or belief that initially "justified" the processing bias. In so doing, subjects exposed themselves to the familiar risk of making their hypotheses unfalsifiable—a serious risk in a domain where it is clear that at least one party in a dispute holds a false hypothesis—and allowing themselves to be encouraged by patterns of data that they ought to have found troubling. Through such processes laypeople and professional scientists alike find it all too easy to cling to impressions, beliefs, and theories that have ceased to be compatible with the latest and best evidence available (Mahoney, 1976, 1977).

Polarization: Real or Merely Reported?

Before further pursuing the broader implications of the present demonstration, it is necessary to consider an important question raised by our procedure: Did our subjects really show change (i.e., polarization) in their private beliefs about the desirability and deterrent efficacy of capital punishment? Certainly they told us, explicitly, that their

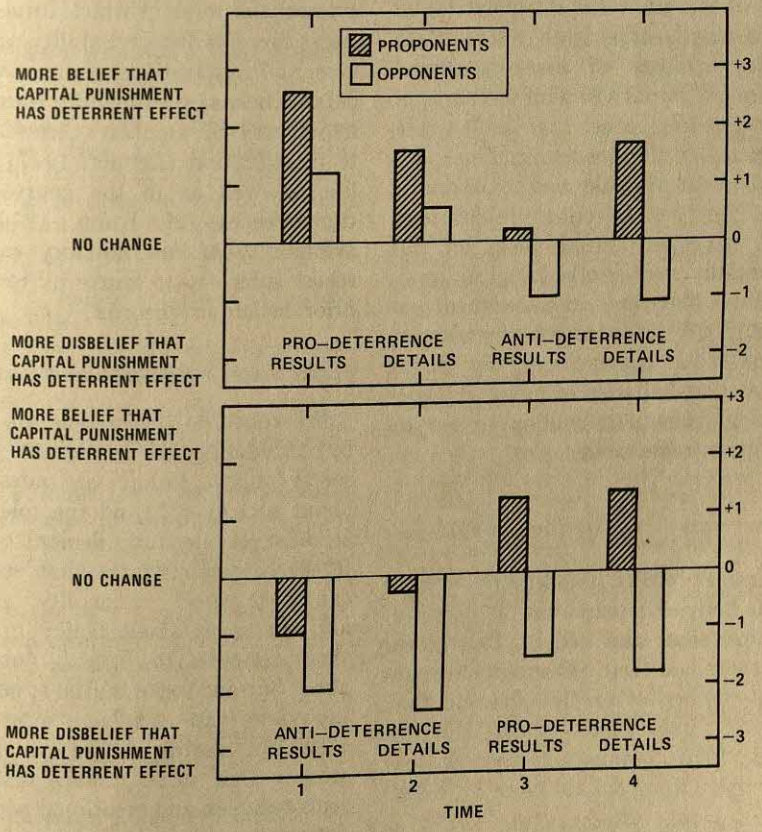


Figure 2. Top panel: Belief changes on capital punishment's deterrent efficacy relative to start of experiment as reported across time by subjects who received prodeterrence study first. Bottom panel: Belief changes on capital punishment's deterrent efficacy relative to start of experiment as reported across time by subjects who received antideterrence study first.

attitudes and beliefs did change after each new piece of evidence was presented, and from the beginning to the end of the experiment. Moreover, they did show a willingness to report a shift in their attitudes in the direction of findings that were contrary to their beliefs, at least until those findings were exposed to methodological scrutiny and possible alternative interpretations. Nevertheless, it could be argued that subjects were not reporting real shifts in attitudes but instead were merely reporting what they believed to be a rational or appropriate response to each increment in the available evidence. Although we believe that it remains an impressive demonstration of assimilation biases to show that contending factions both believe the same data to justify their position "objectively," the potential limitations of the present measures should be kept in mind in evaluating the relationship of this study to prior polarization research. As noted earlier (see Footnote 1) our intended strategy of assessing direct changes from our initial selection measures of attitudes and beliefs, rather than asking subjects to report such changes within the experiment, was neither feasible nor appropriate, given the necessity of selecting subjects with strong and consistent initial views on this issue. Potentially such methodological problems could be overcome in subsequent research through the use of less extreme samples or, perhaps more convincingly, by seeing whether biased assimilation of mixed evidence will make subjects more willing to *act* on their already extreme beliefs.

Belief Perseverance and Attribution Processes

The present results importantly extend the growing body of research on the perseverance of impressions and beliefs. Two of the present authors and their colleagues have now amassed a number of studies showing that, once formed, impressions about the self (Ross et al., 1975; Jennings, Lepper, & Ross, Note 2; Lepper, Ross, & Lau, Note 3), beliefs about other people (Ross et al., 1975), or theories about functional relationships between variables (Anderson, Lepper, & Ross, Note 4) can survive the total discrediting of the evidence that first gave rise to such beliefs. In essence, these prior studies demonstrate that beliefs can survive the complete subtraction of the critical formative evidence on which they were initially based. In a complementary fashion, the present study shows that strongly entrenched beliefs can also survive the addition of nonsupportive evidence.

These findings pose some fundamental questions for traditional attribution models. To the extent that beliefs and impressions can be shown to persevere in the face of subsequent challenging data, we need a "top down" rather than—or perhaps in conjunction with —a "bottom up" approach (cf. Bobrow & Norman, 1975) to the question of how individuals extract meaning from their social environment. Instead of viewing people as impartial, data-driven processors, the present research suggests our models must take into account the ways in which intuitive scientists assess the relevance, reliability, representativeness, and implications of any given sample of data or behavior within the framework of the hypotheses or implicit theories they bring to the situation (Lepper, 1977). In everyday life, as well as in the course of scientific controversies (cf. Kuhn, 1970), the mere availability of contradictory evidence rarely seems sufficient to cause us to abandon our prior beliefs or theories.

Social Science Research and Social Policy

We conclude this article, as we began it, by considering the important links between social policy, public attitudes and beliefs about such policy, and the role of the social scientist. If our study demonstrates anything, it surely demonstrates that social scientists can not expect rationality, enlightenment, and consensus about policy to emerge from their attempts to furnish "objective" data about burning social issues. If people of opposing views can each find support for those views in the same body of evidence, it is small wonder that social science research, dealing with complex and emotional social issues and forced to rely upon inconclusive designs, measures, and modes of analysis, will frequently fuel rather than calm the fires of debate.

Reference Notes

1. Kinder, D. R., & Kiewiet, D. R. *Economic grievances and political behavior: The role of personal discontents and symbolic judgments.* Paper presented at the meeting of the American Psychological Association, San Francisco, August 1977.
2. Jennings, D. L., Lepper, M. R., & Ross, L. *Persistence of impressions of personal persuasiveness: Perseverance of erroneous self assessments outside the debriefing paradigm.* Unpublished manuscript, Stanford University, 1978.
3. Lepper, M. R., Ross, L., & Lau, R. R. *Persistence of inaccurate and discredited personal impressions: A field demonstration of attributional perseverance.* Unpublished manuscript, Stanford University, 1979.
4. Anderson, C. A., Lepper, M. R., & Ross, L. *Theory perseverance: The persistent influence of discredited information.* Unpublished manuscript, Stanford University, 1979.

References

Abelson, R. P. Modes of resolution of belief dilemmas. *Conflict Resolution*, 1959, *3*, 343–352.

Abelson, R. P. Are attitudes necessary? In B. T. King & E. McGuinnies (Eds.), *Attitudes, conflict, and social change*. New York: Academic Press, 1972.

Abelson, R. P. Script processing in attitude formation and decision making. In J. S. Carroll and J. W. Payne (Eds.), *Cognition and social behavior*. Hillsdale, N.J.: Erlbaum, 1976.

Allport, G. W. *The nature of prejudice*. Reading, Mass.: Addison-Wesley, 1954.

Asch, S. Forming impressions of personality. *Journal of Abnormal and Social Psychology*, 1946, *41*, 258–290.

Bacon, F. *The new organon and related writings*. New York: Liberal Arts Press, 1960. (Originally published, 1620.)

Bobrow, D. G., & Norman, D. A. Some principles of memory schemata. In D. G. Bobrow & A. Collins (Eds.), *Representation and understanding: Studies in cognitive science*. New York: Academic Press, 1975.

Bruner, J. S., & Potter, M. C. Interference in visual recognition. *Science*, 1964, *144*, 424–425.

Campbell, D. T. On reforms as experiments. *American Psychologist*, 1969, *24*, 409–429.

Chapman, L., & Chapman, J. The genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology*, 1967, *72*, 193–204.

Chapman, L., & Chapman, J. Illusory correlation as an obstacle to the use of valid psychodiagnostic signs. *Journal of Abnormal Psychology*, 1969, *74*, 271–280.

Furman v. Georgia, 408 U.S. *238*, 1972.

Jones, E. E., & Goethals, G. R. Order effects in impression formation. Attribution context and the nature of the entity. In E. E. Jones et al. (Eds.), *Attribution: Perceiving the causes of behavior*. Morristown, N.J.: General Learning Press, 1971.

Katz, E. The two-step flow of communication: An up-to-date report on an hypothesis. *Public Opinion Quarterly*, 1957, *21*, 61–68.

Kuhn, T. S. *The structure of scientific revolutions* (2nd ed.). Chicago: University of Chicago Press, 1970.

Lepper, M. R. Cognition and social behavior. *Science*, 1977, *196*, 765–766.

Luchins, A. S. Mechanization in problem solving: The effect of Einstellung. *Psychological Monographs*, 1942, *54*, 1–95.

Luchins, A. S. Experimental attempts to minimize the impact of first impressions. In C. I. Hovland (Ed.), *The order of presentation in persuasion*. New Haven, Conn.: Yale University Press, 1957.

Mahoney, M. J. *Scientist as subject: The psychological imperative*. Cambridge, Mass.: Ballinger, 1976.

Mahoney, M. J. Publication prejudices: An experimental study of confirmatory bias in the peer review system. *Cognitive Therapy and Research*, 1977, *1*, 161–175.

Merton, R. K. The self-fulfilling prophecy. *Antioch Review*, 1948, *8*, 193–210.

Nisbett, R., & Ross, L. *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, N.J.: Prentice-Hall, in press.

Rosenhan, D. L. On being sane in insane places. *Science*, 1973, *179*, 250–258.

Ross, L. The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10). New York: Academic Press, 1977.

Ross, L., Lepper, M. R., & Hubbard, M. Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm. *Journal of Personality and Social Psychology*, 1975, *32*, 880–892.

Sarat, A., & Vidmar, N. Public opinion, the death penalty, and the Eighth Amendment: Testing the Marshall hypothesis. *Wisconsin Law Review*, 1976, 171–197.

Sellin, T. (Ed.). *The death penalty*. New York: Harper & Row, 1967.

Snyder, M., Tanke, E. D., & Berscheid, E. Social perception and interpersonal behavior: On the self-fulfilling nature of social stereotypes. *Journal of Personality and Social Psychology*, 1977, *35*, 656–666.

Received February 16, 1979 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 11, 2110-2120

Sex Differences in Bystander Intervention in a Theft

William Austin
University of Virginia

A series of studies demonstrated a strong relationship among the situation-defining variable of degree of harm to victim, sexual configuration among participants, and bystanders' willingness to intervene to stop a theft. A pretest showed that a prior verbal commitment was absolutely necessary for intervention. The remaining data showed that high harm to a victim produced a high rate of intervention and showed strong sex differences in helping behavior in low-harm conditions. A high percentage of female bystanders helped in both low- and high-harm situations, whereas frequent helping by males was observed only when harm to the victim was high. Female victims elicited a significantly greater amount of helping, and sex of thief had no effect. A significant sex of bystander, sex of victim, and harm to victim interaction best describes the data. Results are interpreted in terms of different motivational sets held by males and females when they are responsible for the fate of others. Results also support the utility of an interactionist approach to the question of how individual and situational variables influence prosocial action.

In recent years social psychologists have generated a voluminous amount of research on the topic of helping behavior in an attempt to understand the dynamics of prosocial action. Two general and interrelated questions seem to underlie most studies. The first question posed by researchers points to the situation-defining properties of variables. It asks what personal and situational factors encourage people to define a situation as one in which help is needed or absolutely required (e.g., an emergency). The second question is behavioral and asks what factors facilitate or inhibit helping once a situation is defined as one in which help is needed or required.

Most variables studied by researchers are pertinent to answering both questions, which taken together constitute a unitary process of prosocial action. For example, the ambiguity of situational cues such as a victim's distress (Clark & Word, 1974), number of bystanders present (Latané & Darley, 1970), and degree of victim need (Ashton & Severy, 1976) has been shown to directly affect the definition of a situation as an emergency and to indirectly activate helping behavior.

Both questions also find their way into all of the popular models of helping behavior, including the decision-making models of Latané and Darley (1970) and Schwartz (1977), the cost model offered by J. A. Piliavin and Piliavin (1972), and Staub's (1978) Person × Situation interactive approach. In addition, these models introduce variables and concepts designed to explicate the intervening processes that stand at the interface between the two questions.

The two general concepts of responsibility and cost predominate as the main explanatory tools in these models. Both concepts have multiple dimensions of meaning and have been employed in very different ways by researchers.

Responsibility has been given three distinct meanings in the helping literature. Theorists have described (a) the diffusion of respon-

This project was made possible in part by a Russell Sage Foundation residency fellowship in law and social science granted to the author. The author would especially like to thank Kristine Dwornik Bragg, who supervised the gathering of much of the data presented here.

Requests for reprints should be sent along with a stamped, self-addressed envelope to William Austin, Department of Psychology, Gilmer Hall, University of Virginia, Charlottesville, Virginia 22901.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3711-2110$00.75

sibility among multiple observers of an emergency (Latané & Darley, 1970), (b) personality differences in individuals' willingness to assume responsibility for the fate of others (Schwartz, 1970), and (c) the role of responsibility in social norms that prescribe that help be given to people in need (Berkowitz & Daniels, 1964; Walster & Piliavin, 1972). These separate usages of the concept of responsibility complement one another; each is appropriate for different research questions. It is clear, however, that some aspect of responsibility must be taken into account in describing any prosocial act. Responsibility is the sine qua non of helping. Every helping action requires the observer to role play and see the situation from the victim's perspective. In fact, the labeling of a particular action as helpful means not only that one person has made another's plight more pleasant or less painful but, by definition, that the first person has assumed responsibility for the fate of the second.

The concept of cost signifies a behavioral orientation. J. A. Piliavin and Piliavin (1972) presented a parsimonious cost model of helping behavior. This model consists of a 2 × 2 matrix of potential helping situations, with costs for helping matched against costs for not helping. Costs for helping that would typically discourage bystanders from intervening might be time, risk of physical injury, and being rebuked for not "minding one's own business." Costs for not helping can be divided into the risks of external sanctions, such as social ostracism, and the risks of internal sanctions, typically described as guilt or self-blame.

If one assumes that most of the cost factors relevant to a helping situation can be measured or reliably guessed, the Piliavin model makes straightforward predictions in two of the four cells. When costs for helping are low and costs for not helping are high, then a high rate of helping is expected. Conversely, when costs for helping are high and costs for not helping are low, then prosocial acts are naturally less frequent. However, predicting responses in the remaining two situational variations in the Piliavin model (high/high, low/low) is problematic.

The high/high situation is particularly interesting to researchers because it poses an interpersonal paradox, a "Catch-22," in which there is no satisfying way to resolve what is a classic case of avoidance—avoidance tension. J. A. Piliavin and Piliavin (1972) speculated that individual differences might be more influential under such circumstances, which could mean personality differences relevant to prosocial action (Schwartz, 1977), sex differences, or Person × Situation interactions such that individual characteristics may alter responses as a function of the relative strength of certain situation-defining variables (Staub, 1978).

This study began with the assumption that both responsibility and cost are necessary explanatory concepts in most helping situations. The research reported here was designed to test the effect of one situation-defining variable (degree of harm to a victim) and to explore possible sex differences under conditions that on one hand are conducive to helping (prior commitment obtained and person alone), yet pose the high/high cost dilemma on the other.

The type of helping behavior called for was intervention to stop a theft. Degree of harm to victim was manipulated by varying the value of the items stolen. Sex differences were studied by varying the sex of the bystander, thief, and victim. The experimental procedure first attempted to create pressures so that bystanders would have to make a considerable effort to deny personal responsibility for intervening. Subjects were alone so as to rule out diffusion of responsibility (Latané & Darley, 1970) and were engaged in casual activity (usually reading) so as to eliminate lack of helping due to competing commitments. In addition, each bystander subject was required to make a verbal commitment to watch the stolen items in the victim's absence. This setting of a clear obligation seems to create the high/high cost dilemma when coupled with the realities involved in stopping a thief. There are costs for not helping (incurring the victim's wrath, guilt for being complacent) and anticipated costs for helping (risk of physical injury, embarrassment, or incurring the wrath of a wrongfully

accused thief). I thus assumed the setting was ripe for investigating whether a clear situation-defining variable such as degree of harm, bystander/thief/victim sexual configuration, or both factors in combination would play a significant role in bystanders' attempts to resolve the high/high cost dilemma under conditions of high personal responsibility.

Predictions

1. I predicted a significant positive relationship between degree of harm to a victim and frequency of helping. I reasoned that greater harm would enhance the salience both of the bystander's personal responsibility and of the potential costs for not helping. The concept of degree of harm should not be confused with the similar but distinct concepts of need and dependency, which have both been shown to be correlated with helping (see Berkowitz & Daniels, 1963; J. A. Piliavin & Piliavin, 1972). Degree of harm concerns specific consequences for not helping, whereas need seems to address the probability of negative consequences if help is not forthcoming, and dependency concerns the degree of responsibility for intervening resulting from an existing relationship between victim and bystander. By my procedure I attempted to create the impression that negative consequences would certainly befall the victim while holding need and dependency constant.

The effect of degree of harm on helping under conditions of personal responsibility and high cost has not been studied previously. One study (Schwartz, 1970) varied degree of harm (or salience of consequences) in a study of willingness to donate bone marrow. Schwartz reported a curvilinear relationship between consequences and helping, but his manipulation of intensity of appeal to make a commitment and probability of the donor actually being called makes it difficult to compare his research with the present research.

2. This study was conceived as exploratory with respect to possible sex differences and Sex × Harm to Victim interactions. Although I was confident the high personal responsibility and high/high cost setting was best suited for discovering sex differences, the lack of a sound theory of sex differences in moral behavior and equivocal research findings precluded my stating a priori propositions (see Staub, 1978). Though researchers have reported some sex differences in prosocial behavior, the literature is far from consistent on the topic. Further difficulty stems from the usual problems in making between-studies comparisons of results in light of the different variables and operations used. In particular, researchers have not fully examined the potential effect of the sexual configuration of all of the participants. Numerous studies exist in which the sex of the bystander, the victim, or the wrongdoer has been studied individually, but only a few researchers have simultaneously varied the sex of more than one participant, creating a situation in which very little is known about the role of sexual configuration in helping. To my knowledge, prior to the research reported here, no one has orthogonally varied the sex of bystander, wrongdoer, and victim in a single study. Despite the lack of reliable empirical generalizations in this area I was hopeful that insights could be gleaned from previous studies and that these possibilities might combine with the operational setting employed in this research to shed new light on sex differences in prosocial action.

There is reason to expect sex differences in helping in the high/high cost situation created here. A previous study (Austin & McGinn, 1977) reported sex differences in reward allocation behavior and significant differences in verbal reports of motivations for allocation decisions. Specifically, under conditions of expectation of future interaction, males reported that their behavior was more influenced by cost avoidance considerations, whereas females rated friendship or interpersonal harmony factors as more important. There is some justification for extrapolating from reward allocation to helping, in that one could argue that both situations involve questions associated with treating another person fairly and what outcomes he or she will receive (see Walster & Piliavin, 1972). Accordingly, in the present research, females might be expected to show an overall higher rate of helping because of a greater sensitivity

to interpersonal harmony vis-à-vis the victim (i.e., costs for not helping). If this assumption is correct, then females might be willing to risk instigating conflict with a presumed thief because of a concern with harmony, whereas males might be more willing to risk harmony to avoid costs involved with intervention.

Some support for this conjecture can be found in helping studies. Most encouraging are studies that purport to show greater helping by females because of their heightened affective sensitivity, including empathy (Kilham & Mann, 1974) and guilt (Wallington, 1973). These findings are backed by Schwartz and Clausen's (1970) finding that females were more likely to intervene in an epileptic seizure when they were alone and by Bleda, Bleda, Byrne, and White (1976), who reported that females alone were more likely to report cheating when an identifiable victim was hurt by the dishonesty. Unfortunately, these studies shed little light on the role of sexual configuration. For example, Austin and McGinn (1977) studied only same-sex relationships, and Bleda et al. varied sex of bystander and wrongdoer, but not sex of victim. In addition to this qualification, many other studies have shown either no sex differences or greater helping by males. Latané and Darley (1970), Moriarity (1975), and Shaffer, Rogel, and Hendrick (1975) reported no sex differences in the willingness of bystanders to intervene in a theft. I. M. Piliavin, Rodin, and Piliavin (1969) reported that females were less likely to aid a person who collapsed in subway in a group setting,[1] and Gelfand, Hartmann, Walder, and Page (1973) found males more likely to report shoplifting behavior. However, once again, these studies did not fully study the sexual configuration, and the operational settings varied considerably. For example, Moriarity manipulated sex of thief, but victim/thief configuration was always cross-sex and sex of bystander was not taken into account; Shaffer et al. manipulated sex of victim and thief together, but the configuration was always cross-sex for all participants.

On the basis of previous research, predicting a Sexual Configuration × Degree of Harm interaction is also problematic. My initial thought that females would be victim oriented might suggest a high rate of helping regardless of degree of harm and might suggest that a majority of males would help only under substantial harm to the victim (because high harm might imply higher costs for not helping). These propositions are consistent with the speculations of Gelfand et al. (1973) that females are less inclined to report shoplifting because the victim is an impersonal business establishment, which thus does not activate their heightened affective sensitivity. Similar findings have been reported by Bleda et al. (1976), who found higher prosocial action by females only when there was an identifiable victim, and by Schopler and Bateson (1965), who explained a somewhat higher rate of female helping in terms of an apparent greater concern by males with their own outcomes.

This tentative prediction of greater overall helping by females, which should be most apparent when the degree of harm to a victim is modest, must again be held in abeyance because the supporting research did not examine sexual configurations.

Prior to conducting a field experiment that fully examined the effect of degree of harm and sexual configuration among participants, one pretest and two pilot studies were conducted. The results of these preliminary data-gathering efforts are briefly described below. All studies used identical procedures. Only the experimental designs differed.

General Procedure for All Studies

All subjects were college students at a large southern university. Subjects in the pretest, Pilot 1, and the field experiment attended the

[1] I. M. Piliavin et al. (1969) explained their results in terms of fewer costs for females for not helping because of their *role*. These researchers conjectured that the female role prescribes less initiative in emergencies and absolves women of the worst of public censure for not helping. This explanation is interesting in that it demonstrates the close relationship between the concepts of cost and responsibility in helping situations: Not helping is less costly to females because they are not responsible; they are not responsible because of their role (see Hamilton, 1978, on the importance of role in attributions of responsibility). Unfortunately, the data reported subsequently are exactly opposite to these conjectures.

same university; subjects in Pilot 2 attended a different university in the same state.

Subjects were located by selecting college students who were relaxing (usually reading) and seated by themselves in a corridor of a large university classroom building. One experimenter served as the potential victim and approached the bystander subject. He or she first gained a personal commitment from the bystander to watch his or her personal belongings. The experimenter paused directly in front of the bystander and asked, "Do you mind watching my things? I'll be back in a couple of minutes." All subjects consented in all studies. The experimenter then set several items next to and slightly forward of the bystander. This made it almost impossible for the subject not to notice what the items were and the theft that was about to take place.

The first experimenter (i.e., the victim) then moved off to a distant location to record the subject's response to the forthcoming crime. (Only a few subjects watched the first experimenter leave.) A second experimenter, who served as the thief, then casually approached the bystander and paused to examine the articles placed next to him or her. The thief passed directly in front of the subject and stopped for approximately 10 sec (as if to convey indecision as to his or her impending deed) before picking up the item. The theft was designed to be conspicuous.

The dependent variable was percentage of subjects in each condition who intervened to stop the theft. After intervening, or failing to intervene, subjects were told they had just been in an experiment. All subjects were unaware that this was the case. Subjects were then orally administered an open-ended questionnaire. The questions differed slightly for subjects who helped and those who failed to help. The questions assessed whether subjects noticed the theft, when they first knew a theft was going to occur, their first reaction to it, why they did not stop the thief, why the commitment to safeguard the item was not honored, and how they felt about the experimenter. The purpose of the study was then explained to subjects, and permission to use their data was obtained.

Pretest: Effect of Prior Commitment

In planning this research I first looked at the findings reported by Moriarity (1975) and Shaffer et al. (1975) on the importance of prior commitment for bystanders who observe a theft. Using different settings (public beach, Laundromat, or library) but stolen items of comparable value (portable radio or watch), Moriarity reported an intervention rate of 90%–100% for committed bystanders and 0%–20% for the uncommitted; Shaffer et al. reported a comparable 70% and 10% for the committed and uncommitted, respectively.

These polarized rates of intervention as a function of commitment impressed me so much that I first attempted to confirm whether they were a reliable baseline for reactions to a theft. Because I hold a rather humanistic view of people I wanted to take another look at the reactions of the uncommitted bystanders to see if they were really so passive in the face of a crime. By using the procedure described above, 20 bystander subjects (10 male, 10 female) were located and exposed to the theft of an electronic calculator—an item of considerable value (high harm to victim) and comparable with the items stolen in the Moriarity (1975) and Shaffer et al. (1975) studies. No prior commitment was obtained from these subjects. The victim was always female and the thief male. The victim stopped directly in front of the bystander and set a folder full of papers and a calculator down next to the bystander and then left, glancing back at her belongings. The results were dramatic. Not a single uncommitted bystander intervened to stop the theft. I interpreted this pretest data as corroborating previous research. Commitment appears to be a necessary condition for prosocial action in responses to the type of theft situation created here. Commitment, however, may not be necessary in all theft situations.

Pilot Study 1

This study was conducted as a group project for an undergraduate course. It allowed me to perfect the experimental procedure and provided a means for comparing the average

rate of helping of the subject population with data gathered in other theft studies. The item stolen was a calculator. The thief was always male (M) and the victim female (F). The sex of the bystander was varied (32 males, 36 females), making the bystander/thief/victim configuration M/M/F for half the subjects and F/M/F for the other half. Size of the thief was also varied. One half of the bystanders witnessed an extremely large male, and half viewed an average-sized male. This variable was designed to manipulate cost for helping.

I expected a Sex of Bystander × Size of Thief interaction in a 2 × 2 design. I assumed that risk of a physical confrontation was greater for males accusing other males of stealing. Thus, I reasoned fewer males would help when the thief was huge. My predictions were not supported. In a chi-square analysis no significant effects were found.[2] The rate of intervention was extremely high, ranging from 64.7% to 70.6%, and was fairly uniform. Females showed a slightly higher rate of helping, but it did not approach significance.

These null results suggested several empirical possibilities. First, when an initial commitment to help exists and the harm to the victim is substantial, then bystanders focus on costs for not helping, ignore the risk of physical danger for helping (e.g., size of thief), and show a high rate of helping. Perhaps if the risks of physical danger were more apparent, such as the presence of a gun, one could expect different results. However, the similarity of the rate of helping found here to the rates found in other theft studies (Latané & Darley, 1970; Moriarity, 1975; Shaffer et al., 1975) suggests that the present findings are reliable, though constrained within the limits of the apparent degree of physical danger that was apparent to the bystanders. Second, no sex differences were apparent.

Pilot Study 2

This study broke further ground for the field experiment. In a 3 × 2 design I examined the effect of harm to victim and sex of thief on helping. Degree of harm was operationalized by the value of the items stolen. The bogus victim placed books and folders (low harm), books, folders, and several record albums (moderate harm), or books, folders, and an electronic calculator (high harm) near the bystander. I predicted a significant main effect for harm, reasoning that commitment by itself may not be sufficient to produce the high rate of helping found in Pilot 1. Perhaps substantial harm to the victim is also necessary to activate the bystander's sense of personal responsibility (Schwartz, 1977) because of the potentially high costs for not helping. If this interpolation is correct, then low harm to the victim should beget a low rate of helping and high harm should elicit a high rate of helping.

This study was exploratory with respect to sex differences. I speculated that it would be more costly for a male to intervene with another male because of the physical confrontation, which I assumed would imply a competitive definition of the situation between males, but would be neutralized, or at least attenuated, in a cross-sex confrontation. If this assumption is correct, then one should observe males helping less when the thief is male and harm to victim is low. In this pilot I varied sex of thief, but used male bystanders only ($N = 72$). Once again the victim was always female, which produced M/M/F and M/F/F sexual configurations of bystander/thief/victim, though it was obvious that the results could only be analyzed in terms of sex of thief.

As Table 1 shows, the pilot data confirmed my prediction for the effect of degree of harm. As was the case in Pilot 1, male bystanders showed a high rate of helping (83.3%) when harm was high, but few bystanders helped when harm to victim was low (16.7%) or moderate (29.2%). This difference is reflected

[2] The frequency data for all studies were analyzed with a computer program (*Everyman's Contingency Table Analysis*, Note 1) that implements the log-linear methods for analyzing dichotomous data pioneered by Goodman (1972). It partitions explained variance in a manner similar to a squared multiple correlation regression analysis and yields analogues of main and interaction effects in the analysis of variance. I thank Thomas Guterbock for his assistance in making this program available.

Table 1
Effect of Degree of Harm to Victim and Sex of Thief on Male Bystanders' Helping Responses (Pilot 2)

Sex of thief	Degree of harm			Total
	Low	Moderate	High	
Male	2/12[a]	3/12	10/12	15/36
%	16.7	25.0	83.3	41.7
Female	2/12	4/12	10/12	16/36
%	16.7	33.3	83.3	44.4
Total	4/24	7/24	20/24	
%	16.7	29.2	83.3	

[a] The numerator denotes the number of bystanders who intervened; the denominator denotes the total number of bystanders.

in a significant main effect for harm, $\chi^2(2) = 26.19$, $p < .001$, although it is clear from Table 1 that substantial helping occurred only in high harm to victim conditions.

The overall finding of Pilot 2 is consistent with my reasoning that commitment to help by itself is not sufficient to activate prosocial action. In this case degree of harm was an important situation-defining factor. One could argue that substantial harm caused an increase in bystanders' sense of personal responsibility and their perception of costs for not helping.

Pilot 2 also found no sex differences. My speculation as to qualitative differences in same-sex and cross-sex confrontations was disconfirmed. This finding, when paired with Pilot 1 results, was discouraging as to the possibility of sex differences, given that the high/high cost situation combined with manipulation of harm to victim seemed to be well suited for discovering such differences if they indeed existed (e.g., J. A. Piliavin & Piliavin, 1972).

Field Experiment

Pilot 1 manipulated sex of bystander and found no significant effect. Pilot 2 found male bystanders to be unaffected by the sex of thief by itself or in combination with degree of harm to the victim. These results were perplexing. I thought I had created ideal conditions for the expression of sex differences. Subjects were alone; they were personally committed to assuming responsibility; they faced a personal paradox of high costs both for helping and for not helping; and they were exposed to different strengths of a potent situation-defining variable, which should have affected the intensity of personal responsibility they experienced and tipped the high/high cost balance in different directions.

It is clear, however, that the pilot studies were at best superficial tests of the role of sex in helping. Drawing on Austin and McGinn (1977), who suggested that males and females hold different motivational orientations toward others when they are responsible for another person's outcome, I designed a final field experiment, which provided a truer social psychological test for sex differences by manipulating the total sexual configurations among participants and degree of harm to victim.

I assumed females would be more victim oriented and concerned with interpersonal harmony under conditions of personal commitment and high/high cost. I reasoned that this possible "nomothetic predisposition" might translate into a higher rate of helping when harm to the victim was low. Pilot 2 showed that for males commitment by itself was not enough to produce help. Costs for intervening apparently outweighed costs for not helping when harm to the victim was low; males were willing to sacrifice the victim's belongings when their value was not high even when a commitment had been made. Females may not be so inclined. If they are indeed more victim oriented, then one should expect at least a moderate amount of help in the low-harm conditions.

Method

All eight possible bystander/thief/victim sexual configurations were employed along with low (folders and books) and high (folders and calculator) degree of harm conditions. The experimental design can be alternatively described as an 8×2 design (if sexual configuration is considered a unitary variable) or a $2 \times 2 \times 2 \times 2$ one (if the effect of sex of each participant is analyzed separately). Subjects were 352 male and female college students ($n = 22$ per cell). This design is capable of yielding a complete estimate not only of differences in bystanders' helping responses as a function of the

strength of a situation variable (i.e., harm) but also of separate effects for the role of sex of victim and thief.

Results

Table 2 shows that my tentatively held proposition that female bystanders would show greater overall helping and that this would be most apparent in low harm to victim conditions is correct. This observation is confirmed by log-linear regression analyses that uncovered a number of significant effects.

First, a significant main effect for degree of harm to victim, $\chi^2(1) = 31.13$, $p < .001$, reproduces the finding of Pilot 2 on the important situation-defining function of this variable. Table 2 shows greater helping in all of the high-harm conditions within each sexual configuration. In overall terms, 45.5% of bystanders intervened when harm was low and 75.6% when harm was high. Second, analyzing the helping data as a 2×8 design produces a highly significant main effect for sexual configuration, $\chi^2(7) = 32.85$, $p < .001$. However, this result only means that *some* aspect of the total sexual configuration among participants made a difference. Looking at the separate effects due to the sex of the individual participants in a $2 \times 2 \times 2 \times 2$ partitioning of the design produced a significant main effect for sex of bystander, $\chi^2(1) = 26.86$, $p < .001$, a marginally significant main effect for sex of the victim, $\chi^2(1) = 3.63$, $p < .07$, and a Sex of Bystander $\times$ Victim interaction, $\chi^2(2) = 30.90$, $p < .001$. Sex of thief had no effect on helping, $\chi^2(1) = .16$, *ns*. These findings are easily summarized by percentages. Collapsing across harm, 73.9% of female bystanders intervened, whereas only 47.2% of males did so; 65.9% of the female victims received help, but only 55.1% of the male victims were as fortunate. The Sex of Bystander $\times$ Victim interaction reflects the especially high number of females helped by other females: 80.7% of female bystanders helped female victims, and 51.1% of male bystanders helped females; 67.0% of female bystanders helped males in need, and 43.2% of male bystanders helped another male.

The third and most important finding of the field experiment is the highly significant

Table 2
Effect of Sexual Configuration Among Participants and Degree of Harm to Victim on Helping (Field Experiment)

Sex of bystander/ thief/victim	Degree of harm		Total
	Low	High	
M/M/M	4	14	18/44
%	18.2	63.6	40.9
F/M/M	13	18	31/44
%	59.1	81.8	70.5
M/F/M	4	16	20/44
%	18.2	72.7	45.5
F/F/M	12	16	28/44
%	54.5	72.7	63.6
M/M/F	8	15	23/44
%	36.3	68.2	52.3
F/M/F	17	20	37/44
%	77.3	90.1	84.1
M/F/F	6	16	22/44
%	27.3	72.7	50.0
F/F/F	16	18	34/44
%	72.7	81.8	77.3
Total	80/176	133/176	
%	45.5	75.6	

Note. M = male; F = female; $n = 22$ per cell.

Harm to Victim $\times$ Sexual Configuration interaction, $\chi^2(8) = 68.24$, $p < .001$. I had tentatively proposed (based on one related study and a literature review) that female bystanders show a much higher rate of helping in low harm to victim conditions. This hypothesis was confirmed. I found, as I had in Pilot 2, that male bystanders helped relatively infrequently in the low-harm condition (25.0%) but substantially more often in the high-harm condition (69.3%). The most impressive finding was the substantial rate of helping by female bystanders in both low- (65.9%) and high-harm (81.8%) conditions. The rate of helping by females in low-harm conditions fell just short of the proportion of males helping in high-harm conditions. Breaking this overall interaction effect into component parts produced significant Sex of Bystander $\times$ Harm, $\chi^2(2) = 61.04$, $p < .001$, and Sex of Victim $\times$ Harm, $\chi^2(2) = 35.49$, $p < .001$, interactions. Based on the significant Sex of Bystander $\times$ Sex of Victim interaction (reported earlier as part of the sexual configuration main effect) and the absence of any significant effect due to sex of thief, the

model that best fits these data (and therefore the most important empirical finding of this study) is a Sex of Bystander × Sex of Victim × Harm interaction, $\chi^2(3) = 65.92$, $p < .001$.

Interview responses. The postexperimental interview was open ended in character. It attempted to ascertain how the situation was defined by all subjects, those who helped and those who did not, and which, if any, justifications were used by nonhelpers. A similar pattern of responses emerged in each of the three studies. Bystanders' responses in the field experiment are summarized here in full appreciation of the tenuous nature of post hoc interview data.

All bystanders noticed the theft. Among bystanders who intervened (e.g., 213, or 60.5%), 170, or 79.8%, indicated that they noticed the theft as soon as the subject picked up the stolen object; the remainder did so as the thief walked away. A substantial number of the bystanders who helped indicated some hesitancy to intervene (97, or 45.5%), most often saying they thought the thief might really be the true owner of the objects (39, or 40.2%). More helpers indicated they "were aware of the potentially negative consequences" of not helping (103, or 48.4%) than of helping (74, or 34.7%). Some bystanders said both types of costs were important to them (62, or 29.1%). The remainder did not believe they considered either type of risk (36, or 16.9%). Among female helpers (130), 82, or 63.1%, mentioned costs for not helping; 27, or 20.8%, mentioned costs for helping; and 44, or 33.8%, mentioned both. Among male helpers (83), 47, or 56.6%, mentioned costs for not helping; 21, or 25.3%, mentioned costs for helping; and 18, or 21.7%, mentioned both. Thus, costs for not helping seemed to dominate responses among helpers, though the costs of intervention could not be ignored. This pattern suggests that the high/high cost tension was tipped in favor of costs for not helping among helpers. This finding is demonstrated by the remarks of one bystander: "What a fool I'd look like if she came back and somebody had ripped off her stuff. I know if I was in her place I would have been really steamed. And I said I'd watch her things anyways." A number of helpers (23, or 10.8%) said they would not have intervened if the stolen object had been of lower value (e.g., harm to victim). Finally, only a few helpers (12, or 5.6%) indicated they would have intervened without a prior commitment to guard the stolen object.

Among bystanders who failed to intervene (139, or 39.5%), a very substantial number indicated they "thought about saying something to the thief" (97, or 69.8%). One of these bystanders remarked; "What a relief! It's only an experiment!" Although a majority of the inactive bystanders were thus aware that a theft was quite likely taking place, only a handful (14, or 10.1%) failed to provide a justification for their inaction. The three most common justifications voiced by subjects were that they thought the thief was a friend of the victim (38, or 30.4%), the thief was the true owner of the objects (45, or 36.0%), or the victim sent the thief to retrieve the items (43, or 34.4%). Ninety-three male bystanders did not help; 15, or 16.1%, mentioned costs for not helping; 37, or 39.8%, mentioned costs for helping; 9, or 8.6%, mentioned both; and 41, or 44.1%, said both types of risk were unimportant to them. Forty-six female bystanders failed to intervene; 15, or 32.6%, mentioned costs for not helping; 12, or 26.1%, mentioned costs for helping; 2, or 4.4%, mentioned both; and 19, or 41.3%, indicated neither type of risk was a significant factor. Several patterns are clear from responses to these questions. First, both types of costs were seldom mentioned together, which was not the case with those who helped. Second, male nonhelpers indicated costs for helping were most important, and female nonhelpers mentioned costs for not helping slightly more often than males. Third, a substantial portion of nonhelpers of both sexes indicated risk was not a factor. This may be because of their awkward position of trying to justify a blameworthy action. Finally, a large number of bystanders (81, or 58.3%) said they would have intervened if the stolen items had been of higher value, which again probably reflects a need to justify blameworthy action. It should be kept in

mind that these patterns pertain only to the type of situation created in these studies.

Discussion

This research sheds some light on sex differences in bystanders' responses to a theft and on how sex interacts with situational variables. Researchers may find the postulated different motivational sets of males and females to be a plausible interpretation of these data. Since the experimenters were aware of the experimental conditions, there is a danger of experimental bias. However, the rates of helping compare favorably with the findings of other theft studies that used items of similar value under conditions of personal commitment (Moriarity, 1975; Shaffer et al., 1975). This comparability seems to argue against experimenter bias. The data reported here in combination with these previous theft studies also appear to provide an ecologically robust estimate of the rate of intervention to be expected in a theft under conditions of personal commitment, when the item is of reasonable value but the situation does not appear life threatening (e.g., when the types of potential costs involved are probably psychological rather than physical in nature).

Summary of Findings

1. When harm to the victim was substantial (theft of calculator), bystanders of both sexes showed an extremely high frequency of intervention (Pilot 1 and field experiment). This finding may be due to an increase in the salience of costs of not helping in the high-harm situation.

2. Degree of harm to victim by itself produced a significant increase in intervention to stop a theft (Pilot 2 and field experiment). This effect adds further credence to the interpretation that high harm added to personal commitment produces an increase in personal responsibility (because of the salience of costs for not helping) that is sufficient to guarantee a high intervention rate.

3. The most important result was the sex differences found when the total sexual configuration among participants was taken into account. The significant overall Harm × Sexual Configuration interaction demonstrates the utility of the interactionist approach to prosocial action recommended by Staub (1978) and Schwartz (1977). Sexual configurations exerted differential effects as a function of the strength of the situation-defining variable of degree of harm. I found that sex of thief made no difference in bystanders' behavior, but sex of bystander and sex of victim were influential by themselves, in combination with each other, and in combination with harm to victim. The Sex of Bystander × Harm interaction constitutes the most parsimonious fit to these data and accounts for a very large portion of the explained variance, $\chi^2(2) = 61.04$, $p < .001$, where the overall test produced $\chi^2(15) = 74.95$, $p < .001$, and the test for an overall Harm × Sexual Configuration produced $\chi^2(8) = 68.24$, $p < .001$. However, the theoretical significance of the greater helping elicited by female victims, alone and in combination with other variables, leads me to conclude that the three-way interaction between sex of bystander, sex of victim, and harm to victim, $\chi^2(3) = 65.92$, $p < .001$, best summarizes the empirical findings of this research.

Reference Note

1. *Everyman's contingency table analysis.* Chicago: University of Chicago, Department of Statistics.

References

Ashton, N. D., & Severy, L. J. Arousal and costs in bystander intervention. *Personality and Social Psychology Bulletin*, 1976, *2*, 268–272.

Austin, W., & McGinn, N. C. Sex differences in choice of distribution rules. *Journal of Personality*, 1977, *45*, 379–394.

Berkowitz, L., & Daniels, L. R. Responsibility and dependency. *Journal of Abnormal and Social Psychology*, 1963, *66*, 429–436.

Berkowitz, L., & Daniels, L. R. Affecting the salience of the social responsibility norm. *Journal of Abnormal and Social Psychology*, 1964, *68*, 302–306.

Bleda, P. R., Bleda, S. E., Byrne, D., & White, L. A. When a bystander becomes an accomplice: Situational determinants of reactions to dishonesty. *Journal of Experimental Social Psychology*, 1976, *12*, 9–25.

Clark, R. D., & Word, L. E. Where is the apathetic bystander? Situational characteristics of the emer-

gency. *Journal of Personality and Social Psychology*, 1974, *29*, 279–288.

Gelfand, D. M., Hartmann, D. P., Walder, P., & Page, B. Who reports shoplifters? A field-experimental study. *Journal of Personality and Social Psychology*, 1973, *25*, 276–281.

Goodman, L. A. A modified regression approach to the analysis of dichotomous variables. *American Sociological Review*, 1972, *37*, 28–46.

Hamilton, V. L. Who is responsible? Toward a social psychology of responsibility attribution. *Social Psychology*, 1978, *41*, 316–328.

Kilham, W., & Mann, L. Level of destructive obedience as a function of transmitter and executant roles in the Milgram obedience paradigm. *Journal of Personality and Social Psychology*, 1974, *29*, 696–702.

Latané, B., & Darley, J. M. *The unresponsive bystander: Why doesn't he help?* New York: Appleton-Century-Crofts, 1970.

Moriarity, T. Crime, commitment, and the unresponsive bystander: Two field experiments. *Journal of Personality and Social Psychology*, 1975, *31*, 370–376.

Piliavin, I. M., Rodin, J., & Piliavin, J. A. Good Samaritanism: An underground phenomenon? *Journal of Personality and Social Psychology*, 1969, *13*, 289–299.

Piliavin, J. A., & Piliavin, I. Effect of blood on reactions to a victim. *Journal of Personality and Social Psychology*, 1972, *23*, 353–361.

Schopler, J., & Bateson, N. The power of dependence. *Journal of Personality and Social Psychology*, 1965, *2*, 247–254.

Schwartz, S. H. Elicitation of moral obligation and self-sacrificing behavior: An experimental study of volunteering to be a bone marrow donor. *Journal of Personality and Social Psychology*, 1970, *15*, 283–293.

Schwartz, S. H. Normative influences on altruism. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 10). New York: Academic Press, 1977.

Schwartz, S. H., & Clausen, G. T. Responsibility, norms, and helping in an emergency. *Journal of Personality and Social Psychology*, 1970, *16*, 299–310.

Shaffer, D. R., Rogel, M., & Hendrick, C. Intervention in the library: The effect of increased responsibility on bystanders' willingness to prevent a theft. *Journal of Applied Social Psychology*, 1975, *5*, 303–319.

Staub, E. *Positive social behavior: Social and personal influences* (Vol. 1). New York: Academic Press, 1978.

Wallington, S. A. Consequences of transgression: Self-punishment and depression. *Journal of Personality and Social Psychology*, 1973, *28*, 1–7.

Walster, E., & Piliavin, J. A. Equity and the innocent bystander. *Journal of Social Issues*, 1972, *28*, 165–189.

Received February 28, 1979 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 11, 2121-2136

The Effects of Systematic Variations in Information on Judges' Descriptions of Personality

Daniel S. Weiss

Institute of Personality Assessment and Research
University of California, Berkeley

The use of observers in personality assessment has recently been subjected to intensive scrutiny and criticism by several different commentators on the field. These criticisms have called into question some of the fundamental assumptions of personality assessment, and the evidence presented has been used to dismiss the findings that result from the use of human observers. The present study presents results demonstrating that observers can and do provide personality descriptions that reliably reflect variations in available information. Six different sets of information about 82 persons were furnished to a panel of 52 judges whose descriptions were then based on only one of the six information sets. The results of a series of analyses suggested that agreement among judges is not related to the amount of information used to formulate descriptions. Nonetheless, judges' descriptions did show a relationship between information and variation of specific descriptors, in that variation decreases in conditions where information is minimal. Finally, it was found that interperson similarity of description was associated with information, decreasing when judgments were based on personal contact and/or interviews and increasing when based only on stereotypic data.

Strong warnings about the continued use of observers in personality assessment have recently been sounded (e.g., Berman & Kenny, 1976; Bourne, 1977; Chapman & Chapman, 1967, 1969; D'Andrade, 1965, 1974; Fiske, 1971, 1973, 1974, 1976; Mischel, 1968, 1973; Shweder, 1975, 1977). The essential criticism has been that in making personality assessments, the observer is not sensitive to stimuli or, if sensitive, is unable to respond in ways that demonstrate such sensitivity. Further, critics claim that judges are unable to attend to or to report validly the actual relationships among the behaviors on which descriptions are based. Instead, implicit personality theory and linguistic accounts of personality are offered to explain the responses of personality assessors. If these criticisms are accepted, personality assessment methodology and the host of empirical findings dependent on these methods are suspect.

The most frequently cited evidence that observers' descriptions are invalid is the set of studies showing that despite differences in information about ratees, observers' trait ratings produce extremely similar factor structures of trait terms. Frequently referenced is the study by Passini and Norman (1966), in which undergraduate subjects were asked to rate each other on a set of traits previously selected and factorially grouped by Norman (1963). The members of this group of undergraduate students had had no previous contact or acquaintance with each other but did have several minutes together before the rating task began. During this period facts about

This study is based on a dissertation submitted in partial fulfillment of the requirements for the PhD degree at the University of California, Berkeley. The author acknowledges and appreciates the helpful suggestions and comments of Gerald Mendelsohn and Jack Block.

Requests for reprints should be sent to Daniel S. Weiss, 401 Parnassus Avenue, University of California, San Francisco, California 94143.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3711-2121$00.75

demeanor, dress, and bearing could be observed. Ratings were factor analyzed, and the resultant factor structure was compared to Norman's factor structure based on ratings where raters were more familiar with the persons being rated. The two factor structures were similar, leading Passini and Norman to conclude that *these* data were more a function of the rater than of the ratee. That is, judges in this study, who could not have known the characteristics of their targets, apparently used as a base their own implicit personality theories when they described their targets.

Using this basic approach, other researchers have shown similarity of factor structures across information conditions and have likewise concluded that personality ratings are invalid (e.g., Shweder, 1975). These conclusions, however, have been based on analyses of the relationships among stimuli, rather than on the ratings themselves. With respect to the wide citation and uncritical acceptance of the Passini and Norman study, a second paper that substantially moderates the Passini and Norman conclusions does exist and should be acknowledged. A curious aspect of this part of the controversy, however, is that this second relevant article, in the same journal, published during the same year and coauthored by one of the two authors of the Passini and Norman paper, is less frequently noted. This article (Norman & Goldberg, 1966) requires that the conclusions of the Passini and Norman study be severely delimited.

In this follow-up article, Norman and Goldberg further analyzed the Passini and Norman data and other data and found impressive evidence that the level of interrater agreement was a function of the degree of acquaintance. This finding was important because Norman and Goldberg recognized that the level of interrater agreement is relevant to the claim that implicit personality theories were the basis for the raters' descriptions. If the variance of ratings across ratees is in excess of that expected by chance, then this evidence can be taken to "imply that the raters are to some degree responding in like fashion to *some characteristics* of the ratees" (p. 684). When the reliabilities of both the composite and component ratings were calculated for Monte Carlo data, the original Passini and Norman data, a Peace Corps sample, and a fraternity senior sample, a correspondence was obtained between amount of reliability and amount of previous intimacy and knowledge of ratees. Noting that these findings provided necessary but not sufficient evidence for the validity of the rating data, Norman and Goldberg then demonstrated that with other convergent validational information (self-ratings and predicted peer ratings) in the Peace Corps and Passini and Norman sample, the validity diagonals of the Campbell and Fiske multitrait multimethod matrix yielded coefficients that lent "support not simply [to] the inference of differential ratee relevance in the data from the two groups, but also differences in the pertinence these data have to the personality characteristics of the ratees in the two cases" (p. 690). Given these data, the mere finding of similarity of factor structures of trait rating terms across samples of differing acquaintance cannot be taken as evidence of the insensitivity of the ratings to the ratees.

Wiggins and Blackburn (1976) presented data quite similar to that of Norman and Goldberg's. As part of their investigation of implicit personality theories, they too found that the factor structure of the traits was not different as a function of knowledge of the target being rated (in this case, stranger vs. friends). It was also shown that individual differences existed in the use of implicit personality theories, but the data did not allow the correlates of such differences to be clearly specified.

Despite the findings of similarity of factor structures, there are studies suggesting that the effects of implicit personality theories are not overwhelming. Block (1971), for example, has shown that a panel of judges' descriptions of a set of targets are more similar to each other within targets across judges than they are similar to descriptions across targets within judges. That is, Judges A, B, and C agree more about Target X than Judge A, B, or C agrees with himself or herself about targets X, Y, and Z. Nonetheless, there are many reasons why some research has found some personality judgments wanting.

A fundamentally important variable in per-

sonality rating studies is the amount and kind of information that judges have available. Thus, when comparing several different research studies, commentators have often lumped together types of judgment tasks that differ in the amount of information provided to the judges. Generalizations about the sensitivity of observers are inappropriate if it can be shown that the amount of information available to judges affects their judgmental processes and responses. Thus, the conclusions about the competencies of observers may need modification in the light of the limits imposed on judges by restrictions in the amount and kind of information available to them. Most prior work has not directly examined the variations in judgment due to variation in the amount and kind of information provided to judges. Studies that have examined information as a variable have generally cast results in terms of the validities of the judgments, despite the fact that the variables serving as criteria were often little better than the variables serving as predictors (see e.g., Golden, 1964; Kostlan, 1954; Sines, 1959). On balance, these studies did suggest differences in the validity of the judgments as a function of the amount and kind of information on which a judge based his or her predictions. It may also be noted that the conclusions of these studies were not based on analyses of the structure of personality ratings.

A summary of the research reviewed up to this point suggests several strong and clear conclusions. First, the robustness of trait rating factor structures to changes in information seems beyond dispute, especially when the set of trait terms employed has been previously selected on factorially derived grounds. Second, the robustness of such factor structures provides only equivocal evidence for conclusions about (a) the process used to make ratings or (b) the validity of the ratings that eventually produce the factor structure. Third, the modifications of personality ratings produced by variations in information have generally gone uninvestigated. One may assume however, that if the data base contains very little information, as have the data bases used by some critics, then raters have very little on which to base their ratings other than implicit personality conceptions. Thus, conclusions about the sensitivity of raters are unwarranted when based only on data gathered under meager informational conditions. If, however, amount and kind of information are systematically varied, and the ratings themselves, rather than their structure, are examined, then researchers have the opportunity to evaluate unbiased and relevant evidence of the sensitivity of observers of personality. The present study was undertaken to provide such evidence: The amount and kind of information given to observers was systematically varied, and the effects of information variation on personality descriptions were evaluated.

Method

The data used in this research derived from two sources. The first category contained data in the archives of the Institute of Personality Assessment and Research (IPAR). These data were personality descriptions of assessees made by staff members during a series of 1-day assessments held at the Institute. The second category contained data collected explicitly for this study. The collection of these new data required the use of other archival materials housed at IPAR,[1] namely the biographical records and interview protocols that new judges used to formulate personality descriptions of the assessees.

Subjects

One of the assessment projects carried out at IPAR was a study of attitudes and behaviors in the domain of population psychology (see Gough, 1973, 1975, for representative findings). A heterogeneous group of 402 persons living in communities surrounding the San Francisco Bay area was studied. Of this group, 82 (41 couples) came to the Institute during eight different 1-day assessment programs in 1974, 1975, and 1976. This group represented about 15% of all those to whom invitations to participate were made. All but one couple were married. The mean age for the 41 males was 31.12 ($SD = 5.26$), and for the 41 females was 29.20 ($SD = 4.51$). Their educational level in years was 13.73 ($SD = 1.80$) and 13.49 ($SD = 1.75$) respectively. The Hollingshead social class of the husband was 3.15 ($SD = .99$), where 1 is upper class and 5 is lower class.

---

[1] The author wishes to thank the Institute's director, Harrison G. Gough, for graciously making these data available.

Procedure

During the assessment day, each assessee was interviewed by two different staff members. One interview focused on the assessee's life history; the other, on the assessee's personal values with questions directed toward, but not limited to, issues in the realm of family and larger life goal orientations. Additional assessment procedures included a leaderless group discussion, an improvisational role-playing procedure, a couple interaction procedure, luncheon, a cocktail hour, a sit-down dinner, and finally a modified form of the party game charades. All these additional procedures were observed by all staff members, among whom were the individuals who conducted the interviews.

After the assessments were concluded, personality descriptions of each assessee were completed by the staff. The data utilized in this study came from two response instruments, the Adjective Check List (ACL; Gough & Heilbrun, 1965) and the California Q-sort (Block, 1961). Descriptions of each assessee were made on the ACL and Q-sort by both interviewers and by two or three other staff members who had previously been assigned to record their understanding of that assessee. These descriptions formed the materials in the archival data set. Thus, for any assessee, there were ACLs and Q-sorts completed by (a) a life history interviewer (for 2 of the 82 assessees, these data were missing), (b) a values interviewer (for 5 of the 82 assessees, these data were missing), and (c) two (for 30 of the 82 assessees) or three (for 52 assessees) staff members who had not interviewed that assessee but who had observed him or her during the assessment day (noninterview staff). Neither the life history nor the values interviewer was blind to the rest of the assessment procedures. Each was free to, and presumably did, draw upon information gleaned from the other assessment procedures to compile his or her ACL and Q-sort portraits. On the other hand, the noninterview staff did not have access to either interview protocol.

To extend the range of variation of amount and kind of information further, a second phase of data collection was undertaken. In this phase, three additional conditions or sources of information for judges' descriptions were provided. For each of these three conditions, one male and one female judge completed ACL and Q-sort materials for each case.

The first variant of what will be called the "protocol" data set was a *life history protocol condition*. For each case, the written life history interview protocol was read in its entirety by each of the two judges assigned to that case. In addition, demographic data (to be described below) were also given to the judges. Independently of each other, both judges then completed the ACL and the Q-sort for the case.

The second condition of the interview protocol data set was a *values protocol condition*. The same procedure used for the life history protocols was followed for the values protocols. One values interview from a male assessee in 1974 and one from a female assessee in 1976 were missing, reducing the n for the analyses on the values protocols to 80. The interview protocols for both the life history and values conditions consisted of handwritten records of the subjects' responses to a standardized set of questions. The interviewer's records were as nearly verbatim as possible. The actual questions used may be found in Weiss (1978).

The final condition in the study will be referred to as the *stereotypic-demographic condition*. The information available to judges in this condition consisted of sex, number of children, age, education, race, occupation, place of birth, and type of community of origin (either rural, suburban, or urban). Information about religious preference, a standard stereotype variable, was not included because these data were not available for all cases. In this condition, then, judges had neither direct nor indirect personal contact with the assessees. Their responses can be based only on their conceptions of personality characteristics and differences associated with *classes* of individuals.

Judges

The persons who comprised the assessment staff in 1974, 1975, and 1976 were either senior PhD faculty, full-time research personnel, visiting scholars, or advanced graduate students in personality or clinical psychology. The logic of the design requires that judgments in the various conditions vary only as a function of information and not also in terms of the panels of judges contributing descriptions in those conditions. Thus, great effort was made to secure protocol and stereotype judges who were similar to the staff members in the assessments.

There were 45 assessment staff members whose ACL and Q-sort descriptions were used in the assessments. Of these 45 judges, 21 were still at Berkeley and available and willing to participate as judges in the new study. In addition to these 21 persons, 7 advanced graduate students who had not participated in the assessments were asked to serve in the new study. These 7 judges were comparable to the advanced graduate students who had participated in the earlier assessments. Thus there were 28 judges in all who did Q-sorts and ACLs in the protocol and stereotype portions of the study. A total of 52 judges was employed for the combined assessment, protocol, and stereotype portions of the study.

If no case had been missing, there would have been a total of 902 descriptions. Thus, 34 descriptions or 4% of the theoretically complete file of data were missing, a minute fraction. Contributions to the total of 868 descriptions made by any single judge ranged from a high of 59 cases (7%) to a low of 2 cases (.2%). No case was described more than once by any judge. The mean number of descriptions for the total pool of 52 judges was 16.7 ($SD = 12.9$), or approximately 2%. That is, on the average, each judge contributed about 2% of the judgments on which subsequent analyses were conducted. Thus,

any biasing effects of idiosyncratic judges should be minimal.

Analyses and Results

Three separate, but related, sets of analyses were performed, each evaluating a different facet of the effect of information on observers' descriptions of personality. If the kind of information made no difference to the judges, then the results found in any one condition should be essentially identical to those found in any other condition. That is, if the information condition is irrelevant, then the degree to which assessees are differentiated from each other should not vary from one condition to another. Likewise, the mean level and variation of interobserver agreement should not be related to the information condition. The analyses were designed to provide evidence contradicting the position that information condition does not affect personality descriptions. Further, they sought to show that the manner in which information does affect description can support the conclusion that judges are reliable and sensitive "transducers" of variations in information and that their personality assessments can faithfully reflect the characteristics of the information available to them.

Interobserver Agreement

Studies employing rating tasks or procedures should report estimates of the reliability of the ratings. Because the information available to judges varied in this study, one might expect the amount of interobserver agreement also to vary with condition, as Norman and Goldberg (1966) had found. Because judges provided descriptions of target persons on two descriptive devices, the opportunity to evaluate interobserver agreement using different response formats was present. Three interobserver indices of reliability were calculated.

The first index involved the Q-sort. For each pair of raters for every case, the correlation over the 100 Q items was obtained. For the protocol and stereotype conditions, this correlation indexed interobserver agreement. For the two interview conditions, where only one judge described each case, observer agreement was calculated between the two interviewers. For the noninterview condition, there were 10 male and 13 female assessees for whom only two judges contributed descriptions; the remaining cases were described by three judges. For the former cases the correlation between the pair of judges was used as the index of agreement; for the latter cases, the mean correlation of the three pairs of judges was employed (in this analysis, and every analysis reported henceforth, all calculations involving the averaging of correlation coefficients employed Fisher's r-to-z transformation).

The second index of agreement was calculated for the judges' responses on the ACL. A correlation coefficient across the 300 items comprising the ACL was computed for each pair or set of raters for each case in the same fashion as was employed for the Q-sort. The dichotomous response format of the ACL means that the coefficients produced are actually phi coefficients. This index was therefore termed the ACLPhi index.

The ACL does not require judges to make a judgment for every item; that is, judges are free to check as many or as few adjectives as they believe to be descriptive. Since there are strong individual differences in the number of adjectives a judge is likely to check, regardless of information condition, this format poses some problems, being based on a fourfold tally of a present versus absent dichotomy. To overcome this problem, a third index of observer agreement was defined. The ACL items can be scored for 24 scales, some of which are empirically derived dimensions (e.g., self-control) and others of which are rationally derived scales based on Murray's need-press system. In the scoring algorithm the total number of adjectives checked is used as a correction factor when the raw scale scores are converted to standard scores. The net effect of such a procedure is to remove the differences on the scale dimensions that are attributable only to the differences in the total number of items checked. Thus, a correlation between two judges over the 24 scale scores will not be unduly influenced by differences in the total number of items

Table 1
Analysis of Variance of Reliability Indices

Index and condition	*n*	*M*	*SD*	Range	*F*
Q-sort[a]					
Life history/values interviews	75	.41	.19	−.09–.66	—
Life history protocol	82	.38	.28	−.50–.76	—
Values protocol	80	.38	.27	−.42–.73	.41
Noninterview	82	.41	.21	.07–.73	—
Stereotypic-demographic	82	.39	.22	−.11–.70	—
ACLPhi[b]					
Life history/values interviews	75	.35	.20	−.27–.63	—
Life history protocol	82	.27	.17	−.06–.58	—
Values protocol	80	.24	.18	−.13–.55	4.87*
Noninterview	82	.31	.17	−.02–.64	—
Stereotypic-demographic	82	.26	.17	−.10–.59	—
ACLScale[c]					
Life history/values interviews	75	.58	.57	−.76–.94	—
Life history protocol	82	.53	.56	−.80–.97	—
Values protocol	80	.45	.48	−.70–.90	.74
Noninterview	82	.53	.48	−.49–.91	—
Stereotypic-demographic	82	.54	.46	−.53–.94	—

Note. Means are *z*-transformed values; values for range are raw correlations. All values are between *pairs* of judges. ACLScale = Adjective Check List Scale. ACLPhi = Adjective Check List Scale with phi coefficient.
[a] Correlations based on 100 items; index shows heterogeneity of variance. [b] Correlations based on 300 items. [c] Correlations based on 24 scales.
* p (4, 396) < .001.

checked. Descriptions of the target cases were converted to the 24 standard scale scores using the conversion formula appropriate for each sex. For those cases described by three noninterview judges, the average correlation was used as the index. This index was termed ACLScale.

The interobserver correlations provided the raw data for the first group of analyses. A one-way analysis of variance was performed to evaluate the differences in level of agreement across conditions, for each index. The results of these analyses are reported in Table 1.

Q-sort reliability results. The first finding of interest is that for the Q-sort, there are no significant differences in interobserver agreement as a function of information condition. This result is not in accord with the Norman and Goldberg (1966) study. In the present research, judges reported that they found some conditions of information to be distinctly more helpful than others. One therefore might expect to find the strength of agreement among the judges to be related to the kind of information provided. That this expectation was not confirmed can be attributed in part to the nature of the Q-sort and in part also to the nature of the stereotypic-demographic information. Further, one may well ask if the expected differences were in fact warranted. A more detailed examination of this question will be deferred until the discussion section.

Of note, however, is the finding that the accompanying tests for heterogeneity of variance were significant, indicating that differences in the spread of agreement across the different conditions did occur. Agreement in the assessment conditions was never lower than −.09, whereas for the protocol conditions agreement went as low as −.50. On the other hand, in the protocol conditions, coefficients as high as .76 were obtained. The highest value in the assessment condition was .66.

ACL reliability results. The one-way analysis of variance of the mean interobserver agreement for the ACLPhi index did yield a significant overall difference, although infor-

mation condition accounted for only 4% of the variance in reliability coefficients. The ordering of the mean levels of interobserver agreement for the ACLPhi coefficient was associated with the greater information in the assessment conditions. Nonetheless, the lack of strength of the results does not strongly demonstrate that information condition affects reliability of description.

The ACLScale index, like the Q-sort index, did not show significantly different means as a function of information condition. The mean level of agreement for the ACLScale index, however, is higher than that for the other two indices. This finding is not unexpected, since the standardization procedure reduced the effects of number checked.

Differential Usage of Descriptors by Conditions

The second set of analyses concerns the use of the individual items of the two instruments (Q-sort and ACL). For the Q-sort it will be shown that judges' descriptions of assessees are more differentiated at the item level when amount and variety of information increase. For the ACL, it will be shown that judges use more descriptors with more varied information.

Q items. This analysis concentrated on item variances within condition; that is, on the variance of assessees' scores on a particular item within a particular condition. The basic array of scores consisted of 600 variances; 100 items over six conditions.

The six standard deviations for each item were calculated using Re-Q'd [2] data. Then tests of homogeneity of variance across the six information conditions were conducted. The Bartlett-Box F test and Cochran's C method (Winer, 1971) were used. Twenty-one of the 100 items showed significant heterogeneity of variance ($p < .05$) according to both tests. Eighteen more items, making a total of 39, showed significant departures from homogeneity of variance on one or the other test.

It was expected that the two interview conditions would, on the average, show the greatest variance of item placement across the 82 assessees because these two conditions were the ones with the most abundant and varied information available to each judge; that is, when information is most varied, judges should be able to make the most differentiated responses and hence produce the largest variance. To evaluate this anticipation, an anlysis that examined the relationship between amount of variation and information condition was conducted. The data for this next analysis were derived from the 6×100 array of standard deviations. For Q Item 1, the six within-condition standard deviations were ranked from 1 (the smallest SD) to 6 (the largest SD). This ranking procedure was repeated for the remaining 99 items. Thus, each of the 100 items was assigned a profile of six ranks. In effect, each profile indicates the ordering of the six information conditions in terms of the degree of variation in judges' descriptions. An analysis of variance of ranks across the six information conditions (Winer, 1971) was performed and evaluated by the Friedman test for the significance of the difference of mean ranks. The results of this analysis are presented in Table 2.

According to the Friedman test, $\chi^2(5) = 84.35$, $p < .0001$. The information condition that shows the highest rank and thus the most differential usage of items is the life history interview condition; the values interview and life history protocol conditions follow. Also of note is that the stereotypic-demographic condition shows the smallest mean rank of any of the information conditions, a result fully in accord with the impoverished information in that condition.

For the sets of 21 and 39 items showing significance on both or one of the heterogeneity of variance tests, results were also significant, with $\chi^2(5) = 38.39$ and $\chi^2(5) = 71.89$, $p < .007$, respectively. Moreover, these two sets reveal an important point. The SD of the mean rank for the stereotypic-demographic condition is, for both the 21 and 39

[2] It will be recalled that the number of judges varied from one to three within information conditions. In those conditions where there was more than one judge, the data were Re-Q'd to standardize all Q-sort distributions. This was accomplished by summing across judges and recasting the array of 100 sums into the forced distribution of values used for an individual judge's sort.

Table 2
Means and Standard Deviations of Q-Item Variance Ranks

Condition	*M*	*SD*
All items (n = 100)		
Life history interview	4.40	1.48
Values interview	4.21	1.38
Life history protocol	3.65	1.65
Values protocol	3.32	1.66
Noninterview	3.12	1.71
Stereotypic-demographic	2.30	1.49
Items with both heterogeneity of variance tests significant (n = 21)		
Life history interview	5.00	1.30
Values interview	4.78	1.03
Life history protocol	3.57	1.43
Values protocol	2.71	1.71
Noninterview	3.33	1.71
Stereotypic-demographic	1.90	0.99
Items with one heterogeneity of variance test significant (n = 39)		
Life history interview	4.85	1.14
Values interview	4.89	1.21
Life history protocol	3.59	1.52
Values protocol	2.74	1.58
Noninterview	3.59	1.71
Stereotypic-demographic	1.74	0.91

item sets, the lowest *SD* of the six conditions. The next lowest *SD* is, in both cases, the *SD* of the mean rank for an interview condition. Thus, not only are these conditions extreme in terms of variation of item placements across assessees as a function of information condition (as evidenced by the highest and lowest means), but they are also more consistently extreme (as evidenced by the low *SD*s) than are the other conditions. This result does not necessarily follow from the other findings in the analysis and thus underscores the effect of information on the variation of item placement within condition.

Although it is evident that the assessees are most differentiated in the interview conditions and least differentiated in the stereotype condition, it remains to be shown that the type of items sorted with the most variation are items for which the particular condition provided relevant information. There are no statistical procedures for determining this fit; goodness of fit must be judged psychologically.

Tables 3 through 8 report the Q items that showed the greatest variance in each of the six information conditions. That is, Table 3 reports the items that showed their greatest variation in the life history interview condition, Table 4 reports those items that showed their greatest variation in the values interview condition, Table 5 reports on the life history protocol condition, Tables 6, 7, and 8 contain items from the values protocol, noninterview, and stereotypic-demographic conditions, respectively.

Several points of interest may be noted in Table 3. First, 27 of the 100 Q items revealed their greatest variation in the life history interview condition; this was the largest number observed for any condition. The item content appears to be appropriate for what may be obtained in a life history interview. Four clusters of items may be delineated in the set of 27. Several deal with behavioral style and tempo (Items 4, 17, 20, 26, 30, 31, 77). Four refer to matters of sexuality (Items 58, 73, 77, 80). Five refer to social and interpersonal relationships (Items 17, 48, 49, 54, 89). The final category contains 12 items that refer to temperament and character (Items 9, 30, 41, 60, 62, 68, 70, 75, 77, 86, 97, 99).

Table 4, containing items showing their largest variation in the values interview condition, has 22 entries. As anticipated, this is the second largest number. The item content overlaps somewhat with that found in the life history interview condition, but this similarity is not unexpected. Personal contact and the interview transaction were as much a part of the one condition as the other. Items in Table 4 can be classified into (a) the social and interpersonal sphere (Items 27, 35, 36, 37, 38, 61, 64, 92), (b) character structure (Items 14, 39, 40, 42, 45, 72, 74, 98), and (c) expressive behavior (Items 43, 92, 98). A fourth category, (d), distinct from those observed in the life history interview condition, is the expected set of items on personal values (Items 22, 84, 91).

Table 5 lists the 19 items showing their highest variation in the life history protocol condition. All items, with the exception of item

52, which refers to assertive behavior, can be accommodated in the categories postulated for the life history interview condition. An important difference is that in the present listing the expressive behavior category has no entries. This omission is reasonable, as judges in this condition did not see the interviewee.

Items with largest variance in the values protocol condition are presented in Table 6. As anticipated, items referring to personal values are included, more, in fact, than for the values interview condition. This may have occurred because in contrast to the values interview condition, where *both* personal observations *and* factual information about personal values were available to a judge, in the values protocol condition only the latter were used. Items referring to personal values were thus more salient for the protocol than for the interview. In addition to the seven items dealing directly with values (Items 3, 7, 51, 63, 66, 83, 90), there were also four items referring to temperament (Items 16, 44, 46, 76) and two referring to intellectual functioning (Items 8, 51).

The items showing their largest variation in the noninterview condition are listed in Table 7. With the exception of a single item dealing with impulse control (Item 25), the remaining items can be classified under the heading of modes of interaction. This finding too is appropriate; that is, in this condition judges saw the target persons in casual social interaction throughout the assessment day but did not obtain detailed facts in the realms of values, life history, or character. These judges should then make most differential use of items pertaining to social style and physical demeanor.

Table 8 presents the items that showed their largest variation in the stereotypic-

Table 3
Q-Sort Items Showing Greatest Variance in Life History Interview Condition

4.	Is a talkative individual.[a]
9.	Is uncomfortable with uncertainty and complexities.
2.	Tends to be self-defensive.[a]
7.	Behaves in a sympathetic or considerate manner.
0.	Has rapid personal tempo; behaves and acts quickly.[b]
3.	Extrapunitive; tends to transfer or project blame.
6.	Is productive; gets things done.
0.	Gives up and withdraws where possible in the face of frustration and adversity. (N.B. If placed high, implies *generally defeatist*; if placed low, implies *counteractive.*)
1.	Regards self as physically attractive.[a]
2.	Seems to be aware of the impression he makes on others.
1.	Is moralistic. (N.B. Regardless of the particular nature of the moral code.)
8.	Keeps people at a distance; avoids close interpersonal relationships.
9.	Is basically distrustful of people in general; questions their motivations.
4.	Emphasizes being with others; gregarious.[b]
8.	Enjoys sensuous experiences (including touch, taste, smell, physical contact).[a]
0.	Has insight into own motives and behavior.[a]
2.	Tends to be rebellious and nonconforming.
8.	Is basically anxious.[b]
0.	*Behaves* in an ethically consistent manner; is consistent with own personal standards.
3.	Tends to perceive many different contexts in sexual terms.[a]
5.	Has a clear-cut, internally consistent personality. (N.B. *Amount* of information available before sorting is not intended here.)
7.	Appears straightforward, forthright, candid in dealings with others.[a]
0.	Interested in members of the opposite sex. (N.B. At opposite end, item implies *absence* of such interest.)[a]
6.	Handles anxiety and conflicts by, in effect, refusing to recognize their presence; repressive or dissociative tendencies.[a]
9.	Compares self to others. Is alert to real or fancied differences between self and other people.
7.	Is emotionally bland; has flattened affect.
9.	Is self-dramatizing; histrionic.[a]

[a] Both heterogeneity of variance tests significant ($n = 10$). [b] One test significant ($n = 3$).

Table 4
Q-Sort Items Showing Greatest Variance in Values Interview Condition

2. Is a genuinely dependable and responsible person.[b]
14. *Genuinely* submissive; accepts domination comfortably.
22. Feels a lack of personal meaning in life.[a]
27. Shows condescending behavior in relations with others. (N.B. Extreme placement toward uncharacteristic end implies simply an *absence* of condescension, not necessarily equalitarianism or inferiority.)
35. Has warmth; has the capacity for close relationships; compassionate.
36. Is subtly negativistic; tends to undermine and obstruct or sabotage.
37. Is guileful and deceitful, manipulative, opportunistic.
38. Has hostility toward others. (N.B. Basic hostility is intended here; mode of expression is to be indicated by other items.)[b]
39. Thinks and associates ideas in unusual ways; has unconventional thought processes.
40. Is vulnerable to real or fancied threat, generally fearful.
42. Reluctant to commit self to any definite course of action; tends to delay or avoid action.
43. Is facially and/or gesturally expressive.[a]
45. Has a brittle ego-defense system; has a small reserve of integration; would be disorganized and maladaptive when under stress or trauma.[a]
61. Creates and exploits dependency in people (N.B. regardless of the techniques employed, e.g., punitiveness, overindulgence). (N.B. At other end of scale, item implies respecting and encouraging independence and individuality of others.)
64. Is socially perceptive of a wide range of interpersonal cues.
72. Concerned with own adequacy as a person, either at conscious or unconscious levels. (N.B. A clinical judgment is required here; Number 74 reflects subjective satisfaction with self.)[b]
74. Is subjectively unaware of self-concern; feels satisfied with self.[a]
78. Feels cheated and victimized by life; self-pitying.[b]
84. Is cheerful. (N.B. Extreme placement toward uncharacteristic end of continuum implies unhappiness or depression.)[b]
91. Is power oriented; values power in self or others.
92. Has social poise and presence; appears socially at ease.[b]
98. Is verbally fluent; can express ideas well.

[a] Both heterogeneity of variance tests significant ($n = 4$). [b] One test significant ($n = 6$).

Table 5
Q-Sort Showing Greatest Variance in Life History Protocol Condition

1. Is critical, skeptical, not easily impressed.
5. Behaves in a giving way toward others (N.B. regardless of the motivation involved).
10. Anxiety and tension find outlet in bodily symptoms. (N.B. If placed high, implies bodily dysfunction if placed low, implies absence of autonomic arousal.)[a]
11. Is protective of those close to him. (N.B. Placement of this item expresses behavior ranging from overprotection through appropriate nurturance to a laissez-faire, underprotective manner.)
24. Prides self on being "objective," rational.
29. Is turned to for advice and reassurance.
34. Overreactive to minor frustrations; irritable.[b]
47. Has a readiness to feel guilty (N.B. regardless of whether verbalized or not).
50. Is unpredictable and changeable in behavior and attitudes.
52. *Behaves* in an assertive fashion. (N.B. Item 14 reflects underlying submissiveness; this refers to overt behavior.)
53. Various needs tend toward relatively direct and uncontrolled expression; unable to delay gratification.
55. Is self-defeating.
59. Is concerned with own body and the adequacy of its physiological functioning.[a]
65. Characteristically pushes and tries to stretch limits; sees what he can get away with.
67. Is self-indulgent.[b]
79. Tends to ruminate and have persistent, preoccupying thoughts.[b]
82. Has fluctuating moods.
87. Interprets basically simple and clear-cut situations in complicated and particularizing ways.
94. Expresses hostile feelings directly.

[a] Both heterogeneity of variance tests significant ($n = 2$). [b] One test significant ($n = 3$).

Table 6
Q-Sort Items Showing Greatest Variance in Values Protocol Condition

7.	Has a wide range of interests. (N.B. Superficiality or depth of interest is irrelevant here.)
7.	Favors conservative values in a variety of areas.
8.	Appears to have a high degree of intellectual capacity (N.B. Whether actualized or not; originality is not necessarily assumed).
3.	Is thin-skinned; sensitive to anything that can be construed as criticism or an interpersonal slight.
6.	Is introspective and concerned with self as an object. (N.B. Introspectiveness *per se* does not imply insight.)
4.	Evaluates the motivations of others in interpreting situations. (N.B. Accuracy of evaluation is not assumed.) (N.B. Again, extreme placement in one direction implies preoccupation with motivational interpretation; at the other extreme, the item implies a psychological obtuseness; subject does not consider motivational factors.)
6.	Engages in personal fantasy and daydreams, fictional speculations.
1.	Genuinely values intellectual and cognitive matters. (N.B. Ability or achievement are not implied here.)
3.	Judges self and others in conventional terms of "popularity," "the correct thing to do," social pressures, and so forth.
6.	Enjoys aesthetic impressions; is aesthetically reactive.[a]
9.	Is sensitive to anything that can be construed as a demand. (N.B. No implication of the kind of subsequent response is intended here.)
6.	Tends to project his own feelings and motivations onto others.[b]
8.	Able to see to the heart of important problems.
0.	Is concerned with philosophical problems, e.g., religions, values, the meaning of life, and so forth.[a]

[a] Both heterogeneity of variance tests significant ($n = 2$). [b] One test significant ($n = 1$).

emographic information condition. Ideally, o item should have shown greatest variance n this condition. In fact, none showed signifiant heterogeneity of variance, and their anking highest in this condition may best e viewed as a minor fluctuation around the nore or less equivalent variances of the six onditions.

We may now summarize the Q item variance results. Overall, the pattern is a strong ne; the three assessment conditions, in which ssessees were actually observed, contain all he items making reference to physical qualties of appearance and bearing. This finding s hardly surprising, but its absence would ave raised serious doubts about the effects f information on personality descriptions. he noninterview condition, where no intiate contact occurred but persons were nonetheless personally observed, contained items elating to social interaction and social skills. he values protocol condition contained the bulk of the values items. The two life history nformation conditions (interview and protocol) contained a preponderance of items dealng with personality style, socialization, adjustment, and sexuality, all topics that were explicitly discussed in the interview. Finally, the stereotypic-demographic information condition contained only five items, all lacking significant heterogeneity of variance.

Viewing the six tables as a whole reveals

Table 7
Q-Sort Items Showing Greatest Variance in Noninterview Condition

15.	Is skilled in social techniques of imaginative play, and humor.[b]
18.	Initiates humor.[a]
19.	Seeks reassurance from others.
21.	Arouses nurturant feelings in others.
25.	Tends toward overcontrol of needs and impulses; binds tensions excessively; delays gratification unnecessarily.[b]
28.	Tends to arouse liking and acceptance in people.[b]
33.	Is calm, relaxed in manner.[b]
56.	Responds to humor.
57.	Is an interesting, arresting person.
81.	Is physically attractive; good-looking. (N.B. The cultural criterion is to be applied here.)[a]
88.	Is personally charming.[a]
93.	*Behaves* in a masculine style and manner *or Behaves* in a feminine style and manner. (N.B. The cultural or subcultural conception is to be applied as a criterion.)[b]
95.	Tends to proffer advice.

[a] Both heterogeneity of variance tests significant ($n = 3$). [b] One test significant ($n = 8$).

Table 8
Q-Sort Items Showing Greatest Variance in Stereotypic-Demographic Condition

6.	Is fastidious.
71.	Has high aspiration level for self.
85.	Emphasizes communication through action and nonverbal behavior.
96.	Values own independence and autonomy.
100.	Does not vary roles; relates to everyone in the same way.

another interesting and important result. As information decreases, so does the number of items that show significant heterogeneity of variance. So, for example, in Table 3, which tallies items for the life history interview condition, nearly half of the items show significant heterogeneity according to one or both tests. This contrasts with Table 8, which shows no significant items in the stereotypic-demographic condition. The number of significantly heterogeneous items is proportional to the total number of items in each condition, which is in turn related to the amount and kind of information available in each condition. This pattern, then, further corroborates the abilities of judges to respond to and reflect the variations in information on which their descriptions are based.

ACL items. To show that the effects of information are not confined to a single descriptive method, an analysis of the ACL data paralleling that for the Q-sort was performed. This analysis involved comparing across conditions the total number of adjectives checked about assessees. In the assessment conditions, where judges had direct contact with their targets, the mean number of adjectives checked per assessee should be higher than for the protocol and stereotype conditions, where contact with the target was only through written documents. The total number of adjectives checked for each case by each judge was computed. Differences between conditions were evaluated by a one-way analysis of variance. The results are presented in Table 9.

Table 9
Analysis of Variance of Total Adjective Check List Items Checked

Condition	*n*	*M*	*SD*	*F*
Life history interview	80	89.96	20.54	—
Values interview	78	85.20	18.91	—
Life history protocol	164	58.27	20.37	92.40*
Values protocol	160	54.96	22.04	—
Noninterview	224	83.93	19.67	—
Stereotypic-demographic	164	51.90	23.25	—

* p (5, 864) = .0001.

The findings for the ACL parallel those found with the Q-sort. The total number of adjectives checked about a target person varies substantially with information condition, which accounts for 34% of the variance. When judges had personal contact with their targets, they checked approximately 85 items as descriptive; judges without such contact checked only about 54 adjectives. Scheffé's test for differences among the means of the total number of adjectives checked reveals two homogeneous subsets significant at the .05 level. As predicted, the protocol and stereotype conditions form one group and the assessment conditions form the other. Thus on a second and independent measuring instrument, judges in this study demonstrated that information does indeed affect description: When information available to judges is more abundant and varied, their descriptions are correspondingly fuller and more detailed.

Within-Condition Homogeneity of Description

In the analysis just reported, differentiation among assessees was examined at the level of individual items. In the analysis to follow, differentiation among assessees will be examined at a more global level. Specifically, the Q correlations between assessees provide the basic data. For each sex separately, six matrices of interperson correlations were obtained, one in each condition. As Block (1961) notes, the Q correlation is a convenient index of interperson similarity. If information condition had no effect, then the mean level of similarity between assessees would be no greater in any one condition than in any other. If, however, information does affect description, then a relationship between condition and similarity should obtain. The

stereotypic-demographic condition should produce the most homogeneous (i.e., the least differentiated) set of descriptions, and the interview conditions the least homogeneous set. In Table 10 the mean interperson Q correlation is shown for each information condition.

The patterning of the means of interperson similarity emphasizes the results from the analysis of variance of ranks; the least differentiation among assessees was obtained in the stereotypic-demographic condition, where the mean of all correlations between male assessees was .40 and the corresponding mean for females was .44. In marked contrast, in the life history interview condition, the mean interperson correlation for males was .15 and for females was .14. Information condition accounts for 33% of the variance in correlational similarity. Additionally, the Bartlett-Box F test shows significant ($p < .01$) heterogeneity of variance. The life history interview condition shows less spread than the other conditions do. An interesting, albeit unanticipated, finding was that there was a significant difference in mean similarity between males and females in the stereotypic-demographic condition. The females were described more homogeneously than the males, $t(80) = 2.60$, $p = .01$.

In bold terms these analyses show that when little or no information is available, differentiation among assessees is low. On the other hand, when information is detailed, intimate, and characterologically revealing, judges describe their targets in a far more differentiated fashion. Finally, when the amount of information available is intermediate, the level of differentiation is intermediate. This is perhaps the clearest evidence that the degree to which judges can differentiate among assessees is a direct function of the degree to which differentiating information is available to them.

Table 10
Analysis of Variance of Interperson Similarity

Condition	n	M	SD	F
All assessees (df = 5, 479)				
Life history interview	80	.14	.09	—
Values interview	77	.19	.15	—
Life history protocol	82	.24	.13	49.20*
Values protocol	82	.28	.14	—
Noninterview	82	.29	.14	—
Stereotypic-demographic	82	.44	.13	—
Male assessees (df = 5, 235)				
Life history interview	40	.15	.08	—
Values interview	37	.16	.12	—
Life history protocol	41	.26	.13	24.59*
Values protocol	41	.29	.12	—
Noninterview	41	.28	.12	—
Stereotypic-demographic	41	.40	.13	—
Female assessees (df = 5, 238)				
Life history interview	40	.14	.10	—
Values interview	40	.21	.16	—
Life history protocol	41	.23	.13	26.85*
Values protocol	41	.28	.15	—
Noninterview	41	.30	.15	—
Stereotypic-demographic	41	.47	.13	—

* $p < .001$.

Discussion

The study reported above has been concerned with a basic and fundamental problem in personality psychology: Are human observers capable of giving reliable and sensitive descriptions of personality? Recent writing in personality assessment has questioned the continued use of raters of personality, claiming that the data they produce are biased, unreliable, and insensitive to the "actual behavior" of the ratee. The central assertion of the criticisms of personality ratings is that the ratings do not reflect the "actualities" of the stimuli upon which they are based. The processes by which observers make their judgments are hypothesized to reflect various kinds of artifacts (e.g., logical error, illusory correlation). This pessimistic conclusion has been based on several types of analyses, most of which have considered solely the patterning or intercorrelations of the ratings. The evidence from the ratings themselves, however, has not generally been evaluated. This step logically comes prior to examining the structure of the ratings, and if convincing evidence of the tenability of ratings can be demonstrated, then the results derived from analyses of the patterning of ratings become phenomena to be explained, rather than

demonstrations of the irrelevance of personality ratings.

To demonstrate the tenability of observers' ratings of personality in this study, the amount and kind of information provided to judges was systematically varied, and the effects on the subsequent ratings were examined. If judges' ratings were insensitive to the actualities of the information, then no systematic effects should have been produced by the systematic variation in information. If, however, variations in information produce clear, comprehensible, and predictable variations in personality descriptions, the inescapable conclusion is that judges can and do respond sensitively to the information from which descriptions are made.

We may now discuss the evidence in this study that supports an optimistic view of the capacities of observers. A primary demonstration of the effects of information on description examined the differential use of descriptions. The analysis of variance of ranks of Q item variances and the analysis of the mean number of total adjectives checked on the ACL both provided strong evidence that judges perceived, integrated, and responded to the information each condition provided.

An alternative explanation for the differential variation of Q item placements across conditions might be that adventitious aspects of the assignment of cases to judges within condition, rather than responsiveness to changes in information, produced the observed results. But this reasoning would not explain why items that showed differential variation were items with content that was relevant to the information condition in which they showed increased variation. This second aspect of the data argues strongly against such an alternative explanation. It argues for judges' competent use of information.

The ACL demonstrates the effects of information on description in a different fashion. Information in the protocol and stereotype conditions led judges to reduce markedly the sheer number of items judged to be descriptive of an assessee. Thus, without a forced format, as in the Q-sort, the judges' own descriptive behavior on the ACL demonstrates their responsiveness to variations in information. When the information available lacks immediacy, i.e., when the person is not physically present with the judge, descriptions are less full and rich.

The analysis of interperson similarity probably provided the most vivid and comprehensible demonstration of the effects on personality assessment of variations in information. If differentiation in personality assessment is a function of information, then the degree of homogeneity in description should be related to information condition. The analyses showed that the amount of similarity of description among persons was inversely related to the amount of information. Additionally, the homogeneity of variance of mean interperson correlation decreased in the condition where information was greatest—the life history interview condition. This finding, perhaps self-evident, is nonetheless heartening to those who have argued for the continued use of the human observer in personality assessment. Because the target sample was the same in each condition, because the measure used to describe the targets was the same in each condition, and because the amount of judge overlap was great, the findings are best interpreted as showing that judges altered their descriptive behavior depending on what information was provided.

The analysis of the level of interobserver agreement as a function of information condition yielded less straightforward results. Differences between conditions were not of large magnitude for the ACL and did not occur with the Q-sort; differences in the homogeneity of agreement coefficients were present, however. Several reasons may be suggested for finding equivalent levels of reliability with the Q-sort:

1. The Q-sort distribution is forced and therefore if a core of agreement exists, it will be magnified as a function of the equalization of means and standard deviations for all judges, regardless of the information they had.

2. The item content of the Q-sort is quite variable; some items are rarely sorted as strongly characteristic of anyone. That is, the location of placements for some items is highly modal, and agreement is therefore heightened relative to a set with more normally distributed items, whatever the condition of information.

3. For the stereotypic-demographic condition, in which agreement was as high as it was for the other conditions, it may very well be the case that judges possess a shared cultural stereotype and can, if need be, draw upon this for their descriptions when they have no alternative. Nonetheless, this shared stereotype does not provide the detailed information needed to make item placements for many of the Q items. Hence, the level of agreement reached is only moderate.

The meaning of the level of agreement that exists for the stereotypic-demographic condition is revealed by the analyses demonstrating that the agreement reached under impoverished information was reached at the price of differentiated descriptions of the targets. Thus judges who do not have personally relevant information agree with each other about the same target as much as life history interview judges do; further, across assessees they agree to a much larger degree. In fact, the stereotypic-demographic condition is the only condition in which there is not a single negative mean interperson correlation for any assessee. This is what one would expect if the judges did not have the requisite information to differentiate among assessees but nonetheless had knowledge about characteristics associated with classes of persons.

The crucial question in regard to interobserver agreement is whether equivalent levels across information conditions undercut conclusions about the sensitivity of observers. Agreement for all conditions was only moderate. Parenthetically, it should be noted that the reliabilities reported are analogous to interitem correlations. The reliabilities of the composite descriptions furnished by four or five judges for these 82 cases had a mean value of .702 and a median value of .729 on the Q-sort. These composite descriptions are routinely employed in assessment studies rather than the descriptions from pairs of judges. Nonetheless, it is still the case that the level of agreement is equivocal with respect to the sensitivity of observers. For example, what should be said about a 32-year-old housewife and mother of two, having 16 years of education and working part-time as a substitute teacher? Artifactual accounts of personality assessment based on implicit personality theory or semantic similarity of trait terms seem to require a much higher degree of consensus than that which in fact occurs. This suggests that shared assumptions about personality are not as pervasive as some commentators have indicated (Shweder, 1975, 1977).

An implication of the findings on agreement is that there may be both a floor and a ceiling, below or above which consensus cannot extend. Given less than total but more than minimal information about a large set of personality variables for which judgments will be requested, judges may well be able to agree to a certain point, regardless of the type of information with which they are furnished. For example, if it can be shown that in conditions of impoverished information, agreement among judges is principally a function of strong consensus about a small number of descriptors, whereas in conditions of more abundant information agreement stems from more mild consensus about a larger number of descriptors, our understanding of the nature of interjudge agreement would be deepened. Alternatively, it may be the case, as some of the clinical judgment studies show, that more information leads to the discounting of base-rate information that is actually more accurate and predictive. The role of extent of information is psychologically complex, as this second formulation makes clear.

In arguing for an acceptance of the data that observers provide about personality, we do not suggest that this method is the only appropriate method for study in personality. Nor do we suggest that the use of observers has reached its practical limits in regard to reliability and validity. Training, more familiarity with the response instruments, and a larger number of observers are all ways of increasing the reliability and accuracy of observers' judgments. It is, of course, obvious that the results discussed here do not certify that the ratings of these assessees are valid. That question must await further analyses of the data. But it is reasonable to assume that the prerequisites for validity are present, and that a clear possibility for impressive nonrating correlates exists for these data. A key consideration, moreover, in establishing accuracy is that for many, if not most or all,

studies using other methods of assessment, ultimately validity rests on the correspondence of observations with the criteria laid down by observers. As argued elsewhere (Block, Weiss, & Thorne, 1979), appeals to behavior counts are no less dependent on observers than are ratings. Disagreement among researchers arises regarding the meaning of the observations not their necessity.

The present report has asserted that the use of the human observer as a personality assessment tool is reasonable and useful and that personological descriptions furnished by observers are reliable and often acutely sensitive to the nature of the stimuli on which they are based. The conclusion is that observers are sensitive encoding "instruments." Despite the known biasing effects on ratings of halo and the logical error, for example, it is unlikely that there exists at present a method more economic, precise, and valid for gathering personality assessment data than that utilizing the human observer. The basic soundness of the data in this study supports this observation.

References

Berman, J. S., & Kenny, D. A. Correlational bias in observer ratings. *Journal of Personality and Social Psychology*, 1976, *34*, 263–273.

Block, J. *The Q-sort method in personality assessment and psychiatric research*. Springfield, Ill.: Charles C Thomas, 1961.

Block, J. *Lives through time*. Berkeley, Calif.: Bancroft Books, 1971.

Block, J., Weiss, D. S., & Thorne, A. How relevant is a semantic similarity interpretation of personality ratings? *Journal of Personality and Social Psychology*, 1979, *37*, 1055–1074.

Bourne, E. Can we describe an individual's personality? Agreement on stereotype versus individual attributes. *Journal of Personality and Social Psychology*, 1977, *35*, 863–872.

Chapman, L. J., & Chapman, J. P. The genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology*, 1967, *72*, 193–204.

Chapman, L. J., & Chapman, J. P. Illusory correlation as an obstacle to the use of valid psychodiagnostic observations. *Journal of Abnormal Psychology*, 1969, *74*, 271–280.

D'Andrade, R. G. Trait psychology and componential analysis. *American Anthropologist*, 1965, *67*(5, pt. 2), 215–228.

D'Andrade, R. G. Memory and the assessment of behavior. In H. M. Blalock (Ed.), *Measurement in the social sciences*. Chicago: Aldine, 1974.

Fiske, D. W. *Measuring the concepts of personality*. Chicago: Aldine, 1971.

Fiske, D. W. Can a personality construct be empirically validated? *Psychological Bulletin*, 1973, *80*, 89–92.

Fiske, D. W. The limits for the conventional science of personality. *Journal of Personality*, 1974, *42*, 1–11.

Fiske, D. W. Can a personality construct have a singular pattern? Rejoinder to Huba and Hamilton. *Psychological Bulletin*, 1976, *83*, 877–879.

Golden, M. Some effects of combining psychological tests on clinical inferences. *Journal of Consulting Psychology*, 1964, *28*, 440–446.

Gough, H. G. A factor analysis of contraceptive preferences. *Journal of Psychology*, 1973, *84*, 199–210.

Gough, H. G. An attitude profile for use in studies of population psychology. *Journal of Research in Personality*, 1975, *9*, 122–135.

Gough, H. G., & Heilbrun, A. B., Jr. *Manual for the Adjective Check List*. Palo Alto, Calif.: Consulting Psychologists Press, 1965.

Kostlan, A. A. A method for the empirical study of psychodiagnosis. *Journal of Consulting Psychology*, 1954, *18*, 83–88.

Mischel, W. *Personality and assessment*. New York: Wiley, 1968.

Mischel, W. Towards a cognitive social learning reconceptualization of personality. *Psychological Review*, 1973, *80*, 252–283.

Norman, W. T. Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings. *Journal of Abnormal and Social Psychology*, 1963, *66*, 574–583.

Norman, W. T., & Goldberg, L. Raters, ratees, and randomness in personality structure. *Journal of Personality and Social Psychology*, 1966, *4*, 681–691.

Passini, F. T., & Norman, W. T. A universal conception of personality structure. *Journal of Personality and Social Psychology*, 1966, *4*, 44–49.

Shweder, R. A. How relevant is an individual difference theory of personality? *Journal of Personality*, 1975, *43*, 455–484.

Shweder, R. A. Illusory correlation and the MMPI controversy. *Journal of Consulting and Clinical Psychology*, 1977, *45*, 917–924.

Sines, L. K. The relative contribution of four kinds of data to accuracy in personality assessment. *Journal of Consulting Psychology*, 1959, *23*, 483–495.

Weiss, D. S. *The effects of source variation in personality assessment*. Unpublished doctoral dissertation, University of California, Berkeley, 1978.

Wiggins, N. H., & Blackburn, M. C. Implicit theories of personality: An individual differences approach. *Multivariate Behavioral Research*, 1976, *11*, 267–285.

Winer, B. J. *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill, 1971.

Received March 2, 1979

Announcing a new Quarterly Journal . . .

JOURNAL OF

Occupational Behaviour

Editors:

Cary L. Cooper
Department of Management Sciences, University of Manchester Institute of Science and Technology

James C. Taylor
Center for Quality of Working Life, University of California, Los Angeles

An international journal of social science research on human behaviour in the work place and on related social change. For all social, behavioural and industrial scientists, teachers and practitioners who are concerned with occupational behaviour and management

Typical areas of coverage in the *Journal of Occupational Behaviour* will be:

★ Innovations in job design and enrichment
★ Work experience, occupational satisfaction and stress
★ Industrial participation and democracy in the work place
★ Career development, retraining and retirement
★ Work in its legal and socio-economic context
★ Implications of Work for Leisure, the Family and Public Life

Subscription Details: Volume 1 (1980) Quarterly UK £22.50 Elsewhere US $55.00
Special Subscription Rate for Personal Subscribers: UK £12.50 Elsewhere US $30.00
(Air-speeded delivery to USA and Canada)
For a free specimen copy, or for further information, please write to the Subscription Department at the more convenient address below.

John Wiley & Sons Ltd.
Baffins Lane
Chichester
Sussex PO19 1UD
England

John Wiley & Sons Inc.
605 Third Avenue
New York
N.Y. 10016
U.S.A.

Publisher with major reference book program in the humanities is expanding in the social sciences.

Editors

are being sought to serve as consultants for programs in their fields of expertise. Books in these programs will include bibliographies, guides to information sources, directories, and dictionaries. Prompt publication and generous royalties are guaranteed. Subsidies from authors are not accepted. Inquiries should be addressed to:

Social Science Reference Book Editor
Garland Publishing, Inc.
136 Madison Avenue
New York, New York 10016

The Journal Supplement Abstract Service is an information service providing ready access to psychology-related materials not available through conventional communication channels.

JSAS Selects and Disseminates:

- Educational materials
- Massive data collections
- Methodological techniques and procedures
- Major projects in progress
- Descriptions of effective techniques or programs
- Technical reports
- Fresh looks at controversial issues
- Invited lectures
- Management of psychological resources
- APA task force reports
- Bibliographies
- Literature reviews

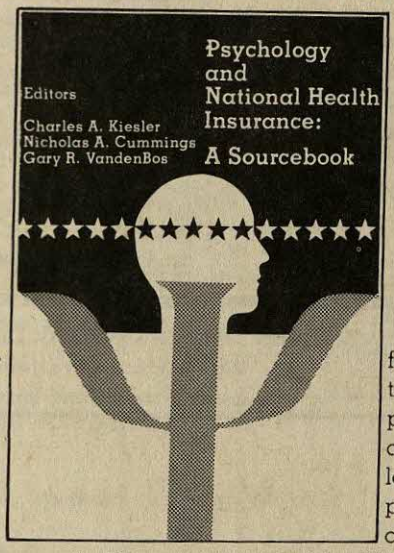
Manuscripts are accepted by a panel of editors. They meet the same high standards set for APA's journals, but are not bound by conventional restrictions of format, length, or editorial policy.

Available titles are listed in the quarterly JSAS *Catalog of Selected Documents in Psychology.* The catalog includes abstracts of newly accepted manuscripts along with manuscript length, price, and ordering information. Each issue of the catalog contains about 50 abstracts. Full-text manuscripts may be purchased individually in paper or microfiche.

Subscription rates for the quarterly catalog are $6 for APA members and $14 for nonmembers (foreign, $16). Please include full remittance for orders of $25 or less. To order, make check payable to American Psychological Association.

American Psychological Association Subscription Dept., 1200 17th Street, NW, Washington, DC 20036

Psychology and National Health Insurance



Psychology and National Health Insurance: A Sourcebook is the only collection of current readings on the status and future of psychology in national health insurance programs. Emphasizing the valuable services offered by psychologists and examining professional standards, the editors provide legislators, psychologists, and the general public with a collection of articles reprinted from a variety of sources as well as new material.

Included are discussions of the following topics: • Surveys on the numbers, locations, and practices of health service providers in psychology • Evaluation; accountability, service review issues • Redistributive aspects of NHI • Current legal and legislative status of professional psychology • Issues in professional training • National health insurance plans • Psychotherapy outcomes and psychological interventions in physical health • Value of prevention services in national health insurance programs.

A detailed index and full citation of all materials are also provided.

Edited by **Charles A. Kiesler, Nicholas A. Cummings,** and **Gary R. VandenBos,** 1979, 680 pages, 7" x 10", index. Price: $19.50 hard cover, $15.00 soft cover.

To order, write to: American Psychological Association, Order Dept. • 1200 Seventeenth Street, N.W. • Washington, D.C. 20036.

Orders of $25 or less must be prepaid.

New Titles in Psychology from O,G & H

SKILL IN NONVERBAL COMMUNICATION

Individual Differences

Edited by

ROBERT ROSENTHAL,

Harvard University

There are increasing indications that interpersonal interactions may be affected in important ways by the interactants' ability to decode and to encode nonverbal cues. These papers offer major evidence for important research investigating the ways in which these skills in decoding and encoding nonverbal cues are related to personal characteristics and to the outcomes of social encounters. Their various findings support, or amplify, or challenge views held in the field of nonverbal communication — some speak to issues rarely raised before — all of them contribute to the understanding of nonverbal communication.

288 pages Available $18.50

BASIC DIMENSIONS FOR A GENERAL PSYCHOLOGICAL THEORY

Implications for Personality, Social, Environmental, and Developmental Studies

ALBERT MEHRABIAN,

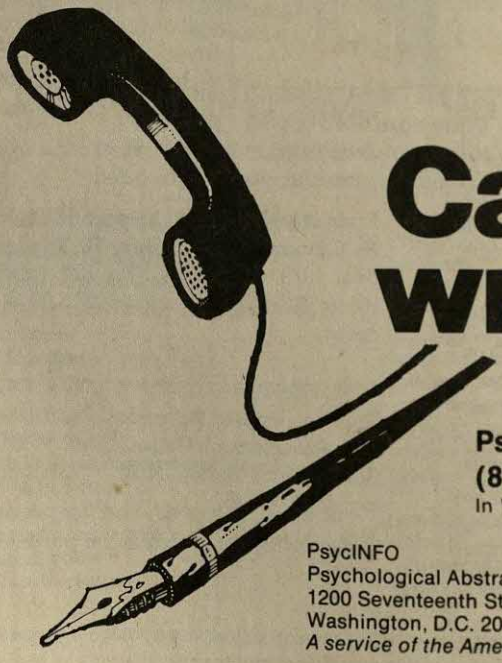
University of California, Los Angeles

This book proposes an approach to the study of human emotions based on a concise description of emotional states which uses a three-dimensional framework of temperament (pleasure, arousability, dominance) for describing emotional traits. This framework provides a set of independent factors for studying numerous problems in personality, social, developmental, and environmental psychology. The author examines the implications of this framework for understanding differential preferences for various situations, explores the ways in which emotional states can lead to seeking or avoiding different means of chemically modifying the bodily environment and concludes with a discussion of various theoretical and experimental studies based on these theories.

ca. 416 pages Available in December $20.00

O G & H Oelgeschlager, Gunn & Hain, Inc. *Publishers*

1278 Massachusetts Ave., Harvard Square, Cambridge, MA 02138



Call or write!

PsycINFO Action Line

(800) 336-4980

In Virginia call (202) 833-5908

PsycINFO
Psychological Abstracts Information Services
1200 Seventeenth Street, N.W.
Washington, D.C. 20036
A service of the American Psychological Association

PsycINFO, the database of record in psychology and related disciplines

(Continued from front cover)

Emotion Recognition: The Role of Facial Movement and the Relative Importance of Upper and Lower Areas of the Face 2049
John N. Bassili

Psychological Masculinity and Femininity and Typical and Maximal Dominance Expression in Women 2059
Helen M. Klein and Lee Willerman

Effects of Person Salience Versus Role Salience on Reward Allocation in a Dyad 2071
Elena M. Carles and Charles S. Carver

Allocation of Attention and the Type A Coronary-Prone Behavior Pattern 2081
Karen A. Matthews and Bradford I. Brunson

The Power Motive as an Influence on Use of Power 2091
Eugene M. Fodor and Dana L. Farrow

Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence 2098
Charles G. Lord, Lee Ross, and Mark R. Lepper

Sex Differences in Bystander Intervention in a Theft 2110
William Austin

The Effects of Systematic Variations in Information on Judges' Descriptions of Personality 2121
Daniel S. Weiss

Corrections to DePaulo, Rosenthal, Eisenstat, Rogers, and Finkelstein 1965

New JPSP Format Delayed Until April 1992

Instructions for Submission of Manuscripts 2013

17/4
80

INNER PERSPECTIVES FROM DORSEY

Personality Theories: A Comparative Analysis, 4th Edition

Salvatore R. Maddi, University of Chicago

Thoroughly absorbing, this text compares and contrasts many theoretical viewpoints with a comprehensive, consistent, and original framework for the personality field. Pertinent issues are posed, and research studies are examined for their value in resolving the issues. Teacher's Manual.

Coming in 1980.

Understanding Social Psychology, Revised Edition

Stephen Worchel, University of Virginia, and Joel Cooper, Princeton University

The traditional areas of social psychology are covered in both theory and research along with demonstrations of how social psychology can and cannot be used to understand "real world" incidents. This edition includes a new chapter on socialization. Student Study Guide. Teacher's Manual.

Published 1979.

Personality: Strategies and Issues, 3rd Edition

Robert M. Liebert, State University of New York at Stony Brook, and Michael D. Spiegler, Providence College

This volume provides a widely resepcted and highly readable introduction to major strategies and issues in the psychological study of personality. Students can see clearly both strengths and weaknesses of each viewpoint. New discussions of attribution theory and cognitive behavior modifications are included. Instructor's Manual.

Published 1978.

THE DORSEY PRESS

HOMEWOOD, ILLINOIS 60430

DP

Request examination copies for adoption consideration. Please indicate course title and text presently used.

JOURNAL OF

Personality AND Social Psychology

VOLUME 37 NUMBER 12 · DECEMBER 1979

Settings of a Professional Lifetime 2137
Roger G. Barker

Roger G. Barker and Behavior Settings: A Commentary 2158
Phil Schoggen

The Identification of Persons as Supersets and Subsets in Free-Response Personality Descriptions 2161
Michael A. Gara and Seymour Rosenberg

Internalization Versus Identification in the Laboratory: A Causal Analysis of Attitude Change 2171
Daniel Romer

Attitudes and Cognitive Response: An Electrophysiological Approach 2181
John T. Cacioppo and Richard E. Petty

Remembering Schema-Consistent Information: Effects of a Balance Schema on Recognition Memory 2200
Keith P. Sentis and Eugene Burnstein

Locus of Control, Interpersonal Trust, and Assertive Behavior Among Newlyweds 2212
William J. Doherty and Robert G. Ryder

An Examination of Self-Perception Mediation of the Foot-in-the-Door Effect 2221
William DeJong

Salience and the Cognitive Mediation of Attribution 2240
Eliot R. Smith and Frederick D. Miller

Effects of "Voice" and Peer Opinions on Responses to Inequity 2253
Robert Folger, David Rosenfield, Janet Grove, and Louise Corkran

Interpersonal Attraction and the Polarity of Similar Attitudes: A Test of Three Balance Models 2262
Abbas Tashakkori and Chester A. Insko

Personality Structure and the Circumplex 2278
Douglas N. Jackson and Edward Helmes

Externality as a Function of Obesity in Children: Pervasive Style or Eating-Specific Attribute? 2286
Philip R. Costanzo and Erik Z. Woody

(Continued on inside back cover)

Published monthly by the

AMERICAN PSYCHOLOGICAL ASSOCIATION

ISSN 0022-3514

Acting Editor

Clyde Hendrick, *University of Miami*

Action Editors

Irwin Altman, *University of Utah*
Charles S. Carver, *University of Miami*
Richard J. Harris, *University of New Mexico*
John H. Harvey, *Vanderbilt University*
William Ickes, *University of Missouri–St. Louis*
George Levinger, *University of Massachusetts—Amherst*
Brendan Gail Rule, *University of Alberta, Edmonton, Canada*
David R. Shaffer, *University of Georgia*
Cookie White Stephan, *New Mexico State University*
Barbara Strudler Wallston, *George Peabody College for Teachers*

Consulting Editors

Robert Steven Baron, *University of Iowa*
Paul A. Bell, *Colorado State University*
Ellen Berscheid, *University of Minnesota (Minneapolis)*
Michael H. Birnbaum, *University of Illinois at Urbana-Champaign*
Jack Brehm, *University of Kansas*
Philip Brickman, *Northwestern University*
J. Merrill Carlsmith, *Stanford University*
Paul Christopher Cozby, *California State University, Fullerton*
Rick Crandall, *Texas Christian University*
James M. Dabbs, Jr., *Georgia State University*
John M. Darley, *Princeton University*
Andrew R. Davidson, *University of Washington*
James H. Davis, *University of Illinois at Urbana-Champaign*
Kay Deaux, *Purdue University*
Marshall Dermer, *University of Wisconsin—Milwaukee*
Edward Donnerstein, *Iowa State University*
Ebbe B. Ebbesen, *University of California, San Diego*
Yakov M. Epstein, *Rutgers—The State University*
Irene Hanson Frieze, *University of Pittsburgh*
J. Martin Giesen, *Mississippi State University*
Lewis R. Goldberg, *University of Oregon*
William G. Graziano, *University of Georgia*
Alan E. Gross, *University of Maryland (College Park)*
Nancy Hirschberg, *University of Illinois at Chicago Circle*
Chester A. Insko, *University of North Carolina at Chapel Hill*
Alice M. Isen, *University of Maryland, Baltimore County*
Martin F. Kaplan, *Northern Illinois University*
Samuel S. Komorita, *University of Illinois at Urbana-Champaign*
Vladimir J. Konečni, *University of California, San Diego*
Ellen Langer, *Harvard University*
Mark R. Lepper, *Stanford University*
Darwyn E. Linder, *Arizona State University*
Dale T. Miller, *University of British Columbia, Vancouver, Canada*
Judson Mills, *University of Maryland (College Park)*
Warren T. Norman, *University of Michigan (Ann Arbor)*
Michael Ross, *University of Waterloo, Waterloo, Canada*
Michael F. Scheier, *Carnegie-Mellon University*
Barry R. Schlenker, *University of Florida*
Kelly G. Shaver, *College of William & Mary*
Phillip Shaver, *New York University*
Marvin E. Shaw, *University of Florida*
Steven J. Sherman, *Indiana University*
Harold Sigall, *University of Maryland (College Park)*
Mark Snyder, *University of Minnesota (Minneapolis)*
Melvin L. Snyder, *University of Texas at Austin*
Shelley E. Taylor, *Harvard University*
Stephen G. West, *Florida State University*
Steve Worchel, *University of Virginia*
Mark P. Zanna, *University of Waterloo, Waterloo, Canada*

Assistant to the Editor

Sally Miller

APA Journal Staff

Anita DeVivo, *Executive Editor*
Ann I. Mahoney, Manager, *Journal Production*
Marjorie M. Henry, *Production Supervisor*
Robert J. Hayward, *Advertising Representative*
Juanita Brodie, *Subscription Manager*

The Journal of Personality and Social Psychology publishes original reports of research, methodology, theory, criticism, and literature review in the fields of personality and social psychology of normal processes and populations. Submitted manuscripts should delineate an important question and provide a clear answer to it. In the case of research reports, this will often require demonstration that the reported results are generalizable beyond a single research setting. Consult the editorial in the January 1976 issue (Vol. 33, pp. 1–7) for a detailed statement of policies and instructions to authors.

Manuscripts: Submit manuscripts in quadruplicate, prepared according to the *Publication Manual of the American Psychological Association* (2nd ed.). Copies should be highly legible and will not be returned to authors unless they are original typewritten copies. Mail manuscripts according to instructions on p. 1439 of the August issue.

Required supplementary information: Manuscripts cannot be reviewed for publication unless the author provides supplementary information in the following categories: (a) a cover letter that indicates the relation (or the lack of relation) of the submitted manuscript to other papers by the same author(s) that are currently submitted, in press, or recently published; (b) one copy of each such related paper; (c) one copy of a statement of the procedures of the research reported in the submitted manuscript in sufficient detail to serve as a guide for conduct of an exact replication; and (d) one copy of the summary tables (e.g., correlation tables, analysis of variance summaries) for the *major* analyses reported. If it appears that the supplementary materials in categories (c) or (d) might be bulky, the author should exercise judgment and restraint in selecting what to send. It is not necessary to send supplementary material that repeats verbatim manuscript content.

After a manuscript is accepted, authors will be expected to provide assurance that their data are available in unanalyzed form for reanalyses by other investigators and that their research procedures were consistent with the American Psychological Association's published principles of research ethics, except insofar as may be detailed otherwise in the manuscript.

Copyright and Permission: All rights reserved. Written permission must be obtained from the American Psychological Association for copying or reprinting text of more than 500 words, tables, or figures. Permission is normally granted contingent upon like permission of the author, inclusion of the APA copyright notice on the first page of reproduced material, and payment of a fee of $10 per page, table, or figure. Abstracting is permitted with credit to the source. Permissions and fees are waived for authors who wish to reproduce their own material for personal use; fees only are waived for authors who wish to use their own material commercially. Permission and fees are waived for the photocopying of isolated articles for nonprofit classroom or library reserve use by instructors and educational institutions. Libraries are permitted to photocopy beyond the limits of U.S. copyright law: (1) those post-1977 articles with a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through the Copyright Clearance Center, P.O. Box 765, Schenectady, NY 12301; (2) pre-1978 articles, provided that the per-copy fee stated in the Publishers' Fee List is paid through the Copyright Clearance Center, P.O. Box 765, Schenectady, NY 12301. Address requests for reprint permission to the Permissions Office, APA, 1200 Seventeenth Street, N.W., Washington, D.C. 20036.

Subscriptions: Subscriptions are available on a calendar year basis only (January through December). Nonmember rates for 1979: $60 domestic, $64 foreign, $6 single issue. APA member rate: $20. Write to Subscription Section, APA.

Change of Address: Send change of address notice and a recent mailing label to the attention of the Subscription Section, APA, 30 days prior to the actual change of address. APA will not replace undelivered copies resulting from address changes; journals will be forwarded only if subscribers notify the local post office in writing that they will guarantee second-class forwarding postage.

Reprints: Authors may order reprints of their articles from the printer when they receive proofs.

Back Issues and Back Volumes: For information regarding back issues or back volumes write to Order Department, APA.

Microform Editions: For information regarding microform editions write to any of the following: Johnson Associates, Inc., P.O. Box 1017, Greenwich, Connecticut 06830; University Microfilms, Ann Arbor, Michigan 48106; or Princeton Microfilms, Princeton, New Jersey 08540.

Published monthly in one volume per year by the American Psychological Association, Inc., 1400 North Uhle Street, Arlington, Virginia 22201. Printed in the U.S.A. Second-class postage paid at Arlington, Va., and at additional mailing offices.

Diary No. 242
Date 20.12.80
File No. Library.on...
Bureau Ednl Psy. Research.

JOURNAL OF
Personality
AND
Social Psychology

Volume 37 December 1979 Number 12

Settings of a Professional Lifetime

Roger G. Barker
University of Kansas

This is a venture in ecocentric [sic] autobiography. Behavior settings important to the author during his 50 years in psychology are identified and described, circumstances of his entrances into and departures from the settings are reported, and immediate and enduring consequences for him are set forth. The article is a microsection of the history of psychology, 1928–1978.

I was Professor of Psychology and Department Chairman at the University of Kansas in Lawrence in 1948. On a particular day of that year, I worked on a research report in my Study at Home, answered correspondence, and made some phone calls at my desk in the Chairman's Office, met my Graduate Class in Experimental Social Psychology for two hours, conferred with the Dean of the College of Arts and Sciences in the Dean's Office, chaired a meeting of the Department Faculty, and attended a Departmental Colloquium; these six behavior settings made up my professional environment. During my life in psychology, which began, I shall presume, when I entered upon graduate study at Stanford University in 1928, and has continued to this day, the stream of my behavior as a psychologist has meandered through thousands of behavior settings. These settings have constituted the environment of my professional life, and in this article I shall describe some of their features and consequences for me.

The settings of my environment have had widely varying attributes; I have entered some of them by my own choice and have been coerced in varying degrees to inhabit others; in some settings I have had considerable influence, in others little power. The Chairman's Office differed greatly in these respects from the Experimental Social Psychol-

Editor's note. This is the second article on historical aspects of social psychological research. The series was begun under the general editorship of Anthony Greenwald; the action editor for the series is Irwin Altman.

This offspring of my life has benefited greatly during its long period of gestation and labor from the skills and encouragement of Irwin Altman and Phil Schoggen. In addition, Irwin Altman had an important part in its conception, and Phil Schoggen was presiding pediatrician after its premature birth. To them I extend my most sincere thanks. However, as its sole natural parent, I must answer for its defects.

Requests for reprints should be sent to Roger G. Barker, Box 98, Oskaloosa, Kansas 66066.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3712-2137$00.75

ogy Class: in the Office there was great diversity of both human and nonhuman components (secretaries, faculty, students, typewriters, telephones, files, people conversing, dictating, typewriting, and more), and all of these were incorporated into a relatively stable, multiform pattern of Office activities. In the Class there was less diversity (no secretaries, telephones, typewriters, files, people conversing or dictating) and a radically different, less complex standing pattern of lecturing, listening, and note taking. I entered the Chairman's Office by my own choice, but I was under strong compulsion to turn up for the Class at the scheduled time. I was powerless to change the program of activities of the Dean's Office, but I was able to alter the program of the Class, although my power was by no means unlimited. This setting had been a part of the College and the University before I joined the faculty, and they imposed certain features upon it that I was powerless to alter: I had to teach experimental social psychology, the setting's raison d'être (not ecology, a great interest of mine at the time), and I had to give grades (objectionable to me, but a policy of the University). Still, within these and other limits, I was in command of the program of activities.

When I have inhabited a behavior setting —whatever its attributes, the conditions of my entrance, and my power—I have become one of its component parts, and my behavior and experience have been formed by its ongoing program; these, therefore, have changed appropriately as I have moved from setting to setting. I did not pick up my mail or dictate letters in the Experimental Social Psychology Class or give lectures or quizzes in the Chairman's Office. The impress of some settings upon me has not extended beyond their boundaries; the stamp of others has been long-lasting. The influence of the Chairman's Office ceased abruptly at the door, whereas some attributes of my behavior and experience first occurring in response to forces within the Class have remained with me. In this account, I shall pay special attention to behavior settings that had long-continuing consequences for me.

Unfortunately the data I have to work with are poor. Few contemporaneous accounts of any setting are available, and still fewer are by independent observers. Most of what I am able to report is the fallible memory of how settings appeared to one inhabitant. In an attempt to mitigate the faults of the data, I shall confine my account to those settings for which there are contemporary data or very clear memories and that appear to have been crucial to my scientific activity. Fifteen behavior settings or clusters of settings most adequately meet these requirements. Here they are listed in temporal order of their occurrence and with their institutional connections: Stanford University, 1929–1935 (Terman's Seminar, Miles' Later Maturity Facility, Stone's Animal Laboratory); University of Iowa, 1935–1937 (Lewin's Offices, Nursery School Laboratory, Topology Meetings); Harvard University, 1937–1938 (Murray's Clinic, Child Psychology Class, Boring's Sack Lunch); University of Illinois, 1938–1942 (Study at Home, Extension Classes); Stanford University, 1942–1945 (Office of Disability Survey); Clark University, 1946–1947 (Office at the University); University of Kansas, 1947–1972 (Office of Department Chairman, Field Station in Oskaloosa).

I shall introduce each of these settings with its specifications under six rubrics: program, physical attributes, temporal characteristics, human components, powers of human components, and boundary properties. Then I shall report the forces and circumstances that brought me into them, the actions and experiences they elicited from me when I inhabited them and after I left them, and my power within them.

But first I must describe the 1928 person who was to make the trip through these behavior settings.

The 1928 Person

He was unsound physically, his intellectual powers were unknown, his financial resources were meager, his motivation was strong but unfocused.

An osteomyelitis infection of the left hip and right knee had had many acute phases since its beginning 12 years before. In 1928

the disease was quiescent, but acute episodes were a threat; a flare-up kept him out of school during 1929–1930. In its quiescent periods the disease was mildly debilitative, and its destruction of joints caused some locomotor impairment.

The 1928 person was 25 years old and 3 years behind academically. He attributed this to missing school because of illness, but he did not know what it signified for a career in science. Perhaps his isolation in sickrooms and his 6 instead of 8 years of secondary and college education handicapped him for advanced study; he was poorly grounded in mathematics and languages; he had studied no physics and had had only elementary courses in chemistry and biology. He remembered many evidences of intellectual marginality and inadequacy. There was his experience in the fourth grade, when the teacher could not decide if he belonged with the A group or the B group, so she had him sit alone between the elect and the scrubs. And he could not forget failing the first course in high school algebra, having to repeat it, and edging through the second go by memorizing the answers to likely problems, ending with the unanswered question: "How can it be possible to add, subtract, multiply, and divide letters, and why do it in any case?" Years later, too late to be of any comfort to him, he discovered that George Bernard Shaw had also had trouble with algebra, "mistaking the *a*, *b*, *n* and *x* for goods" such as eggs and cheese "with the result that I rejected algebra as nonsense" (Shaw, 1949, p. 41). Shaw blamed the instruction; "I had been made a fool of"; the 1928 person blamed himself. And there was his mother's aunt, an intelligent and insightful person, who was devoted to him during his adolescence and greatly worried about his future; she summed up her impressions of his assets and liabilities with the advice that he should prepare himself to be a short-order cook in a country restaurant. There was some evidence on the other side. There was the occasional, ambiguous remark about him by partisan adults: Still water runs deep. But this was less convincing than his great aunt's explicit evaluation. There was the thrilling time in the eighth grade when he was called upon to stand before the class and explain how to point off decimals in the quotient of a long division problem, and the teacher praised him. Having graduated from Stanford University with sufficient promise for the Psychology Department to accept him for graduate study would have been reassuring had there been entrance requirements; at least he was not discouraged by Professor Terman, the chairman. By far the most persuasive evidence, which tipped the scale in his self-evaluation, was the engagement between him and Louise Dawes Shedd. Louise Shedd knew him well, she herself had high scholarly credentials and bright goals for a life in the classroom and laboratory, and she was not about to link her life with that of a short-order cook. He was willing, in fact, relieved, to accept her evaluation, and he hoped to justify it.

His financial resources were less than zero, for during his undergraduate years he had accumulated hundreds of dollars in tuition loans that became due to the University in years immediately ahead.

Psychology for him was a means, not an end. His undergraduate introduction to the subject matter and methods of the science had no special appeal to him, but he believed it to be the least onerous route to his goal of "doing good" for mankind. He was a thoroughgoing, naive idealist and ardent reformer. He considered medicine but was sure he would not be accepted as a student, and his long, discouraging experience as a consumer of medical services had alienated him from the profession. He found economics too remote from particular people, and sociology too speculative and wordy. He tried to get a handle on demography but could discover no people or publications to inform him sufficiently about it. He did not consider literature, art, or the physical and biological sciences of any relevance to his goals. Psychology seemed to be a promising route, but even as he began graduate study he did not foreclose other routes, and within psychology's varied and vast domain, he had no preference. Although the zeal of the 1928 person was great, there were clouds on the horizon. He anticipated that the psychology route to good

works would have difficult and unpleasant stretches, and he was keenly aware of the likelihood of recurring acute phases of his illness. His prescience in the latter respect was verified by the flare-up in 1929–1930, and he hesitated about persisting, but for reasons impossible for him to comprehend to this day, his fiancee stood fast, encouraged him, and even married him in the summer of 1930. So in the autumn of 1930 he made a second start and became an anxious, vigilant, expectant psychologist in the making.

Stanford University, 1929–1935

I was a graduate and postgraduate student at Stanford University for 6 years. For 4 years I attended classes, did research under supervision, led "quiz sections" of elementary psychology, did various departmental chores (mimeographed class notes and exams, scored Strong Vocational Interest Tests, and cared for the animal colony), passed examinations, and received the MA degree (1930) and the PhD degree (1934). Being unable to find a regular position in the depression year of 1934, I was fortunate in being kept on as a research associate for 2 more years. The work was rewarding and the salary of $100 a month was satisfactory; with Louise's $1,800 a year as high school biology teacher we were able to pay my tuition loan debts to the University and save some money. Three behavior settings were crucial to me during these Stanford years.

Terman's Seminar

Program: Report by seminar member, usually a student, with interruptions by others discussing or questioning points made. *Physical attributes:* Locus, Terman's home on the Stanford University campus, in an attractive living room of sufficient size for a single-row circle of 25 persons. *Temporal characteristics:* Weekly meetings (with occasional skips) for about 2 hours during most autumn, winter, and spring terms; approximately 20 meetings most years. *Human components:* Dr. Lewis M. Terman, one or two other faculty members, occasional guests, 10–20 graduate psychology students. *Powers of human components:* Program determined at three levels of power. Terman's power extended over the entire setting (he selected the speaker, approved the topic, opened and closed the meeting); the speaker's power was supreme over the content and method of his report; the power of the other inhabitants was limited to supplementing, correcting, criticizing, and approving the speaker's presentation. *Boundary properties:* Admission at the invitation or urging (in the case of students) of Dr. Terman.

Reports usually described completed research or plans for research; occasionally books or monographs were reviewed. Almost all reports were concerned with research, emphasizing methods; theories or general issues were very rarely seminar topics. Interchanges between reporters and other members were usually for more information about procedures. The Seminar was eminently civilized and controlled, the occasional disagreements and arguments being muted. Terman set the tone, being quietly attentive, making few contributions himself other than introducing the speaker and ending the session with general remarks. Despite its appearance of tranquility, many students reported, outside the Seminar, that it was a tense, even traumatic experience for them. Many of the controlled interactions were undoubtedly reactions to the Seminar as a dangerous place for students; they were on trial before powerful present and future evaluators.

When Terman opened the Seminar to me, I eagerly attended. The name of the game for me was to learn as much as possible, to discover if I was likely to make the grade in scientific work, and to improve my chances of doing so by demonstrating what strength I had to Terman, the other faculty members, and my peers. The first and last intentions were not congruent. Optimal learning requires acknowledgment of ignorance, whereas putting one's best foot forward requires masking ignorance. And to walk on thin ice, as I did in the Seminar, always on the brink of both disaster and improved footing, is not conducive to a wide perspective and clear thinking; it is inhibiting. Still, one must act; to remain motionless on the ice, or dumb in the Seminar, is a sure way to disaster. Alert caution is essential. Under these circumstances motives and abilities and disabilities are not openly revealed, and valid judgments of persons are not possible, though reputations are made. I never knew my reputation among the

nembers of the Seminar, but I did know my elf-estimate. I placed myself near the median f the student seminarians, I was uncertain of assing the examinations, and I was quite ure my destiny was at best to become a ournеyman psychologist, perhaps in a junior ollege where, after all, there was plenty of cope for good works. Considering the people gainst whom I was judging myself, this self-stimation is not surprising; during the years f my attendance they included six who be-ame president of the American Psychological ssociation and a number of others who were onored by the Association for their scientific ontributions.

My immediate reaction after the Seminar as frequently one of great dismay: at my norance of what seemed to be common nowledge, at my failure to contribute what fterward appeared to me to be a valuable put, and at my espousal of weak, foolish, or relevant ideas. Louise came to dread the te-night aftermath of the Seminar.

I cannot now remember what I learned; I m quite sure that I encountered no mountain eaks of new understanding. But the Seminar rovided a regular update of what was hap-ening in psychology, especially with respect methods of investigation. More than this, became acculturated into the language, the ores, the ethos of the psychology tribe. And made acquaintances who have continued to e valued personal friends and influential onnections within the profession. On the egative side, I first encountered an aspect of cademic life that has continued to be dis-asteful to me: the ubiquity of judging others passing and failing students, approving and ot approving candidates, promoting and not romoting colleagues, accepting and not ac-epting research proposals, and so forth) and f being judged by others. The fact that most nterpersonal relations within Terman's Sem-nar involved, explicitly or implicitly, judg-nents of personal worth reduced its attraction s a social occasion and eroded its educational enefits.

Miles' *Later Maturity Facility*

Program: To study change in abilities from middle to later years of life. *Physical attributes:* Located near the central business district of Palo Alto in a one-time residence. The former reception hall, parlor, living room, dining room, and bedrooms were modified for use as offices and for administering tests and experiments; office furnishings and equipment for tests and experiments. *Temporal characteristics:* Operated weekdays for several months during academic year 1931–32 from 9:00 a.m. to 5:00 p.m. *Human components:* Walter R. Miles, graduate student research assistants, secretary, subjects. *Boundary properties:* Free access by staff; subjects admitted by appointment. *Distribution of power:* The power of Miles extended over the entire setting and program; he planned the research, secured the financing, selected the staff, recruited the subjects, approved the methods of testing and experimentation. The power of each research assistant was dominant within his laboratory room. The only power of subjects was to refuse to participate in a particular procedure; to my knowledge this did not occur.

I had completed my master's degree work under Miles with a thesis on finger maze learning, and it was now time to undertake research for the PhD degree. So when Miles returned from a trip East, bringing the news of a grant for the magnificent sum of $10,000 to support a study of old age, I eagerly embraced the opportunity to participate. For reasons I cannot recall, I chose to study muscular fatigue. Miles gave me complete freedom to devise a muscle fatigue test suitable for both robust 50-year-olds and frail centenarians, and I took considerable satisfaction in adapting a spirometer to determine hand and arm fatigue quickly. Other graduate students investigated changes in intelligence, learning, memory, motor skills, and sensory acuity. Subjects were recruited from clubs, churches, and living groups by paying the organization for each member who became a subject. They were brought to our experimental rooms by the secretary. At the time, I did not appreciate the luxury of having subjects provided to me with no effort on my part; many times since, I have appreciated Miles' efficient logistics.

I greatly enjoyed working within the Later Maturity Facility; the freedom within my own domain, and the straightforward problem were agreeable. I went home most nights feeling I was making progress toward providing some potentially useful new knowledge.

The more enduring consequences of this setting for me were, first, increased confidence

in my ability to engage in research on my own; after approving the area of my contribution, Miles took no part in my project; it became my own. Second, I learned something about the possibilities and difficulties of a kind of field study; I observed the extensive public relations activities required to secure subjects, and I was especially impressed with the effectiveness of approaching citizens by way of the organizations to which they were devoted. I also noted that the demands of public relations, at which Miles was very effective, divorced him from the details of the research. Because of this experience, I believe it was easier for me, when I later engaged in field studies myself, to realize that they require staffing different from that of laboratory investigations with captive or hired subjects and that community experts are as essential as observers, interviewers, testers, and so forth.

Stone's Animal Laboratory

Program: To investigate motivation in animals, chiefly rats, with emphasis on sexual behavior. *Physical attributes:* Locus, third-floor attic of the building housing the psychology department in an area partitioned into a sky-lighted animal room for about 100 animal cages, perhaps 10 small rooms for assistants and equipment, and Stone's large office. The equipment was primitive by present standards: student-made mazes, jumping apparatus, activity drums, Monroe hand calculators, simple shop equipment. *Temporal characteristics:* Functioned at some time every day. *Human components:* Calvin P. Stone, one or two postdoctoral fellows, two graduate assistants, several thesis students, a few students doing class projects, visitors. *Powers of human components:* Stone's power extended over the entire laboratory; he approved the fellows, selected the assistants, admitted the students and approved their research, set rules and standards. Fellows had complete power over their own research; assistants controlled their particular segment of the total program of the laboratory after consultation with Stone. Visitors' power was limited to viewing approved areas and asking questions. *Boundary properties:* Fellows, assistants, and students admitted by Stone personally; visitors were tolerated, but except for professionals, they were not welcomed.

As in 1931 (the opportunity to participate in the Later Maturity Facility came at a crucial time), I was fortunate in 1933 to be able to work in Stone's Animal Laboratory when I could not find a regular job. I did not choose to spend two years in this setting, but neither did I object; it was inevitable.

I found work in the setting even more satisfying than in the Later Maturity Facility. I had passed all examinations and written my thesis; that tension was gone. The intellectual camaraderie was congenial; postdoctoral fellows brought news and innovations from other laboratories; the pace of activities was not determined by tightly scheduled subject appointments. But there was a vigorous, steady program of activities; almost all the animal work required regular cleaning, feeding, observing, testing, and examining; Stone was a hard, regular worker himself, and everyone knew that he valued industriousness right along with honesty and intelligence. He was known to disapprove strongly of the policy of another university that reportedly provided technicians to do such detail work for graduate students as sectioning and preparing tissues for examination and running rats in mazes. In Stone's Laboratory, students cared for their own animals and spent the boring hours putting them through the necessary procedures. Inasmuch as Stone was in and out of the laboratory many times most days, beginning at 7:30 a.m., his power and reputation were sufficient to maintain a fairly tight ship without many definite rules and regulations.

Special emphasis during this period was on variation in the age of sexual maturity of rats, its genetic basis, and its relation to other behavior and somatic characteristics. Parallel investigations were conducted of the age of maturity of human females and its relation to their size and their interests and attitudes. I came to find much of the work interesting, and some of it had a permanent influence upon me. Even the evening after evening of testing young male rats for age of first copulation became enjoyable when Louise came along and read aloud from a high stool in the center of the animal room. (Conrad, Tolstoy, and Galsworthy in no way diverted the rat subjects from their putative activities.)

It was here that I first became aware of the potential value of archival data for psychological science. Stone was interested in the

intellectual level of children of abnormally early sexual maturity, and he had previously reviewed and summarized the literature. He set me to updating the survey, and I was impressed to find that although most evidence was in the form of reports of single or a very few cases, and none were adequate methodologically in terms of number and selection of cases and methods of testing, still, en masse, these data, reported by many independent investigators at widely varying times, in diverse situations and cultures, provided overwhelming evidence that early intellectual maturation does not accompany abnormally early sexual maturation. I noted that almost no cases were reported in the psychological literature, whereas medical journals served as valuable archives for these rare cases. I began to see that the insistence of academic psychology that every publication conform to all current canons of scientific adequacy was depriving the science of data on important issues. Years later I discovered a pair of identical adolescent twins, one of whom had been seriously crippled since childhood, and I saw this as an exceptional opportunity to obtain data from a very rare natural experiment on the effects of physique on personality; I studied them extensively. But I was advised not to attempt to get the material published because "one case means nothing." I wondered: Would a paleontological journal reject a study of 1 dinosaur egg because the investigator did not present data on 50 eggs? The lesson I began to learn in Stone's Laboratory has stayed with me, and I and my colleagues have not hesitated to collect "inadequate" data under certain circumstances, namely, when the problem to which the data refer is important, when fully acceptable data cannot be obtained due to their temporal or physical dispersion or their rarity, and when the data are such that investigators at other times and in other places can add to them.

It was in Stone's Laboratory that I first encountered a set of problems that have occupied me ever since, namely, environmental influences on behavior. At the time, however, I did not see the particular problem in this context. It arose from our studies of relations between age at menarche and physical and social development in girls. We discovered that in the early adolescent years, early maturing girls are more similar to older females in measurements of physique than are late maturing girls of the same chronological ages, and when we asked if this same relation holds for social interests and attitudes, we found that it does: Postmenarcheal girls 12, 13, and 14 years of age are more similar to older females in their responses to interest and attitude test items than are premenarcheal girls of the same ages. So we faced this question: To what degree are the differences in attitudes and interests due to direct hormone influences, and to what degree to the fact that the physiques of the early and late maturing girls provide the girls and their associates with stimuli with different social significances, thereby imposing different social environments upon them? We could not answer this question with the data at hand. Five years later, at the University of Illinois, I returned to this problem and found evidence that powerful adults (parents, teachers) bring greater pressure upon physically mature girls to engage in mature behavior than upon physically immature girls. And 10 years later, when I was back at Stanford, I returned to the problem in connection with studies with my students of the psychological consequences of physical crippling.

In addition to initiating this particular continuing interest, Stone's Animal Laboratory further strengthened an attitude I carried with me from other Stanford settings, namely, that strong and persisting but rather narrow programs of activities are productive. There were no brilliant performances at Stanford. The research of Terman on the gifted, of Stone on animal motivation, and of E. K. Strong on interests yielded no remarkable breakthroughs, but they were and have continued to be recognized as substantial achievements. This was not an "Aha!" experience for me; but as I look back I can see that it became a deeply rooted conviction that this was to be my way of doing science. Along with this vaguely developing insight there was increased self-assurance. At the end of my 6 years at Stanford, I was more sure that my aspirations for a productive career in science

were not hopeless. However, I did not have a strong commitment to any problem. I had worked on a rather wide range of problems via a considerable number of methodologies, but I had become devoted to none. I was still unfocused, at the beck and call of almost any opening that would allow me to earn a living as a psychologist. In the spring of 1935 an opening came that radically altered my intellectual lifestyle and threw me again into uncertainty about myself.

I took part in many behavior settings at Stanford other than the three I have described: classes, lectures, seminars, laboratories, projects, examinations. They enriched my intellectual life and widened my perspective, but only one of them had particular, identifiable consequences for me. This setting channeled the stream of my behavior as a psychologist into its first great bend. Kurt Lewin came as a visiting professor for the year 1932–1933. I was very busy during this year finishing my doctoral thesis so I was able to attend Lewin's class only as a visitor. He attracted me greatly as a person, but his psychology confused me or, perhaps more correctly, it was incomprehensible to me. Fairies and ectoplasm would have been more comprehensible than life space, valence, psychological force, inner-personal regions, substitute value, psychological satiation, and so forth. Stanford students knew something about Gestalt psychology experiments and theories of perception from the writings of Wertheimer and Köhler, and of development from the works of Koffka, but Lewin's so-called Gestalt psychology seemed to have nothing to do with these, as we expected. And equally disconcerting, Lewin seriously reported experiments with seven subjects, and instead of replicating the experiments, he changed the conditions; even worse, if the results from the altered conditions were in accord with his predictions on the basis of theory, he took this as a verification of the findings and the theory. I suppose Stanford University in those days was among the least auspicious places in the United States for an understanding of Lewin; theory was almost a nonword in the psychology department, although we did use it in connection with Spearman's interpretation of intelligence test intercorrelations (theory of general intelligence) and we read about psychoanalytic theory. But most of us had no background in the philosophy of science, and the place of theory in science. So Lewin's *Dynamic Theory of Personality*, which I reviewed (with Terman) and which he taught in his course, was a transient foreign body, a UFO, to most of us. I cannot recall that any of the students who attended his class took and retained a serious interest in his viewpoint unless they had later association with him. The gulf was too wide to be bridged quickly, and the dissonance was so great that some rejected his ideas out of hand. I was not negative; I was tolerantly baffled.

So when the opportunity arose to join Lewin as a General Education Board Fellow at the University of Iowa for the year 1935–1936 I was intrigued; here was a chance to learn a different kind of psychology, to earn a full living for Louise and me for the first time, and to have the prestige of the fellowship. Indeed, I had no choice; circumstances beyond my knowledge and control were turning the stream of my professional behavior, indeed my whole life, in a new direction. I have no idea to this day how my nomination to this fellowship came about, but I am sure that an important factor on the Stanford side must have been: It is really time for Barker to push off the home place. I thought so, too.

University of Iowa, 1935–1937

Kurt Lewin went to the Iowa Child Welfare Station in the autumn of 1935 on a grant from the General Education Board, with provision for three research assistants. The first assistants were Tamara Dembo (who had been a student of Lewin in Berlin), Herbert F. Wright (who had been a student of Donald K. Adams at Duke University; Adams had studied with Lewin in Berlin), and me. My fellowship was extended in the spring of 1936 for 1 more year. Three behavior settings were of primary importance to me at Iowa.

Lewin's Offices

Program: To plan the research project we would undertake, and to discuss and make decisions about

rocedures and problems when it was underway; to nalyze data from the project and write up reports; o read the writing Lewin was doing, to listen to his roposals for reformulations and additions, and to riticize and make suggestions. *Physical attributes:* ocated in four adjacent office rooms on an upper loor of East Hall, University of Iowa campus; he building had been a hospital and the rooms were ormer single-patient rooms; they were a little rowded with a desk, a bookshelf, a blackboard (in ewin's office), and four chairs when we were all resent in a single office. *Temporal characteristics:* n operation at some time almost every day, with ne or more participants; sessions with Lewin oc- urred whenever the project needed his attention and henever he had something to discuss. *Human com- onents:* Lewin, the three assistants, frequent vis- ors. *Boundary properties:* Assistants selected and dmitted by Lewin; free entrance for visitors. *owers of human components:* Lewin established he program of the setting; he decided that the re- arch would be upon frustration, but the details of rocedure were worked out in group consultation, here he was first among equals; he was, of course, total control of his own writing, the assistants erving more as persons upon whom he could try out is ideas than as consultants or collaborators, al- hough he took seriously all criticisms and sugges- ons.

It is difficult to imagine more different cli- nates for scientific work than those obtaining t Stanford and at Iowa. At Stanford, science athered facts about behavior via experiments nd tests, determined their central tendencies, nd analyzed their interrelations; in Lewin's etup, science explored ideas about behavior ia experiments and observations. At Stan- ord, conclusions were in terms of the means, ispersions, and correlations of samples of *acts* about behavior; my own work discov- red that the rate at which 80-year-olds pump ir into a spirometer declines faster during a ork period than the rate at which 50-year- lds pump air. At Iowa, conclusions were in erms of verified or altered *ideas* about be- avior; in our work there we found (in ac- ordance with a theory) that in psychological rustration, inner-personal systems can be con- idered to be in a state of blocked tension that mounts to functional dedifferentiation, one xpression of which is lowered level of intel- ectual activity, that is, intellectual regression.

In the beginning, the sessions in Lewin's ffice were an ordeal for me; they were be- wildering and tiring. The ideas we were to explore and clarify in connection with the frustration study were too unfamiliar for me to make many contributions; as for the monograph Lewin was working on, *The Conceptual Representation and Measurement of Psychological Forces,* the ideas were utterly baffling. Lewin's eagerness and the energy to back it up seemed boundless, whereas the tension and alertness of the 2-hour sessions of Terman's Seminar left me dog-tired; after meeting with Lewin, Dembo, and Wright from 2:00 to 7:00 in the afternoon (with Lewin reading aloud what he had dictated the day before, interlining new sentences, rearranging the order, violently objecting to a criticism by Dembo, turning to Wright—"Herbert, is she right?"—accepting Dembo's criticism, diagramming a relation on the blackboard, crossing the whole page out, dictating a new version to Dembo, and so forth), I was ready to drop. After five o'clock I would hope beyond hope that my dear, pregnant wife, lonesome at home, would telephone that I was urgently needed. Sometimes she did. A frequent concluding remark by Lewin was "We must think about this," and that after 3, 4, 5 hours of nothing else. Did this add up, perhaps, to a kind of brainwashing? In any case, as the months went by, I began to understand Lewin, and his ideas have remained at the center of all my subsequent work. But equally important to me has been the new, higher level of intellectual effort to which I became adapted. I could never come close to Lewin's intensity; I had to take it much slower, but his refrain, "We must think about this," has stayed with me. Although no one could have been more subordinate to Lewin in terms of knowledge, I was always treated as a colleague, never as a pupil. My contributions, however naive, were always taken seriously, Lewin often seeing in them more than I had intended.

Nursery School Laboratory

Program: To carry out the research on frustration, we planned in the offices in East Hall. *Physical attributes:* Two ground-floor rooms in a remodeled residence across the street from the station's nursery school equipped with a one-way observation booth, child-sized chairs and tables, toys, and a movable barrier in accordance with the research design. *Temporal characteristics:* Setting occurred intermittently during winter and spring of 1936–37, according to a

schedule of appointments. *Human components:* Tamara Dembo and myself, alternating as observer and experimenter, and a single child from the nursery school at each occurrence. *Boundary properties:* Free access to Dembo and me; child subjects, a different one on each occasion, selected and required to attend in accordance with decisions made by us and nursery school staff. *Powers of human components:* Dembo and I had joint authority over the entire setting, although, according to the experimental design, we did not intrude into regions designated for free play; child subjects had no power except in designated free play regions.

Basic to much of the early research Lewin initiated is the theory that psychological tension systems correspond to intentions to engage in molar actions. The studies of interrupted and substitute tasks issued from this theory. Lewin came to Iowa with the intention of investigating this idea further; he thought that the undischarged tensions that occur in frustration result, via increased rigidity and spread of tension to adjacent regions of the person, in functional dedifferentiation, and that one manifestation of this is intellectual regression to a state normal for an earlier age. The problems for the assistants were to devise a frustrating situation and methods of assessing intellectual level inside and outside this situation. The nursery school setting in which we worked on these problems elicited from me two insights that have remained with me. I discovered the value, for studies of molar actions, of situations where the investigator intrudes not at all or very little, and the value of detailed narrative records of behavior. The frustration study involved, for the most part, a continuation of the child's usual nursery school day, and in his free situation, we were able by means of nonintrusive narrative records to assess intellectual level as accurately as formal intelligence tests do. In later work, my colleagues and I have come to depend very extensively on narrative records of behavior in situations that are completely free of our influence; and we continue to be impressed with the value of ordinary language as a coding system for the subtleties and complexities of behavior. The Nursery School Laboratory was a welcome refuge for me from the conferences in Lewin's Office. Here I found a kind of therapy in applying my Stanford expertise to the gathering and analyzing of masses of data.

Topology Meetings

Program: To discuss Lewinian theory and research findings. *Physical attributes:* Occurred in academic conference rooms of various universities. *Temporal characteristics:* Took place during Christmas vacation for two or three full-day and evening sessions. *Human components:* Lewin and invited guests; the latter were former and present students and colleagues, and a few others interested in Lewin's ideas. *Boundary properties:* Invitations issued by Lewin and by people at host school. *Powers of human components:* Designated persons at host institution controlled local arrangements, Lewin arranged program of papers; free discussion.

The Topology Meetings were as much a part of my Iowa experience as were Lewin's Office and the Nursery School Laboratory, even though the meetings were held at Cornell University and Bryn Mawr College in those years. They agitated me somewhat as Terman's Seminar had, although the testing aspect was less pervasive and the learning aspect more pervasive; still these settings were new and strange, and I was alert and cautious. In them my world expanded greatly, to encompass many new friends and acquaintances, many new concepts applied to new problems, and many new locales and institutions (I had not previously been east of the Mississippi).

The adjustments I made to the three crucial behavior settings at Iowa were only part of the intellectual turmoil I experienced. There were other upsetting settings. In some of them Herbert Feigl expounded physicalism and the views of the Vienna Circle; in others, Spence's kind of Hullian behaviorism was expounded. Iowa, for that brief time, at least, was a bubbling cauldron of antagonistic ideas. With my nontheoretical background, I was severely buffeted by the strongly asserted convictions of these sophisticated advocates. Lewin's handling of these disturbances impressed me and has been a model for me since. Although he defended his viewpoints strongly, he was not rigidly partisan. An early remark he made to a student objector at Stanford, when his English was poor, expressed his position:

Con be, but I sink absolute uzzer." He did ot think current controversy would settle asic issues, that only empirical evidence ould in the long run sift the true from the lse, and that the business of science was to et ahead with empirical tests. He strongly xplored the disruptive partisanship within erman psychology.

Before the end of my second fellowship ear a new opportunity arose that I could not ecline: an instructorship at Harvard Uniersity to teach a course in child psychology. s with the General Education Board Fellowhip, I have no knowledge of the forces that rought this opportunity to me. But again, fter 9 years, the stream of my professional ehavior was to flow in a ready-made chanel. Previously, the Later Maturity Facility, he Animal Laboratory, and the General Edcation Board Fellowship had taken charge f my life in psychology, and now the Harvard sychology Department took over with no ffort by me. But there was a difference: Whereas previously I had been in the role of upil (to Terman, to Miles, to Stone, to ewin), now I was to be on my own in my eaching and in my research.

Harvard University, 1937–1938

I arrived at Harvard under the cloud of an lness, not my old familiar trouble, but susected appendicitis. I presume now, more han I admitted then, that the symptoms were f psychosomatic origin. And why not? I was onfronted with the task of teaching a subject ew both to me and to the Harvard Psychology Department (Child Psychology had not een in its curriculum before) to hypercritcal Harvard students (Boring warned me of hem in a letter in which he also expressed he concern of the Department about its repitation among undergraduates for good teachng) before the eyes of some of psychology's op brass (Boring, Allport, Murray, Lashley) nd its brightest lieutenant colonels (S. S. tevens, B. F. Skinner, Robert W. White).

Murray's Clinic

Program: To carry out studies of personality; the program was minimal this year, however, as Murray was on sabbatical leave. *Physical attributes:* Located in a converted residence on Plimpton Street, a block and a half from Emerson Hall, headquarters for the Department; clinic office, staff offices, library, experimental rooms, shop, kitchen; the library and several offices were elegantly furnished with fine rugs and period furniture. *Temporal characteristics:* In regular operation weekdays and at any time for particular staff members. *Human components:* Acting director, three or four staff and/or psychology department members, a postdoctoral fellow, several students, clients and subjects, secretary, shop man. *Boundary properties:* Free access to official personnel, clients and subjects by appointment. *Powers of human components:* The acting director was Robert W. White, who had final control of the entire setting, although staff members and the fellow were quite independent; the secretary and shop man were semiautonomous within their areas; powers of clients and students were limited by the staff members with whom they were associated.

Murray's Clinic was my salvation: It was leisurely and quiet; my office was large, with comfortable chairs and a couch; after the hectic pace at Iowa, it was a refuge. Here I could sit and think about where I had been and where I was going, emerging only at intervals to grapple with my one class. Part of where I had been was with me in the form of a partially completed manuscript about the Iowa frustration experiment; sitting on my desk in a big box, it was a continual irritant, but it did not claim or oppress me. I tested where I was going by doing an experiment on conflict resolution; it was a Lewinian experiment and it was congenial to me, but I did not see it as a channel to a lifetime of research; I did only one other study of conflict. I believe I became dimly aware, then, that further investigation of conflict resolution would lead to finer grained and more complex theories and more detailed and precise experimental procedures that did not suit me. This may have been the beginning of my aversion to the reduction of theoretical explanations of molar events to theories of their more molecular components, with the implication that the latter are the more fundamental. Murray's Clinic was my first experience of a behavior setting without a coercive program. Whereas the settings at Stanford and Iowa took me in hand and put me through their paces, Murray's Clinic protected me from impositions. Settings of this kind have fortu-

nately occurred at regular intervals in my career.

Child Psychology Class

Program: To teach child psychology to undergraduate students. *Physical attributes:* Classroom for about 40 students in the basement of the Clinic; chairs, podium, blackboard; furnishings and decorations somewhat shabby. *Temporal characteristics:* One hour's duration, three times weekly for one semester. *Human components:* Instructor and 35 students. *Boundary properties:* Access only for instructor and registered students. *Powers of human components:* Program under complete control of instructor; powers of students limited to expressing approval or disapproval.

The sensitivity of the Psychology Department to the students' regard for the teaching it provided caused me to approach this, my first teaching, with more than my usual trepidation. I worked hard welding my Stanford and Lewinian background into what I am sure was a unique course of study. There were a few rough places in this setting, marked by some foot shuffling, but it ended with an acceptable round of applause. So I emerged with some confidence in my teaching ability and with a course outline that was a long-time asset.

Boring's Sack Lunch

Program: To combine an economical lunch with friendly conversation among colleagues. *Physical attributes:* Locus, Boring's office in Emerson Hall; table around which eight people could sit comfortably; sack lunches; in the background a desk and shelves filled with books. *Temporal characteristics:* 1:00 to 2:00 p.m. most weekdays. *Human components:* Boring and, usually, five or six staff members and visitors. *Boundary properties:* Access at Boring's invitation. *Powers of human components:* Boring's benign influence pervaded the setting.

I was flattered to be invited. I expected stimulating conversation of the kind I had heard took place at the high tables in the English colleges, but in this I was disappointed. Despite the high caliber and great achievements of a number of the attenders (B. F. Skinner, S. S. Stevens, Boring, occasionally Lashley and Beebe-Center), small shoptalk about equipment, library resources, particular research results, publishing problems, together with bantering but restrained gossip about current university, community, and national affairs prevailed. I emerged from this setting with some impressions of eminent and soon-to-be eminent people, but nothing of professional value to me. It did strengthen my belief, originating at Stanford and strengthened at Iowa, that the effective men in the science were blue-collar, hard-hat workers, not gentlemen scholars. Here was Boring, well past middle age, the Mr. Psychology of the profession in America, eating his cold egg sandwich with the boys and discussing the nuts and bolts of research and writing, not lunching gracefully in the Harvard Faculty Club with his professional and administrative peers on the Club's famous horsemeat steaks.

My time at Harvard was one of greatly expanded perspective, not of psychology as a field of study, but of the people, institutions, and cultural milieu of the science. Other settings that contributed to this expansion were conferences sponsored by the Macy Foundation, where I encountered the eastern wing of the child development movement and for the first time directly heard psychoanalysts expound their views.

Although I was no longer in the role of a student, I was still a probationer. As I understood the Harvard policy, it was: 3 years at most as an instructor and then up or out. As I saw no possibility of my moving up, I determined before the end of the year to seek another position. Behind this decision there were a number of considerations: uncertainty if I would be appointed for a second year and apprehension about finding a place in 2 or 3 years, when I would need one; the belief that 9 years as apprentice (6 years at Stanford, 2 at Iowa, 1 at Harvard) with the continual strain of being tested was enough; and the realization that I was veering sharply away from my goal of using psychology for the direct benefit of people. So when I heard that a position as Assistant Professor of Educational Psychology was open in the College of Education at the University of Illinois, I applied for it and was appointed. As it turned out, another year at Harvard was offered to me, and some surprise was expressed that I

chose Illinois. But Louise and I went with few regrets; we were at last taking a hand in the direction of my professional career, our economic future did not come to a dead end in 1 or 2 years, and education was surely a promising place to apply what I had learned at Stanford and Iowa.

University of Illinois, 1938–1942

We arrived in Urbana, Illinois, in July 1938 with our 2-month-old and 2-year-old children. We rented a house while looking for one to buy; we were ready to settle down.

Study at Home

Program: To get on at last with my own version of psychology. *Physical attributes:* We soon purchased an old house from a retired academic who had made himself a study to match, on a small scale, the library on Plimpton Street: paneled walls, a fireplace that worked, many shelves, stained glass windows, space for a large desk and work table. *Temporal characteristics:* Occurred at my discretion during the Illinois years. *Human components:* Myself, my students, Louise, and our children. *Powers of human components:* I was in overall charge, but in these years Louise began to participate in the research, a partnership that has continued, with some breaks when she has had outside jobs. *Boundary properties:* Penetrable to me, Louise, and children at any time; to students on invitation.

What I had prepared for the Harvard students of child psychology seemed good enough also for the Illinois students of educational psychology, so I turned early from preparations for my class to other undertakings. Two of these were rooted in Stone's Laboratory, two in Lewin's Offices, and one was a "do good" effort.

The question of whether differences in the attitudes and interests of premenarcheal and postmenarcheal girls have a basis in the social significance of their different physiques was left unanswered 5 years earlier, so I returned to it now and found evidence that, indeed, the adult associates of mature adolescent girls provide a different (more "mature") social environment for them than they do for immature girls. And I expanded my concern for the environmental significance of physique to physically disabled people and asked if they also live in a different environment from that of physically normal people by reason of their physiques. I began a long involvement with this question with two case studies. Although the roots of my interest in the environmental significance of physique were in Stone's Laboratory, my interpretations and theories came from Lewin's Office, and undoubtedly my expansion of this interest to physical disability was influenced by my personal experience with the problem.

The unfinished manuscript of the frustration research still occupied a prominent place on my desk, and I continued to wrest time from more immediately interesting new tasks to inch along toward its completion. And the conflict resolution study had become such an uncompleted task for me that I was impelled to replicate (and verify) the findings by an entirely different method.

My "do good" effort was carried out in collaboration with Herbert F. Wright, my fellow Fellow at Iowa, and Jacob S. Kounin, my new colleague at Illinois. This was a collection of reports of psychological research in the field of child development. It was intended to provide a "bookshelf" of primary research for students of child development and thereby promote scientific child study as a field for both practitioners and scientists. It is my impression that this was the first "reader," as they are now called, in child psychology, and perhaps in psychology as well (Barker, Kounin, & Wright, 1943).

My Study at Home was not only a place to get on with my version of psychology as a science; it became a place to teach my version, as well. In the early autumn of 1940 an acute phase of the bone infection occurred, and I was completely incapacitated for the term. Jack Kounin came to my rescue, taking over my classes, presumably for a short time. But as it turned out, I was out of commission for 4 months. Due to Jack's heroic efforts and the tolerance of the College and University, my pay continued. In the second semester, although still in a full body cast, I was able, again with the kind indulgence and special arrangements of the College, University, and students, to meet small classes in my Study at Home. Perched on a high tavern-type chair

within a specially built surrounding pulpit, I was able to expound the word to a score of students at a time.

Again, as with Murray's Clinic, my Study at Home saved me from possible disaster. This one nourished projects I brought to it from Stanford and Iowa, and in it new, long-continued undertakings originated.

Extension Classes

Program: To teach educational psychology at the graduate level to primary and secondary school teachers. *Physical attributes:* Located in classrooms of state teachers' colleges, which did not at that time offer graduate-level instruction. *Temporal characteristics:* Occurred on Saturday mornings for two hours during each semester. *Human components:* Instructor and 20 to 40 students from the schools of the area. *Boundary properties:* Only students satisfying registration requirements of the University were admitted. *Powers of human components:* Program controlled by instructor.

The Saturday trips from Urbana to outlying teachers' colleges, usually by automobile but in some cases by train, were enjoyable and satisfying. The mature, practicing teachers brought me into contact with real teaching problems in the towns where they taught; they educated me perhaps as much as I educated them. And the trips revealed to me something about the rural Midwest. I was intrigued with the small towns, and elsewhere I have related how the general problem that dominated the last 25 years of my professional life occurred to me as I was rushed through them by the train on the trip from Urbana in the center of Illinois to Carbondale in the south (Barker & Associates, 1978). In short, I had an overwhelming *negative* "Aha!" experience: Here I was, a native of the culture and an expert on child behavior (and especially on frustration) who knew no more about the everyday behaviors and environments of the children of the towns than laymen know. I was aware, too, that other child psychologists knew no more than I did, and furthermore, that we had no means of discovering more; no methods of determining the extent and conditions of frustration, joy, anger, success, conflict, problem solving, fear, and so forth among the towns' children. I thought how different the position of an agronomist would be. He would know or could determine the kinds, yields, and qualities of the crops we were passing, the properties of the soils in which they were growing, and the relations between soil conditions and output. This was the beginning of a growing conviction that a science that knows no more about the distribution in nature of the phenomena with which it is concerned than laymen do is a defective science, and it was the beginning of my impression that small towns of the kind I observed in Illinois and learned about from the teachers are favorable places to begin to remedy the defect. It was seven years before the seed planted in the Extension Classes and on my trips across the plains of Illinois began to sprout.

It was not long after we went to Illinois to "settle down" that irritants began to appear. There were the disagreeable administrative impositions upon my teaching that finally became intolerable to me. And I was soon to discover that my flight from a marginal position vis-à-vis the upper uppers of the profession at Harvard had landed me squarely with the lower lowers. I attempted to establish collegial relations with psychologists in the Psychology Department but without success. Part of the difficulty was structural: I was separated from them spacially, temporally, and administratively; I was in a different building, their seminars and other gatherings often conflicted with my staff meetings and other duties, and administrative messages were on different communication networks. And in addition, I found that a psychologist in a college of education is distinctly lower-class. This so impressed me and it was so important to me personally that I used my experience as evidence for an analysis of the values and dynamics involved. Finally, I became discouraged about my ability to combine scientific work with applications. I began to see that it required more than clear lectures to alter the practices of teachers. My own experience, namely, that because of impositions from the encompassing system I could not practice in my own teaching at the University, the principles that my science showed me to

be true should have made this immediately clear to me, but it took time. I struggled on with increasing dissatisfaction. Finally, in the spring of 1942, these multiple environmental stresses became so great that I wrote to Terman telling him of my disenchantment. To the surprise and delight of both Louise (who, surprisingly, did not thrive with a troubled mate) and me, Terman answered immediately with the offer of a place at Stanford for the duration of the war.

At Illinois I learned some important things about myself and about my profession; I closed out some unfinished tasks and initiated some new ones. I look back on the time there as a rough but beneficial passage.

Stanford University, 1942–1945

The Stanford appointment was as Acting Associate Professor. I understood clearly that this was an emergency appointment to bolster a war-depleted staff. At the time its temporary nature did not trouble me or Louise; the advantages were many: I was free to teach as I pleased (the motto for Stanford of its first President, David Starr Jordan, "The winds of freedom are blowing," was fully realized in the classrooms of the Psychology Department); I was again a member of psychology's upper class; we were home among friends; Louise was welcomed back on a part-time teaching basis to her old school; my health was on the upgrade and promised to improve more in the California sunshine. We were in such good shape, in fact, that our third child was born in the spring of 1943. Furthermore, it turned out that I was able to make an important immediate application of my psychological knowledge.

But not all was rosy. The war was ever present on the West Coast, with blackouts, shortages, relatives and friends embarking for Asian combat, and casualties disembarking. The teaching load was heavy; in addition to the regular offerings, short cram courses were given for officers in training. I taught a number of subjects new to me, so I had to do much homework to keep ahead of the classes. Most of these 3½ years at Stanford were a steady grind. Only one setting stands out as of special significance for my career. It came about through a fortunate conjunction of (a) my desire to make a contribution to the war effort; (b) the national need for rehabilitation service for war casualties; (c) my interest in, my exploratory research upon, and my personal experience with, physical disability; and (d) resources that the Social Science Research Council (SSRC) made available. A crucial person in tying together these separate strands was my colleague Quinn McNemar, who was a member of the relevant SSRC committee; it was at his instigation that the Council funded the preparation of a survey of what was known about the psychological aspects of the rehabilitation of the physically disabled (Barker, Wright, Meyerson, & Gonick, 1953). I set up a special study at home for this project.

Office of Disability Survey

Program: Preparation of monograph *Adjustment to Physical Handicap and Illness. Physical attributes:* We were living in a house built by Walter Miles, my former professor and thesis adviser, which had a fine, spacious, isolated study; it became headquarters for the monograph project. *Temporal characteristics:* Occurred whenever I had time during the years 1943–1945. *Human components:* When the magnitude of the project became clear, I was fortunate in obtaining contributions from Beatrice A. Wright, Lee Meyerson, and Molly Gonick. *Boundary properties:* Ready access by the four participants. *Powers of human components:* Overall control was in my hands; others had complete autonomy for their own contributions.

In this setting I experienced some of the great satisfactions of my professional life, and they have continued to this day. In the first place, I felt competent; I believed I was as able any anyone to do the job at that time. Second, I saw this as my first opportunity to bring my skills to bear upon an urgent practical problem. As it turned out, the time was ripe for this; beginning with World War 2, concern and services for the disabled have increased greatly, so any firm foundations laid in 1943 have had continuing consequences. And third, I believed at the time that I was able to contribute to this firm foundation by pointing out that some psychological problems of the disabled are not unique to them, that

adolescents and ethnic minorities, for example, face the same problems; I was able to do this in terms of Lewin's concepts of marginal men and overlapping situations. I have not followed the course of rehabilitation psychology in recent years, so I have no firsthand knowledge of the permanence of this contribution, but I am told that the monograph is still in demand. Clearer evidence of the enduring consequences of the Disability Office is the fact that both Beatrice Wright and Lee Meyerson have become leaders of this field of psychological study.

In 1945 I wore the child psychology label, primarily, and there was not much demand for this branch of the science at Stanford. Members of the Department had ambitions to develop a program of studies in child psychology, but money was scarce and progress slow; their principal efforts went to accommodating regular staff members returning from war service and to filling vacant, established slots. So in October 1945 I was told that there was no hope for me beyond the next spring semester. As of June 1946 I would be through at Stanford. The likelihood of this subsidence of the stream of my professional life had long been within my time perspective, but the reality was more vivid than the distant prospect. It gave a new urgency to the somewhat relaxed inquiries I had initiated before reality descended. I had communicated about possible jobs with a number of schools, including Clark University, and in the summer I had an interview with President Atwood about the vacant G. Stanley Hall Professorship in Child Psychology. I did not take this possibility seriously. So, when a definite offer came from Clark in November, Louise and I were almost overwhelmed by the drastic alteration in our prospects. But we were able to make the transfer; we arrived in Worcester, Massachusetts, for the second semester in January 1946.

Two firsts of some importance occurred during the 3½ years at Stanford. One was advising graduate students on their thesis research. In line with my interest in the disability problem and my work on the survey, three students undertook theses on attitudes toward the disabled. The other first was the establishment of the Disability Survey Office as a major behavior setting with a program after my own specifications; until this time, apart from minor side excursions, I had operated in behavior settings with programs arranged by others. At Stanford I was born as a psychologist, and on my return 15 years later I gained these two evidences that I had at last reached my majority.

Louise and I left Stanford older and with more experience of the stream of university life. We had drifted through the dangerous, white waters of Harvard, portaged to the slow Illinois channel with its sandbars and driftwood, portaged again to the steady, full flow at Stanford, on whose banks we were briefly stranded until an unexpected flash flood carried us to Clark. One more portage was to come.

Clark University, 1946–1947

Clark turned out to be a bayou for us. The top administrators changed six months after we arrived, the psychology department had been understaffed for some time and had an acting chairman, and I was uncertain of the direction of my next efforts, having completed the disability monograph. With so much in the process of change, all of us—the University, the Department, and I—marked time. It was now that the seed planted in my mind on the plains of Illinois began to grow, and I seriously considered a project to discover and describe the living condition and behavior of the children of a small town. I took two actions. I explored the towns around Worcester, and I discussed the project with Kurt Lewin, who was then at the Massachusetts Institute of Technology and was living only 30 miles away. The exploration was discouraging. The small settlements of the region were not self-contained towns as in the Midwest; they were spacially dispersed and often specialized fragments of political units (towns) often straggling many miles along streams or across forested ridges. The children of these fragments did not have a common community environment as did the children of the towns I had seen in Illinois. But the discussions were encouraging, and in fact Lewin was enthusiastic. He, himself, was moving from lab-

oratory experimental research to "action research" in communities and institutions to study the consequences for the inhabitants of induced changes in their structures and programs. So he was supportive of efforts to establish baselines in unaltered communities against which to assess the effects of changes. With his encouragement, I worked on an application for funds for a community study with the intention of doing the research at a distance from Worcester if necessary.

And now luck was with us. The United States Public Health Service was expanding its support of basic research, particularly on children. The advisory committees included persons familiar with my work whom I had met at Stanford, Iowa, Illinois, and the Macy Foundation and Topology Meetings. The project was approved, and the research was funded early in 1947. And then Dean Lawson of the University of Kansas turned up in Worcester. He was looking for candidates for the chairmanship of their Psychology Department. The department was at a low ebb, and the administration was ready to make a new beginning and provide the support required to bring psychology at Kansas to its former rigorous state.

Why should I be interested? I had just come to Clark, to an endowed professorship with some status and to a department that also was set to make new beginnings. One could interpret my career as exhibiting clear signs of instability; a solid and frank New Englander, on hearing from Louise of the towns in which we had lived, remarked, "I see your husband is something of a floater." Surely it was time to settle in and stay on course. But there was the research and the dispersed Massachusetts towns. I told Dean Lawson of the project and the kind of town it required. Did he know of such a town in the Lawrence area? "Yes," he said without any hesitation, "I know the place. I've spoken there several times. Its name is Oskaloosa." I made two trips to Lawrence that spring; on the second trip I asked my former fellowship colleague at Iowa, Herbert Wright, to come along and consider joining in the department and the research. We visited Oskaloosa. He came and he joined. So Louise and I made one more portage.

University of Kansas, 1947–1972

We arrived in Lawrence in late August 1947 in a 115° heat wave. But, no matter, important things were underway. Fritz and Grace Heider were joining us, along with Herbert and Lorene Wright. There were houses to rent, a new department office to establish, instructors, teaching assistants to hire, and the University's rules and regulations to master.

I learned slowly why I had been chosen for this job over some other strong candidates. G. E. Coghill, a stellar name in biology in those days, had recently retired, and psychologists Raymond Wheeler, F. T. Perkins, and J. F. Brown had lately left the University. All of these men were well known for their "organismic" viewpoints. It was the wish of the two regular staff members who remained, Beulah Morrison and Anthony J. Smith, and of the administration, that this tradition be continued. Without making this explicit, Dean Lawson apparently saw that I would do this naturally and with enthusiasm. And in fact he had two immediate signs of the correctness of his insight when I recruited Herbert Wright and Fritz Heider within a couple of months. In the next three years we added Martin Scheerer, Alfred Baldwin, and Erik and Beatrice Wright. The new senior staff members were all from the center or the fringes of the Gestalt psychology movement.

Two behavior settings dominated my professional activities over the next 25 years; one of them was of my own creation, and I had considerable power over the other.

Office of Department Chairman

Program: To invigorate and expand the psychology department and to establish and administer settings subordinate to it according to University and Staff policies. *Physical attributes:* Suite of four rooms in Strong Hall with office furnishings and equipment. *Temporal characteristics:* My occupancy continued for 3 years; open for regular business weekdays 8:00 a.m. to 5:00 p.m., and for special business at any time. *Human components:* Myself, other staff members, students, secretaries. *Boundary properties:*

Open access. *Powers of human components:* The basic policies of the Department were determined by external settings: those of the University administration and the setting Department Staff Meeting. Within the constraints imposed by these settings, the chairman had complete control of the Office.

In my 19 years of academic life the extent of my administrative experience was limited to registering and advising students at Illinois and Stanford and to voting on minor issues in the College Staff Meeting at Illinois. My former colleagues at Stanford expressed anxious surprise that I would or could become an administrator. Their surprise was not justified. They did not know the strength of my commitment to certain academic and educational principles and my eagerness to have more of a hand in promoting them than hitherto; and they did not know that administration was part of the package that included Oskaloosa, the crucial component. There were grounds for their anxiety, though not the ones they probably had in mind. They would know of my deficiencies in keeping appointments, answering correspondence, arranging schedules, making and keeping to budgets, gregariousness, and so forth. But they would underestimate, I think, the power of the behavior setting Office of Department Chairman to take me in hand in these respects. However, their anxiety (and mine, too) would have been justified had they known of the probable conflict between the duties of the Office of Department Chairman and those of the research setting Field Station in Oskaloosa. I lasted 3 years as chairman.

The usual conflict between administration and research was exacerbated in this case by a geographical factor; we found it impossible to do the research in Oskaloosa, 20 miles from Lawrence, without living there and having full-time headquarters there; and the urgencies that occurred in both places could not be scheduled. However, if I had found administration congenial and rewarding, I might have remained longer as chairman. But I did not. For one thing, I did not have easy relations with my superiors. Some sort of a personal status and power problem was involved. I saw myself as being too submissive to policy directives I opposed, and thereafter I felt guilty of betraying my principles. I did not stand up to authority in the way I thought I should. This was partly a matter of my divided commitment; I did not have the time to prepare for hassles with top administrators; but there was also inability to confront superior power. Another problem for me as chairman was the discovery that persons with whom I agreed and had harmonious relations were not always in agreement and harmony among themselves. My ambition was for a unified, amicable department, and with incredible naiveté I supposed that those with whom I was congenial would be congenial inter se. There were no great conflicts within the department, but not the unity I had hoped for. An added circumstance that interfered with maintaining a department with a common point of view on psychological matters was the rapid increase in the staff to serve the great increase in university enrollment after 1947. Inevitably those added were of a variety of psychological persuasions. In my 3 years in the chairman's slot, the Department was invigorated with staff members who received wide recognition as scientific scholars in the next two decades.

These were tremendously busy years for me, with increasing conflict between the demands of the chairmanship and the research, with increasing dissatisfaction with my relations with University administrators, and with some disappointment that my goal of a small, excellent, unified group of scholars was not more completely achieved. So, my occupancy of the chairman's position in the Office of the Department of Psychology brought me both satisfactions and regrets.

Classroom teaching was not clear sailing either. I had decamped when there was administrative interference with the way I wished to function as instructor at the University of Illinois. Now, it was discouraging to me to discover that there were other, more pervasive institutional obstacles to my teaching. The Kansas University administration intruded very little in class operations, but a generally prevailing program of classroom instruction made it difficult for me to establish the kinds of programs I desired. This was my

irst experience with the fact that the settings of an institution may be so interdependent hat deviant settings cannot function. My lasses were arranged to foster students' skill n formulation of questions, finding and using he most relevant evidence, and writing re-orts on the basis of the evidence. Learning acts had no place in my classes; today's facts re obsolete tomorrow, and the current ones re always available if one knows where to ind them and how to use them. My classes ould operate best if the students had some nscheduled time: to think of interesting and nswerable questions, to search for relevant vidence, and to analyze, organize, and pre-ent the findings. But who can think, search, nalyze, and write when confronted with two uizzes a week in one class, 500 pages of eading in another, and 9 hours of laboratory ttendance in another? As so-called standards ent up and competitive grading became more evere, the students lost their freedom to act s inquirers, problem solvers, and expositors o the imposed demands of fact-, page-, and our-oriented classes.

In addition to trouble with the teaching ystem within the University, I found that lassroom settings and research settings do ot mix—that the activities required by class-oom teaching and by research are so conflict-ng as to be mutually injurious. I have ob-erved this to be true for others, too. Both erman and Lewin were excellent classroom eachers when they devoted themselves to it, ut when they were deep in research, as they sually were, their teaching declined pre-ipitately.

From the intrusions of the Illinois Dean I scaped to Stanford; from the interference of he prevailing teaching system and the con-icts with research I escaped to Oskaloosa. ortunately a Research Career Award from he National Institute of Health made this ossible.

ield Station in Oskaloosa

rogram: In the beginning, the program was to escribe the living conditions and behavior of all the hildren of the town; later, all inhabitants were in-luded. ***Physical attributes:*** Suite of offices in Oska-osa, Kansas, with office furnishings and calculators; at intervals a satellite Station was established in Leyburn, Yorkshire, England. ***Temporal characteristics:*** Operated during regular office hours, and often into the evening for particular staff members, from the fall of 1947 through the spring of 1972. ***Human components:*** Three to seven staff members, one to four graduate students, and occasional town residents. ***Boundary properties:*** Staff and graduate students selected and admitted by Station directors; they thereafter had free access; townspeople also had free access. ***Powers of human components:*** During the first 7 years, the Station was administered by the codirectors, Herbert F. Wright and me; we had power over the entire setting; when Herbert left the Station, I was in charge. Next in power were professional scientists (field workers, data analysts) who had complete control of their special operations; after them came the graduate students with power over their projects in consultation with their advisers, and finally the secretaries, who were masters of their own desks and secretarial facilities. Townspeople and other visitors had no power.

Here, at last, in the Midwest Psychological Field Station, the stream of my professional behavior entered a setting constructed to my own design: I initiated it, and its program was arranged cooperatively with Herbert Wright. Most other major settings I had inhabited had ongoing operations into which I was incorporated, with limited power to make alterations. I soon discovered, however, that in the Station I was not a free spirit but was captive of the setting I had established—a setting that embodied a past I could not escape and a future I could not control. Herbert and I had programmed the Midwest Field Station to describe the behavior and psychological habitat of the children of the town individually. With the setting underway (procedures developed and tested, staff trained, computer programs set up, citizens alerted and cooperative), it was not a simple matter when our insights and intentions changed to alter the program to one of describing extraindividual behavior within behavior settings. We had to struggle against our own creation. It makes one wonder how much of the stability of individual behavior has its source in the stability of the behavior settings people create and inhabit, a stability that is sustained by the fact that people establish new settings according to designs they carry with them from previous settings. The origins of some features we built into the

Midwest Field Station are fairly clear: The aim to describe the living conditions and behavior of all of the children originated in the Extension Classes in Illinois; concern with the naturally occurring environment was brought from Stone's Animal Laboratory and the Iowa Nursery School Laboratory; disciplined, persisting concentration on a rather narrow program of activities, the policy that "dogged does it," came aboard in the Stanford settings; the dual importance of precise data and precise theories welded together characteristics of both the Stanford and the Iowa settings; the particular theories that undergirded the program of the Midwest research came from Lewin's Offices in Iowa; the importance attached to archives of atheoretical data originated in Stone's Animal Laboratory; the first data collecting system we installed, the narrative record, dated back to the Iowa Nursery School Laboratory; the importance given to community relations was first met in Miles' Later Maturity Facility. These problems, emphases, methods, and theories with which we endowed the Field Station were imported from earlier settings. We installed the past in the present. But the inheritances were assembled in new relationships, and new elements were added, including staff members, methods, ideas, and data; the Station developed a dynamic of its own with unforeseen consequences for its inhabitants. For me these were far-reaching. In this latest stretch of the stream of my professional behavior, I entered the Field Station as a psychologist aiming to study the naturally occurring behavior and environments of the people of the town as individuals. For a long time I clung to the view that this was the way to observe and explain their everyday activities. But the Field Station finally turned me around and showed me that more than people and the stimuli that impinge upon them individually are required, that an ecobehavioral science of extraindividual behavior and its nonbehavioral context is needed. The course of this change has been presented in a number of publications (Barker, 1960, 1963a,b, 1965, 1968; Barker & Associates, 1978; Barker & Gump, 1964; Barker & Schoggen, 1973; Barker & Wright, 1951, 1955).

Other Places, Other People

I have had to omit many rewarding and pleasant reaches of my behavior stream and most of the people who shared parts of it with me and gave me tows and directions. Space is a factor here, but more important is absence of records of the early days. And there were important hidden settings (committees, administrative offices, and so forth), whose inhabitants I do not know, that gave me free time (Research Career Award: Fellowship, Center for Advanced Study in the Behavioral Sciences), research grants (U.S. Public Health Service, Society for the Aid of Crippled Children, Commonwealth Fund, Ford Foundation, Carnegie Foundation of New York, University of Kansas, Kansas University Endowment Association), honors (Kurt Lewin Award, Society for the Psychological Study of Social Issues; Research Contribution Award, American Psychological Association; G. Stanley Hall Award, Division on Developmental Psychology, American Psychological Association), and summer appointments (Columbia, Oregon, Colorado, California). I regret especially not being able to name the graduate students who studied with me at Stanford University, Clark University, and the University of Kansas, some of whom have become valued colleagues, and many of whom contributed greatly and still contribute to my intellectual development. My story would be too incomplete, however, without four prime influences on the odyssey. Louise Shedd Barker was an eager spectator, a frequent adviser, and an occasional stand-in in all the early settings. But when we discovered that the Station could not be operated from Lawrence, that it required us to live in Oskaloosa, Louise became an essential operative as chief field worker and our main line of communication with the community. She was well prepared. In her own professions she had experience as a field ecologist, first as biologist at the Stanford Marine Station and on the University campus, and then as high school "home visitor," searching out the homes of absent children and investigating the causes of their absence. Fritz Heider's ideas about media and things provided the link between

people and behavior settings that prompted me to see the latter as more than convenient areas for sampling behavior but rather as entities (things) that impose patterns upon their components, including their human inhabitants (media). It was this insight that raised behavior settings for me from a technological convenience to the basic unit of an ecobehavioral science. Herbert F. Wright was co-equal developer of the Field Station's program during its first years, making unique contributions to the study of individual behavior; and Paul V. Gump was equally important in the latter years by bringing the Station's ideas and methods to bear on school and community operation.

Conclusion

I began this trip with the desire to benefit humanity. I have not, myself, been able to satisfy this desire, for I have found that, as with teaching, the time and skills required to make and disseminate applications have so conflicted with those required by the research as to make the former impossible for me. So the zeal and success of a small number of former staff members, students, and others to bring our methods and findings to earth in connection with hospitals, schools, architecture, town planning, child development, and even the movement "small is beautiful" have given me keen satisfaction. I began the journey, too, with reservations about my ability to make it in the scientific world. When I report that the self-doubt has not been dispelled, the response may well be: How greedy can he be? What does he want in brownie points? The answer: The stream of my professional life has skirted the areas of psychology that are currently richly cultivated and harvested, and in fact, it has finally landed me outside the turf of the psychology tribe. Being something of a maverick has its rewards from those (and there are many) who value what they hope will prove to be worthwhile innovations. The innovator hopes so, too, but in the meantime he is at a disadvantage vis-à-vis his mainstream associates, since he cannot keep up with them in the main channel. So, how can he be sure where he stands?

References

Barker, R. G. Ecology and motivation. In M. R. Jones (Ed.), *Nebraska Symposium on Motivation* (Vol. 8). Lincoln: University of Nebraska Press, 1960. 1–49.

Barker, R. G. On the nature of the environment. *Journal of Social Issues,* 1963, *19,* 17–38. (a)

Barker, R. G. (Ed.). *The stream of behavior.* New York: Appleton-Century-Crofts, 1963. (b)

Barker, R. G. Explorations in ecological psychology. *American Psychologist,* 1965, *20,* 1–14.

Barker, R. G. *Ecological psychology: Concepts and methods for studying the environment of human behavior.* Stanford, Calif.: Stanford University Press, 1968.

Barker, R. G., & Associates. *Habitats, environments and human behavior.* San Francisco: Jossey-Bass, 1978.

Barker, R. G., & Gump, P. V. *Big school, small school.* Stanford, Calif.: Stanford University Press, 1964.

Barker, R. G., Kounin, J. S., & Wright, H. F. (Eds.). *Child behavior and development.* New York: McGraw-Hill, 1943.

Barker, R. G., & Schoggen, P. *Qualities of community life.* San Francisco: Jossey-Bass, 1973.

Barker, R. G., Wright, B., Meyerson, L., & Gonick, M. R. *Adjustment to physical handicap and illness: A survey of the social psychology of physique and disability* (Bulletin 55, rev. ed.). New York: Social Science Research Council, 1953.

Barker, R. G., & Wright, H. F. *One boy's day.* New York: Harper & Row, 1951.

Barker, R. G., & Wright, H. F. *Midwest and its children.* New York: Harper & Row, 1955.

Lewin, K. *Dynamic theory of personality.* New York: McGraw-Hill, 1935.

Lewin, K. *The conceptual representation and measurement of psychological forces.* Durham, N.C.: Duke University Press, 1938.

Shaw, G. B. *Sixteen self-sketches.* New York: Dodd Mead, 1949.

Received March 24, 1979 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 12, 2158–2160

Roger G. Barker and Behavior Settings: A Commentary

Phil Schoggen
Cornell University

These comments on the preceding paper by Roger G. Barker suggest that it documents and illustrates Barker's approach to science, his lifelong concern with environment–behavior relations, the major influences that have shaped his thought, and the utility of the behavior setting concept.

Barker's paper impresses me as a remarkably accurate, concise, and revealing reflection of his lifelong approach to science in general and to the study of the relations between molar behavior and its environment in particular. It sketches and illustrates in a fascinating way the development of his most significant conceptual contribution, the behavior setting, from its earliest precursors in Calvin Stone's animal laboratory to its present status in the broader domain that he now refers to as "an ecobehavioral science of extraindividual behavior and its nonbehavioral context." And he does it with the candor and wit that his associates will recognize as so characteristic of Barker. Of course, a good understanding of Barker's work requires study of his several books and papers, but I know of no single paper that is more truly indicative of Barker's thinking than this one. I have tried below to identify some of the main reasons why I have this impression.

This paper, above all, documents well a way of going about scientific work that both Barker and his principal mentor, Kurt Lewin, had in common, namely a spontaneous and omnipresent genius for finding in the mundane experiences of one's own everyday activities inspiration for the development of new concepts or practical tests of hypothetical relationships. For example, Lewin used his World War I battlefield experiences to develop his concept of the psychological environment (Lewin, 1917). His astute observation of a coffeehouse waiter's impressive memory for hours of charges and sudden forgetfulness once the bill was paid served as a basis for his conception of psychological tension systems (Marrow, 1969). Similarly, Barker's experiences on train trips through the countryside of central Illinois spawned his subsequent preoccupation with studies of "the nature of the situations that actually confront children in their daily lives and how they react to them" (Barker & Associates, 1978). Where others saw only pleasant countryside and small midwestern towns, Barker saw a major challenge to scientific psychology. His efforts to cope with illness and physical disability shaped his thinking about the relation between the psychological and the nonpsychological environments. Like Lewin also, Barker encouraged his students and colleagues to use their own direct experiences to further their understanding of fundamental psychological processes. As a graduate student research assistant and resident staff member on the Midwest project in its early days, I regularly attended church and sang tenor in the church choir. On one of our many trips by car between Lawrence and Oskaloosa, I remember expressing puzzled amazement at the faithfulness of the other parishioners—I couldn't understand why they so dependably braved the rigors of Kansas winters and suffered through Sunday after Sunday in the sweltering heat of summer to be present for the morning worship service. In his quiet, hesitant, almost apologetic way, Roger said,

Requests for reprints should be sent to Phil Schoggen, Department of Human Development and Family Studies, Cornell University, Ithaca, New York 14853.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3712-2158$00.75

"Well, I don't know. But, Phil, why do *you* go?"

This paper also documents another important feature of Barker's career in that it shows both his early concern with the problem of environment–behavior relations (he was writing about ecological problems many years before any other psychologist except Lewin and Brunswik) and the impressive singleness of purpose with which he has dedicated himself to this concern, forsaking all others. With his long-time friend and colleague Herbert F. Wright, Barker wrote in 1949 on what they then called *psychological ecology* (Barker & Wright, 1949). This was followed by their major report on the Midwest project (1955), Barker's seminal theoretical paper in the Nebraska Symposium on Motivation (Barker, 1960), his now classical *Ecological Psychology* (Barker, 1968), several other books and papers alone and with colleagues, and most recently, a new work that presents his current thinking about ecobehavioral science, which goes well beyond ecological psychology (Barker & Associates, 1978).

Dedication to his early learning at Stanford that "dogged does it" is apparent also in the fact that, with the single exception of the 1953 revision of his Social Science Research Council monograph on the social psychology of physique and disability (Barker, B. Wright, Meyerson, & Gonick, 1953), every one of Barker's seven books and 22 papers published since 1949 is devoted to the ecology of behavior.

Barker (1963) has credited Kurt Lewin and Egon Brunswik as the intellectual forebears of his work on the ecology of behavior. In the present paper, he also traces his concerns for understanding environmental influences on behavior to some rather surprising beginnings. Barker's thinking reflects still other important influences.

The work of social anthropologists such as Margaret Mead, Clyde Kluckhohn, and Ruth Benedict figured prominently in the thinking of Barker and Herbert Wright in the early days of the Midwest Field Station research. I recall many discussions of the project staff during which great admiration was expressed for the pioneering efforts of the social-anthropologists to study behavior in natural situations. Although Barker and H. Wright wanted to study child behavior and development in ordinary American as opposed to primitive cultures, and although they tried to develop observational and recording methods more systematic than the traditional field notes of the social anthropologists, it is clear that Barker and H. Wright built heavily upon the foundation laid by these workers.

Sociology and particularly human ecology also provided important influences. I remember well my delight in finding that my oral examination for the master's degree surprisingly turned into an interesting debate among Roger Barker, Herbert Wright, and Martin Scheerer as to whether the classical study of mental disorders in different areas of Chicago by sociologists Faris and Dunham (1939) could be considered a study in psychological ecology. This discussion showed clearly that the thinking of Barker and H. Wright was strongly influenced by the work of sociologists. This was evident also from their familiarity with and appreciation of other related sociological literature, for example, Dollard's *Caste and Class in a Southern Town,* the Lynds' *Middletown in Transition,* and Hollingshead's *Elmtown's Youth.*

A less conspicuous but still important influence on Barker's thinking came from the field of biology, which provided him with extremely useful models. Although it was clear to me that Barker had done a considerable amount of reading in biology on his own, I think Louise Shedd Barker's formal training in biology and her central involvement in Roger Barker's thinking throughout his career greatly enhanced the value of concepts, principles, and research strategies from biology in the development of his own ecological approach.

Certainly Barker's concept of ecology came directly from its use in plant and animal ecology. Webster's unabridged dictionary definition suits equally well Barker's use of the term and that of biological ecology, "the totality or pattern of relations between organisms and their environment."

In his early attempts to get some project

staff members to understand the mutual interdependence of behavior and its immediate context, Barker spoke of the well-worn footpath that cut diagonally across a familiar vacant lot near the high school in Oskaloosa. At some earlier time, the path did not exist, and the first persons to cross the lot diagonally shaped that part of the environment in line with their need to reduce the amount of required walking. Yet the path that they created was not completely straight but curved and bent around clumps of bushes, mounds, and depressions. Once established, however, that path shaped the behavior of countless pedestrians, nearly all of whom now follow its curves and bends as faithfully as a train on steel rails.

Eugene P. Odum in *Fundamentals of Ecology* (1971) expresses the same thought in the context of his discussion of the ecology of fresh water ponds: "Not only is the pond a place where plants and animals live, but plants and animals make the pond what it is" (p. 12).

Finally, it is appropriate that Barker chose to write in terms of behavior settings central in his professional career, because the concept of behavior settings will, I believe, turn out to be his most important and enduring contribution. Its potential significance for behavioral science is only beginning to be recognized (e.g., Moos, 1976; Wicker, 1979). To the best of my knowledge, it is the only concept in the behavioral sciences that combines both the ecological environment and behavior in a single unit of empirical analysis. The behavior setting concept derives from and reflects Barker's concept of the environment (Barker, 1963). This concept marks Barker's originality and most clearly distinguishes his position from those of Brunswik and Lewin. Whereas for them and for psychologists generally (e.g., Leeper, 1963), the ecological environment is unstable, characterized by only statistical regularities, and related to behavior only in terms of probabilities, Barker argues and provides impressive empirical evidence that there is order and organization in the preperceptual environment and that its relation to behavior is neither random nor merely probabilistic. Although the laws that govern the phenomena of the nonpsychological environment are different from and incommensurate with the laws that govern individual behavior, Barker and his associates have demonstrated in their work with behavior settings that despite this conceptual incommensurability, our understanding of molar behavior can be improved through systematic studies of its relations with the immediate environmental contexts. Behavior settings and the molar behavior of their inhabitants are mutually, causally related. Barker's review of important behavior settings in his own professional career provides further testimony to this important discovery.

References

Barker, R. G. Ecology and motivation. In M. R. Jones (Ed.), *Nebraska Symposium on Motivation* (Vol. 8). Lincoln: University of Nebraska Press, 1960.

Barker, R. G. On the nature of the environment. *Journal of Social Issues*, 1963, *19*, 17–38.

Barker, R. G. *Ecological psychology: Concepts and methods for studying the environment of human behavior*. Stanford, Calif.: Stanford University Press, 1968.

Barker, R. G., & Associates. *Habitats, environments and human behavior*. San Francisco: Jossey-Bass, 1978.

Barker, R. G., and Wright, H. F. Psychological ecology and the problem of psycho-social development. *Child Development*, 1949, *20*, 131–143.

Barker, R. G., and Wright, H. F. *Midwest and its children*. New York: Harper & Row, 1955.

Barker, R. G., Wright, B. A., Meyerson, L., and Gonick, M. R. *Adjustment to physical handicap and illness: A survey of the social psychology of physique and disability* (Bull. 55, rev. ed.). New York: Social Science Research Council, 1953.

Faris, R. E. L., and Dunham, H. W. *Mental disorders in urban areas*. Chicago: University of Chicago Press, 1939.

Leeper, R. W. Learning and the fields of perception, motivation and personality. In S. Koch (Ed.), *Psychology: A study of a science* (Vol. 5). New York: McGraw-Hill, 1963.

Lewin, K. Kreigslandschaft. *Zeitschrift für Angewandte Psychologie*, 1917, *12*, 440–447.

Marrow, A. J. *The practical theorist: The life and work of Kurt Lewin*. New York: Basic Books, 1969.

Moos, R .H. *The human context*. New York: Wiley, 1976.

Odum, E. P. *Fundamentals of ecology* (3rd ed.). Philadelphia: Saunders, 1971.

Wicker, A. W. *An introduction to ecological psychology*. Monterey, Calif.: Brooks/Cole, 1979.

Received April 15, 1979 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 12, 2161-2170

The Identification of Persons as Supersets and Subsets in Free-Response Personality Descriptions

Michael A. Gara and Seymour Rosenberg
Rutgers—The State University

Using a free-response format, 14 subjects each described 36 target persons. A set-theoretical model was proposed for representing the perceived similarities and differences among the persons described. Using this model, four types of targets were identified for each subject's protocol: supersets, subsets, disliked contrasts, and miscellaneous. Supersets are persons whose perceived characteristics subsume those of many of the other persons described. Subsets are persons described with only a limited portion of the characteristics attributed to the supersets. Disliked contrasts are those disliked persons who were described with a set of terms almost completely different from those used to describe the other targets. Miscellaneous targets are target persons who could not be identified as supersets, subsets, or disliked contrasts. As predicted, the supersets were persons who were perceived by the subjects to be very significant in their lives. Disliked contrasts were persons whom the subjects perceived as least significant; subsets and miscellaneous targets could not be distinguished from one another in this regard. It is suggested that supersets provide the perceptual categories for the construing of persons, and the correspondence of this idea with other extant psychological ideas, such as transference, is discussed.

We are often in the process of comparing people with each other. Our conversations are punctuated with statements such as "Those two sisters could hardly be more different" and "John and Bill are as alike as two peas in a pod." It does not seem unusual, then, that the perceived similarities and differences among the people known by an individual have occupied the serious attention of social psychologists in the area of person perception. These perceived relationships have been examined in a number of studies (Cronbach, 1958; Jackson, Messick, & Solley, 1957; Jones & Young, 1972; Messick, 1961; Rosenberg, 1977).

As part of the analyses of these perceived relationships, a measure of the psychological proximity of each person in a subject's life to each other person was defined in each of the above studies. The measures chosen all satisfied the symmetry assumption. That is, the similarity of Mary to John is defined as equal to the similarity of John to Mary.

A considerable number of theoretical and empirical arguments make the assumption of symmetry questionable. Within the field of person perception, the possibility of asymmetric relationships among personality descriptors has been raised several times at the theoretical level (Hays, 1958; Jackson, 1962; Rosenberg & Sedlak, 1972). At the empirical level, there is a study by Warr and Knapper (1968) in which subjects made inferences on an adjective checklist about a target person who had been initially described by the investigators with six different cue traits. The subject was asked two questions for any given pair of traits: If this person is cynical, how likely is he to be precise. If this person is precise, how likely is he to be cynical? The

This work was supported in part by National Science Foundation Grant BNS 76-10675 to Seymour Rosenberg and grants for computer use from Rutgers Center for Computer and Information Services.

The authors acknowledge the helpful comments given by Sam Kingsley throughout this research.

Requests for reprints should be sent to Michael Gara, Psychology Department, Livingston College, Rutgers—The State University, New Brunswick, New Jersey 08903.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3712-2161$00.75

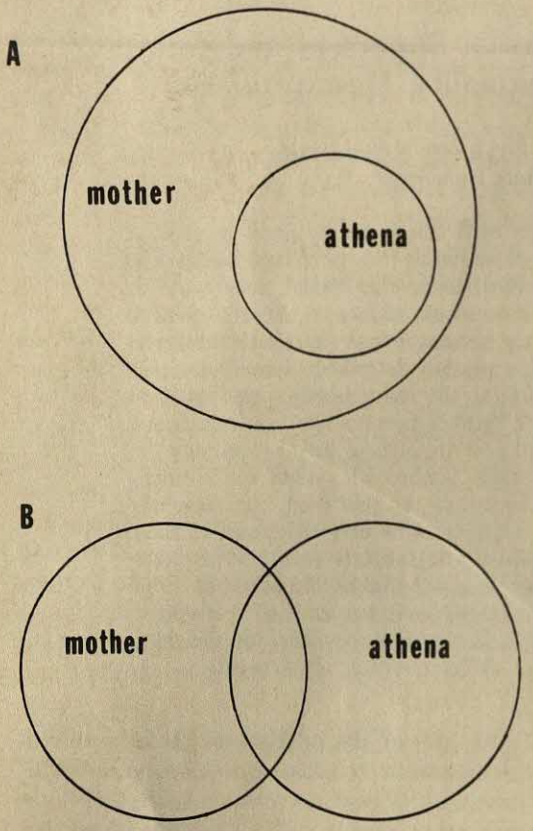


Figure 1. A: Hypothetical superset relationship. B: Hypothetical symmetric relationship.

results indicate marked asymmetries: Being *cynical* implied being *precise,* but being *precise* did not imply being *cynical.*

The notion of asymmetry has also received attention outside the immediate sphere of person perception. Working in the tradition of personal construct theory, which shares methodological and substantive concerns with person perception research, Hinkle (1965) developed a technique (IMPGRID) for assessing the one-way implications between constructs. Furthermore, investigators have found asymmetric similarity judgments with objects such as countries and geometric figures, (Tversky & Gati, 1977), auditory signals (Wish, 1967), and colors, line orientations, and numbers (Rosch, 1975).

In the present study, a set-theoretical model of person perception is used to depict the target persons described by subjects. This model permits asymmetries in the description of the targets to be reflected in the set-theoretical relation of targets to each other. The most extreme type of asymmetry is depicted in Panel A of Figure 1: Athena is a proper subset of Mother. The fact that Athena has an attribute implies that Mother has it as well; however, the converse is not true. Hays (1958) proposed a set-theoretical model for the relationships among personality descriptors, but it received little subsequent attention.

In this study, the set-theoretical relation between targets is measured by a special implication phi (here denoted ϕ_I) that was proposed by Francis (1961). Unlike the standard phi coefficient for the correlation between two dichotomous measures, there are two ϕ_Is for any two target persons, A and B, one for A → B and one for B → A. It is the inequality of these two phis that reflects the degree and direction of asymmetry. When there is a symmetric relationship between targets such as that depicted in Panel B of Figure 1, the two phis will be equal.

The application of a set-theoretical model to the target persons that an individual describes has a special appeal. Some targets (dubbed *supersets*) may bear the superset relation that Mother bears with Athena in Panel A of Figure 1 (although it may not be perfect) to many other target persons described. A group of other targets (dubbed *subsets*) bear relationships akin to that which Athena bears to Mother in the same figure.

It seems reasonable that supersets are persons in an individual's life who provide the perceptual categories for the construing of other persons. The underlying notion is that supersets are persons with whom the individual has interacted in many different types of situations. A good deal is thus known about these persons in terms of personality characteristics, and a portion of these may consequently be assigned to persons known less well. In fact, the personality characteristics referred to by the personality terms that an individual uses to describe people were probably first observed in these supersets.

Consequently, we should find that these same supersets are known more intimately by

the subject than are subsets and any other type of target we identify. In addition, supersets are probably perceived as more complex than other targets, and the affect that is displayed in their presence is more ambivalent (multivalent) than that displayed in the presence of other targets. Further interaction with nonsupersets may result in a significant impression change, whereas the overall impression with regard to supersets should be already well formed and slow to change. In sum, supersets are the most significant people in an individual's life.

To test these ideas we obtained a data base independent of the one we used to identify supersets and subsets. This consisted of ratings of the targets on several properties that we thought would differentiate the targets in terms of how *significant they were in a subject's life*. Then the question was asked: Are supersets the significant persons in the subject's life?

Method

Subjects

Ten female and four male undergraduates enrolled in a psychology course at Livingston College served as subjects. Each subject was paid $36 for participating in the study, which involved approximately 10–15 hours of his or her time. Subjects were told in advance that they would be paid only upon the completion of their work, that incomplete work would receive no payment.

Procedure

The first data that were obtained consisted of each subject's descriptions of a sample of people that he or she knew. This data base was used to identify the supersets as well as the other types of target persons for each subject.

In obtaining these data for each subject, the free-response methods developed by Rosenberg (1977) were followed exactly, except for the way target people were selected. For that purpose the subjects were provided with a list of 36 role descriptions (Table 1). A subject was instructed to read each role description and then list a person in his/her life who seemed to fit that particular description best. Thus, subjects generated their own lists of 36 target persons to describe.

The free-response method for obtaining descriptions of these targets required that the subject generate his/her own descriptive vocabulary. Each subject generated two different kinds of vocabularies: *traits* that the subject perceived as descriptive of the various target people and *feelings* that were elicited in the subject by these target people. The subject was asked to formulate his/her trait perceptions and feelings in the form of discrete units, but not necessarily single words. The subject cumulated a trait and feeling list as he/she described the target persons —that is, when he/she used a trait or elicited feeling to describe one person, he/she was asked to judge the presence of that trait or elicited feeling for all the other target people.

Before systematically describing any of the people on his/her list, each subject prepared two "starter lists," one of traits and one of feelings. To obtain these lists, the subject was instructed to write down on a provided questionnaire the first five traits char-

Table 1
List of Persons Described by Subjects

1. Mother
2. Father
3. Liked brother
4. Least liked brother
5. Liked sister
6. Least liked sister
7. Most liked relative—intimate with
8. Second most liked relative—intimate with
9. Most liked relative whom you are not intimate with
10. Second most liked relative whom you are not intimate with
11. Most disliked relative
12. Second most disliked relative
13. Best friend
14. Next best friend
15. Third best friend
16. Casual friend
17. Casual friend
18. Casual friend
19. Most disliked person you know well
20. Second most disliked person you know well
21. Third most disliked person you know well
22. Most liked person you have met in last six months
23. Second most liked person you have met in last six months
24. Most disliked person met in last six months
25. Second most disliked person you have met in last six months
26. Public figure—most liked
27. Public figure—second most liked
28. Public figure—least liked
29. Public figure—second least liked
30. Me now
31. Me ideal
32. Me negative ideal
33. Spouse or lover—present
34. Spouse or lover—past
35. Mythical figure—liked
36. Mythical figure—disliked

Table 2
General Relationship Between Attributions to Two Target Persons

	Target Person B	
Target Person A	1 or 2	0
1 or 2	*a*	*b*
0	*c*	*d*

Note. 0 = target does not possess attribute; 1 or 2 = target does possess attribute. *a*, *b*, *c*, *d* = number of times a given combination occurs in description of target person.

acteristic of, and first five feelings elicited by, Targets 1, 7, 12, 14, 18, 21, 23, 26, 30, 32, 34, and 35. The starter lists served not only to reduce biases in content that might be associated with the first few people on the list but also to help the subject get started with the description task. Once collected, the two starter lists were each randomized for each subject, and the description task began. The descriptions were collected on a computer terminal tied to the Rutgers time-sharing system. Each subject entered his/her list of 36 target persons and the two starter lists, one of traits and one of feelings. A computer program (Livingston & Kingsley, 1976) was used to (a) present the subject with his or her list of target persons one at a time; (b) present for each target person all descriptors (traits or feelings) the subject had entered; (c) record the subject's judgment (0, 1, 2) of each target person on each descriptor; (d) record any new trait (or feeling) term that the subject wished to add to his or her vocabulary of terms; (e) re-present for judgment the persons already described, when a new trait (or feeling term) was added.

Subjects were instructed to make the trait and feeling judgment on a 3-point scale. For traits, these were:

0 = the trait is not descriptive of the person, but does not necessarily mean that the person possesses the opposite trait (in order to indicate that the person possesses the opposite trait, the subject was instructed to add the opposite as a separate trait term).

1 = the person possesses the trait to a noticeable degree but need not describe the way the person is all the time.

2 = the person posseses the trait to an extreme degree but need not describe the way the person is all the time.

A similar scale, appropriately worded, was used for the feelings.

0 = the person does not make (or has not made) you feel this way.

1 = the person makes (or has made) you feel this way to a noticeable degree.

2 = the person makes (or has made) you feel this way to an extreme degree but need not make you feel this way all the time.

During this entire process of describing the 36 target persons, the trait and feeling vocabularies were kept separate: A subject systematically alternated between trait and feeling descriptions according to a fixed schedule. The subject started with trait descriptions for the first 5 people on his/her list, followed by feeling descriptions for the first 10 people on the list. The subject switched back to trait descriptions and alternated them with feeling descriptions, such that after persons n through $n + 9$ were described by traits, persons $n + 5$ through $n + 14$ were described by feelings.

When the description tasks were all completed, the trait and feeling vocabularies were combined for each subject. Thus the computer contained 0, 1, and 2 entries in a $36 \times n$ descriptor matrix for each subject.

Identification of Supersets and Subsets in a Subjects Protocol

Suppose we represent all possibilities of perceived trait and feeling occurrence in Person A with respect to Person B, using a 2×2 table (Table 2), where 0 represents the rating that the target person (A or B) does not possess an attribute, and 1 or 2 represents the rating that the person does. The cell entries *a*, *b*, *c*, and *d* each represent the number of times that a given combination occurs in the actual description of these two persons.

The relationship expressed with the Venn diagram in Panel A of Figure 1 depicts Mother as a proper subset of the goddess Athena. Let Mother be Person A and Athena be Person B in Table 2. As seen in Panel A of Figure 1, whatever attributes Athena possesses, Mother does also. In other words, *c* must equal zero. For the general case, whenever we can marshall evidence that the *c* cell is indeed equal to zero, we can state that Person B is a proper subset of Person A. Situations where *c* is small but not equal to zero can be placed on a continuum of what Francis (1961) calls "weak implication" and he demonstrates that a ϕ_I coefficient (range from -1 to 1) can be calculated, where

$$\phi_I = \frac{ad - bc}{(a + c)(c + d)}. \quad (1)$$

A high positive value indicates the degree to which Person B's characteristics are a subset of Person A's. A high negative value indicates that Person B implies the opposite of Person A.

For the weak implication case, ϕ_I's were calculated for each subject for all pairs of targets described, yielding a 36×36 matrix of ϕ_I's for each subject. Each entry indicates the extent to which the characteristics of the target in column j are a subset of the characteristics of the target in row i. By definition, the diagonal entries are 1.

To simplify examination of this matrix, the rows and columns were permuted separately to bring into proximity rows or columns with a similar pattern of

ns. This permutation was done according to a hierarchical clustering of the rows of the matrix and then separately of the columns, using as input a profile distance

$$\delta_{ij} = \sum_{k=1}^{36} (\phi_{1,k} - \phi_{1,k})^2 \qquad (2)$$

alculated separately for rows (where k refers to olumns) and columns (where k now refers to rows). his simple type of two-way clustering is described some detail in Rosenberg (1977).

Figure 2 illustrates a sample protocol from one ubject following two-way clustering. All target perons numbers in Figure 2 correspond to the numbers f the role figures in Table 1. In this example, the ow labeled 1 (which is Mother) illustrates the egree ($\phi_1 \times 10$) to which the characteristics of each rget designated in the columns—6, 11, 24, and so rth—are subsets of the characteristics attributed Mother. A minus sign indicates that the ϕ_1 is egative, whereas an asterisk indicates that $\phi_1 = 1.0$.

Distinct clusters of targets in Figure 2 are sepated by horizontal lines for the row-wise clustering nd vertical lines for the column-wise clustering. he following guidelines were used to group the rgets (row-wise and column-wise) for each subct. We chose the level of the hierarchical tree proced by the clustering program that separated the 6 targets into 7 clusters. If there was a comparaively large cluster (in terms of number of targets ntained) remaining at this level of the tree, a gher level of the tree was chosen to break up this uster. However, in no case were more than 11 clusrs chosen.

After the clustering of the targets had been fixed this way, targets were classified as supersets, subts, disliked contrasts, or miscellaneous for each ubject, as shown in Figure 2. (Miscellaneous targets re not labeled in Figure 2.)

Note that target persons in the same cluster share very similar pattern of ϕ_1s. To indicate the genral size of the ϕ_1s within a cluster, the sum of all quared ϕ_1s within a cluster is averaged by the numer of targets within that cluster, and this number shown for each row and column cluster in Figure . To be classified as a superset, it is necessary, but ot sufficient, that a target be included in the luster(s) of targets with a pattern of high ϕ_1s *row-wise*. Similarly, to be classified as a subset, it is ecessary, but not sufficient, that a target be inluded in the cluster(s) of targets with a pattern of igh ϕ_1s *column-wise*.

The reason that the above conditions are necessary ut not sufficient for classifying targets as supersets r subsets is that a given target could occur in both lusters described above, if that target was charcterized by a pattern of high ϕ_1s, both *row-wise* nd *column-wise*. That target would bear an overall *ymmetric* relation, not a superset or subset relation. uch targets rarely occur. In a total of 504 targets 14 subjects × 36 targets per subject) we classified 1 targets as supersets, 71 as subsets, and only 23 as symmetrical. Moreover, the latter were found in only 7 of 14 subjects.

In every case, a cluster(s) of target persons emerges with negative phis in relation to supersets and subsets. This category was dubbed "disliked contrasts" because without exception, only the disliked role figures in Table 1 emerge in this category.

Targets not identified as supersets, subsets, or disliked contrasts are grouped into a miscellaneous category. Included in this category are the symmetric targets, as defined above.

Property Ratings

The other data that were obtained from each subject after he/she had completed the person description task were ratings of each of his/her target people on the properties listed in Table 3. These properties were constructed to assess the significance of each of the target persons in the subject's life. The property ratings were then used to test the idea that supersets, which had been identified in the free-response protocols, were significant people.

Results

In order to relate the interpersonal significance of a target person as assessed by the property ratings with the type of target as identified by the ϕ_1s, a multivariate analysis of variance (MANOVA) was performed on the property ratings. The target type effect (which has four levels: supersets, subsets, disliked contrasts, and miscellaneous) was tested against the variance–covariance matrix for the Target Type × Subject interaction,[1]

[1] Using a test proposed by Huynh and Feldt (1970), we found that certain assumptions were not met that were requisite to the validity of this particular test of the main effect when repeated measures are involved. To ensure the validity of the reported findings, alternative procedures of analysis were pursued. These involved the calculation of three orthogonal contrasts among each subject's four target type means for each property and testing these contrasts for flatness using a Hotelling T^2 and Bonferroni adjusted significance levels. There were significant differences among the four target type means for every property except Complexity and Feeling Ambivalence. Specific contrasts among the target type means for each property were also tested (essentially a single sample t test of the hypothesis that the population value of the contrast is zero) using Bonferroni adjusted critical values (Harris, Note 1). The profile of results using MANOVA is essentially replicated by these procedures, thus only the MANOVA is reported. A summary of these alternative analyses is available from the author upon request. We thank John Miller for his assistance with these analyses.

Persons	6 1 2 2 3 1 4 8 6	3 1 1 2 2 2 9 0	1 3 8 1 0 3	2 1 2 2 2 8 3 6 7	3 1 1 2 4 6 7 1	1 9 4	3 2 2 4 5 7 5 2	3 3 2 1 3 1 3 1 9 0 5 5	ave. $\Sigma(\phi_{\underline{I}}^2)$ for cluster
11	5 * 5 6 4	4 3 5 3	- - - -	- - - - -	- - - -	- -	- - - - - - -	- - - - - -	
24	4 7 * 4 5	4 4 5 4	- - - -	- - - - -	- - - -	- -	- - - - - - -	- - - - - -	
6	* 6 4 5 4	4 4 6 3	- - - -	- 1 - - -	- - - -	- -	- - - - - - -	- - - - - -	4
28	3 4 2 * 4	5 2 3 2	- - - -	- - - - -	- - - -	- -	- - - - - - -	- - - - - -	
36	4 5 5 6 *	2 3 5 3	- - - -	- - - - -	- - - -	- -	- - - - - - -	- - - - - -	
32	2 2 2 4 1	* 2 3 3	- - - -	- - - - -	- -	- -	- - - - - - -	- - - - - -	
19	5 4 4 4 3	4 4 * 4	- - - -	- - - -	- - -	- -	- - - - - -	- - - - - -	3
12	4 4 4 3 3	4 * 5 4	-	- - - - -	- - - -	- -	1 - - - - -	- - - - -	
20	4 5 5 3 3	6 4 6 *	- - - -	- - - -	- - -	- -	- - - - - -	- - - - - -	
3	- - - - -	- 1	3 1 2 2	5 6 4 5 6	3 3 4 4	6 6	* 6 5 4 4 4	3 4 4 4 3 4	
25	- - - - -	- - -	3 1 2 2	6 5 5 6 6	3 3 4 5	6 7	5 * 6 5 5 5	3 3 4 4 4 4	
22	- - - - -	- - -	3 1 2 2	5 5 4 5 6	4 3 4 4	4 6	4 6 * 5 5 5	3 4 3 4 3 3	
18	- - - - -	- - -	1 1 1 1	5 * 4 4 3	3 3 5 3	5 5	3 3 3 3 2 3	1 2 4 3 2 3	
16	- - - - -	1 - -	2 1 1 1	4 6 5 5 5	4 * 6 5	5 5	3 4 3 1 2 2	1 2 3 3 3 2	
17	- - - - -	- - - -	1 1 1 1	5 8 6 6 5	5 6 * 5	5 5	3 4 3 2 3 2	1 2 4 3 3 3	7
23	- - - - -	- - - -	2 2 1 1	5 7 * 5 6	5 4 5 5	6 6	3 4 4 3 3 3	2 3 4 4 4 4	
27	- - - - -	- - - -	2 1 1 1	5 5 5 5 *	3 4 4 4	6 5	3 5 4 3 3 3	2 2 3 4 3 3	
21	- - - - -	- - -	2 2 1 2	5 6 6 6 7	5 6 7 *	4 6	4 6 5 2 4 4	2 3 5 3 4 4	
2	- - - -	- - -	1 1	* 4 3 2 3	3 2 2 2	2 3	2 3 3 2 2 2	2 2 2 2 2	3
34	- - -	1 - - -	1 1 1 1	3 3 4 3 2	* 3 3 3	1 2	1 2 2 1 2 2	1 2 2 1 2 3	
4	- - - - -	- - - -	2 1 1	3 4 3 3 3	1 1 2 1	5 4	2 4 3 * 5 5	2 2 1 3 3 3	
5	- - - - -	- - - -	2 2 1 1	5 4 4 4 3	3 2 3 4	5 7	3 5 5 6 * 6	3 4 4 4 4 4	5
7	- - - - -	- - - -	2 2 2 1	4 5 4 4 4	3 2 3 4	5 6	4 5 5 7 7 *	2 4 3 3 4 4	
1	- - - - -	- 1 - -	* 3 2 4	6 5 6 6 9	5 5 4 6	7 8	6 8 8 6 6 6	5 6 5 6 5 5	
30	- - - - -	- - - -	4 * 3 4	6 7 8 6 6	7 5 6 7	7 *	4 6 6 4 7 6	5 7 7 6 5 5	12
13	- - - - -	- - - -	5 4 4 *	5 7 6 7 7	8 5 5 7	5 6	5 7 9 5 6 5	6 8 7 6 6 6	
8	- - - - -	- 1 - -	3 3 * 5	6 6 6 5 5	7 3 4 6	5 7	5 8 8 4 6 6	6 8 8 5 6 6	SUPERSETS
31	- - - - -	- 1 - -	3 3 3 3	6 7 5 5 5	6 3 4 4	6 8	5 6 7 6 7 7	5 * 7 5 5 5	
33	- - - - -	- - - -	2 2 2 2	2 4 4 4 4	3 2 2 3	5 7	4 5 4 3 4 3	* 5 6 4 4 4	
29	- - - - -	- - -	2 2 2 2	4 7 5 5 5	3 4 5 5	5 6	4 5 4 2 4 3	4 5 * 4 4 4	5
14	- - - - -	- - - -	1 1 1	2 4 3 4 4	1 2 2 1	* 5	2 3 2 3 3 2	1 2 2 4 3 3	
9	- - - - -	- - - -	1 1 1 1	3 5 4 4 4	2 3 3 3	5 *	3 4 3 3 4 3	3 3 3 3 3 3	
26	- - - - -	- - - -	2 2 1 2	5 8 6 * 7	4 5 7 5	8 8	5 6 5 4 4 4	3 3 5 6 5 4	
10	- - - - -	- - - -	3 2 1 2	5 7 7 7 7	3 3 5 4	* 8	4 6 5 5 5 4	4 4 5 * 5 4	10
35	- - - - -	- - - -	2 2 2 2	6 6 6 6 6	4 4 5 5	7 7	3 6 4 5 5 4	3 4 4 5 * 5	
15	- - - - -	- - - -	2 3 2 2	6 6 5 6 6	4 3 5 5	7 6	4 5 6 5 5 3	3 4 4 5 6 *	
ave. $\Sigma(\phi_{\underline{I}}^2)$ for cluster	5	4	3	9	7	12	7	7	

DISLIKED CONTRASTS

SUBSETS

Figure 2. Sample protocol with supersets and subsets identified.

and a significant main effect was obtained, F approximation to Wilks's lambda $(24, 93) = 5.78$, $p < .0001$.

In a multivariate analysis of variance, whenever the number of groups (K) is less than or equal to the number of variables being investigated, there are at most $K - 1$ (in this case, three) potentially interesting dis-

Table 3
List of Properties Used to Rate Targets

Label	Question	7-point bipolar scale
Self-Disclosure	1. How much do you discuss personal problems?	not at all–a lot
Concern	2. How concerned are you about whether this person thinks well of you?	not at all concerned–very concerned
Self-Expression	3. How far do you express feelings of love, approval, or admiration for this person?	try to conceal–express freely
Range of Interaction	4. In how many different types of situations have you interacted with this person?	limited–many
Closeness	5. How close are you to this person presently?	little–very
Complexity	6. How could you characterize the complexity of this person?	simple–complex
Feeling Ambivalence	7. How conflicted or ambivalent are your feelings about this person?	little–very
Impression Stability	8. If you were to interact with this person in many more types of situations than you do now, what do you expect would happen to your impression of this person?	it would change a great deal from what it is now–it would not change much from what it is now

riminant functions that distinguish among he groups. Harris (1976) argued that Wilks's ambda, although providing a valid test for n overall effect, cannot be used as a test or the significance of any particular discrim-nant function, including the first. He recom-nended comparing the characteristic roots ssociated with each discriminant function to he critical value of the greatest-character-stic-root distribution. This procedure was ollowed, using the *s*-parameter specifications uggested by Harris for testing each root and he same *m* and *n* parameters as for the over-ll test. (See Morrison, 1976, p. 189.) The est statistic .420 was extracted for the first iscriminant function and was statistically ignificant ($p < .01$). The second and third iscriminant functions were not found to be ignificant.

The first discriminant function associated with the target type effect accounts for almost 85% of the variation in the property ratings and is the direction in these data that most distinguishes the four target types. To clarify this discrimination, a Duncan multiple com-parison procedure was performed on the multivariate means, which are obtained, one for each target type, by multiplying the prop-erty means collapsed across subjects by the first discriminant function. This procedure revealed that the supersets scored highest ($p < .001$) on the linear combination of the properties yielded by the first discriminant function. There were no significant differences between subsets and miscellaneous targets, and the disliked contrasts scored lowest.

The actual discriminant function coefficients and the loadings of each of the eight properties on this function are presented in Table 4. The loadings are the correlations of each property with the discriminant function and tell us how well we could estimate the discriminant function with a property if only we had that property at hand.

With respect to the discriminant function

Table 4
Coefficients and Property Loadings Associated With Target Type Effect for the First Discriminant Function

Property	Coefficient	Loading
Self-Disclosure	−.000002	.59
Concern	.020061	.94
Self-Expression	.008741	.75
Range of Interaction	−.000437	.45
Closeness	.006339	.72
Complexity	−.002833	.01
Feeling Ambivalence	.003891	−.09
Impression Stability	−.003147	.23

coefficients, the target types are largely distinguished by their ratings on the question: How concerned are you about whether this person thinks well of you? The subjects were very concerned about the opinions of supersets and least concerned about the opinions of disliked contrasts. They did not distinguish subsets and miscellaneous targets in this regard.

The profile of loadings in Table 4 presents a similar picture. Though Self-Disclosure, Self-Expression, Range of Interaction, and Closeness do a fairly good job of predicting the discriminant function that optimally distinguishes the target types, Concern does the best job. These properties are all highly intercorrelated, as we assumed they would be; they were designed to assess the significance of a target in a subject's life, albeit in different ways. There is some evidence that the impressions of supersets are stable relative to the other targets. (Impression Stability has a loading of .23 on the first discriminant function.) There is no evidence that impressions of supersets are perceived as more complex. Further, subjects did not perceive their feelings as especially ambivalent for any one type of target.

In general, the idea that supersets are significant people seems well supported. On the other hand, we were not able to distinguish subsets from targets classified as miscellaneous. This is not to say that there are not important differences here but only that the properties we happened to construct failed to capture them. Further investigation is needed to distinguish these target persons.

Discussion

In this study, target persons were described in a free-response manner by subjects. Four categories were used in the classification of targets. To recapitulate, supersets (subsets) are target persons who characteristically bear superset (subset) relations to other targets, although these set relationships may not be perfect. Disliked contrasts are characterized by disjunctive logical relationships with other targets. Targets not identified as supersets, subsets, or disliked contrasts were grouped into a miscellaneous category.

The fact that these four kinds of targets are structurally distinct does not guarantee that they will be psychologically distinct. To determine the psychological distinctiveness among these target types, independently measured ratings of the targets were obtained. In terms of the eight rated properties that were used to determine the perceived significance of each target person in the life of the subject, supersets were found to be the most significant people, as predicted, and disliked contrasts were distinguished as the least significant people. Subsets and miscellaneous targets could not be distinguished from one another.

The remainder of the discussion focuses on supersets and addresses two questions: What are supersets? Who are supersets?

What Are Supersets?

This question has already been answered, in terms of the logical relation these targets share as the characteristic way in which they relate to most other targets. To elaborate, supersets receive ascriptions from almost every content category in a subject's person descriptions. A *content category* refers to a group of descriptors (traits and feelings) that are used in a similar way by the subject in describing persons, probably because these descriptors roughly refer to some common quality detected in certain persons and/or some common quality detected (felt) in the self when certain persons are present. The use of hierarchical clustering with the free-response person description protocols is one of the methods for determining a subject's content categories (e.g., Rosenberg, 1977). Nonsupersets—especially subsets—are often described with many of the terms in one content category (or more) and few of the terms in other content categories (Gara, Note 2).

What might these structural properties of the subjects' persons descriptions say about the *process* of person perception? A statement about process is suggested from the structure of these protocols, although no data about process were directly obtained in this study.

Supersets, that is, significant persons in one's life, provide the perceptual (content) categories for the perception of other persons. This idea is certainly not new. It is involved, though stated differently, in many extant psychologies of the person (e.g., Freud, 1940/1949; Kelly, 1955; Sullivan, 1953).

Freud's concept of transference will be discussed briefly as an illustration of this idea that significant persons provide the perceptual categories for the perception of other persons. Transference refers to the situation in psychotherapy where the therapist is seen as a partial replication of a significant person in the patient's life, and the patient acts and reacts with the therapist on that basis. In other words, the therapist is a subset of a significant other. Applied to the broader context of person perception in general, rather than therapist perception in particular, transference simply refers to the experience of persons as subsets of significant others. It is interesting to note here that two of the roles that loom large in classical accounts of transference—the self (*Me Now*) and mother—are each found to be supersets in 8 of 14 cases.

Who Are Supersets?

Although more than 50% of the subjects had Mother and/or Me Now as supersets, it would be misleading to think that in terms of the role descriptions presented in Table 1, supersets only cover a small range, such as parental figures. What is true is that supersets are never mythical, public, or disliked role figures. Furthermore, in a preliminary analysis (Gara, 1978) it was found that one group of subjects had only parental figures as supersets, whereas, in another group, supersets appeared to be mostly nonparental and contemporary figures such as lovers and friends. Another group of subjects seemed to fall between these two extremes with regard to the supersets identified in the protocols.

Regardless of these differences in the role relationship of the supersets to the subjects, supersets were universally found to be very significant persons in the subject's life. What is especially striking is that the ratings of Concern best discriminated supersets from other targets. Subjects are very concerned that the supersets think well of them.

What is the nature of this concern? The answer depends on how the subject interpreted the question. Subjects' ratings on other properties show that they feel close to supersets, that they discuss personal problems with supersets, and that they freely express their feelings with supersets. It could be that subjects' ratings of supersets on Concern reflect their desire that these important people in their lives accept them totally, regardless of shortcomings. In other words, the question could have been interpreted as the desire for unconditional positive regard (Rogers, 1959) from supersets.

On the other hand, there is the possibility that some or all subjects interpreted the question as an assessment of how positive they wanted to appear to a target. In this case, when the subjects report, as they do, that they express themselves freely with supersets, the spontaneity of the expression is limited by their concern that they will appear positive. That is, subjects *monitor* themselves in the presence of supersets. Regardless of what role the superset bears with the subject, even if the superset is the self-concept itself, there is always at least one phenomenological representation of a person that, in effect, monitors the individual. The notion that the self-concept itself can be a monitor is not new: "Most of the ways of behaving which are adopted by the organism are those which are consistent with the concept of the self" (Rogers, 1951, p. 507). The notion of "monitoring" per se is involved in the psychoanalytic concept of introjection of the significant other, which results in the formation of the superego. Elaboration of these notions can be found in object relations theory (Fairbairn, 1954).

Future investigations of the set-theoretical representation of target persons in person perception need to be directed at the unraveling of the subjective meaning(s) of Concern vis-à-vis supersets and the investigation of the possibility that supersets are monitors. It would also be important to determine if a subject's general psychological adjust-

ment and living problems can be differentially affected depending on the role relationship he/she bears with his/her perceived monitor. That is, is it better in terms of general adjustment to have a parental figure or a contemporary figure as a superset/monitor? Are there any people with mythical/public figures as supersets (none in this study), and if so, what are the consequences? We should also not exclude the possibility that all representations of significant others are, in a sense, mythical, and the monitoring function per se that those representations serve is what is problematic in human experience.

Reference Notes

1. Harris, R. J. Personal communication, April 7, 1979.
2. Gara, M. A. *The identification of superordinate and prototypical targets in personality impressions.* Paper presented at the Second International Congress on Personal Construct Theory, Oxford University, Oxford, England, July 1977.

References

Cronbach, L. J. Proposals leading to analytic treatment of social perception scores. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford: University Press, 1958.

Fairbairn, W. R. D. *An object-relations theory of the personality*. New York: Basic Books, 1954.

Francis, R. G. *The rhetoric of science: A methodological discussion of the two by two table.* Minneapolis: University of Minnesota Press, 1961.

Freud, S. [*An outline of psychoanalysis*] (J. Strachey, Trans.). New York: Norton, 1949. (Originally published 1940.)

Gara, M. A. *The identification of superordinate and prototypical persons in personality descriptions.* Unpublished master's thesis, Rutgers University, 1978.

Harris, R. J. The invalidity of partitioned-U tests in canonical correlation and multivariate analysis of variance. *Multivariate Behavioral Research,* 1976, *11,* 353–365.

Hays, W. L. An approach to the study of trait implications and trait similarity. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford, Calif.: Stanford University Press, 1958.

Hinkle, D. N. *The change of personal constructs from the viewpoint of a theory of implications.* Unpublished doctoral dissertation, Ohio State University, 1965.

Huynh, H., & Feldt, L. Conditions under which mean square ratios in repeated measurements designs have exact F-distributions. *Journal of the American Statistical Association,* 1970, *65,* 1582–1589.

Jackson, D. N. The measurement of perceived personality trait relationships. In N. F. Washburn (Ed.), *Decisions, values and groups* (Vol. 2). New York: Pergamon, 1962.

Jackson, D. N., Messick, S. J., & Solley, C. M. A multidimensional scaling approach to the perception of personality. *Journal of Psychology,* 1957, *44,* 311–318.

Jones, L. E., & Young, F. W. Structure of a social environment: Longitudinal individual differences scaling of an intact group. *Journal of Personality and Social Psychology,* 1972, *24,* 108–121.

Kelly, G. A. *A theory of personality: The psychology of personal constructs.* New York: Norton, 1955.

Livingston, J., & Kingsley, S. Query: An interactive computer program for the collection of checklist data. *Behavior Research Methods & Instrumentation,* 1976, *8,* 391–392.

Messick, S. J. The perceived structure of political relationships. *Sociometry,* 1961, *24,* 270–278.

Morrison, D. F. *Multivariate statistical methods.* New York: McGraw-Hill, 1976.

Rogers, C. R. *Client-centered therapy.* Boston: Houghton Mifflin, 1951.

Rogers, C. R. A theory of therapy, personality, and interpersonal relationships as developed in the client-centered framework. In S. Koch (Ed.), *Psychology: A study of a science.* New York: McGraw-Hill, 1959.

Rosch, E. The nature of mental codes for color categories. *Journal of Experimental Psychology: Human Perception and Performance,* 1975, *1,* 303–322.

Rosenberg, S. New approaches to the analysis of personal constructs in person perception. In J. Cole (Ed.), *Nebraska Symposium on Motivation* (Vol. 25). Lincoln: University of Nebraska Press, 1977.

Rosenberg, S., & Sedlak, A. Structural representations of implicit personality theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6). New York: Academic Press, 1972.

Sullivan, H. S. *The interpersonal theory of psychiatry.* New York: Norton, 1953.

Tversky, A., & Gati, I. Studies of similarity. In E. Rosch & B. Lloyd (Eds.), *On the nature and principle of formation of categories.* Hillsdale, N.J.: Erlbaum, 1977.

Warr, P. B., & Knapper, C. *The perception of people and events.* London: Wiley, 1968.

Wish, M. A model for the perception of Morse Code-like signals. *Human Factors,* 1967, *9,* 529–540.

Received October 10, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 12, 2171-2180

Internalization Versus Identification in the Laboratory: A Causal Analysis of Attitude Change

Daniel Romer
University of Illinois at Chicago Circle

Two possible mediators of attitude change, internalization and identification, were investigated in a laboratory setting. Internalization was assumed to underlie change when respondents are attracted to others who hold the same opinion *and* can argue in favor of their attitude position; identification was assumed to mediate change when respondents are attracted to similar others but cannot necessarily support their position. A causal analysis based on these assumptions confirmed the independent existence of internalization and identification as mediators of attitude change. The analysis suggested that internalization involves more valid change than identification does and that attraction toward similar others is affected by both internalization and identification. These conclusions are supported in terms of both individual and treatment variation. The results suggest that attraction toward similar others does not necessarily reflect true attitude change but that valid change can be detected even in laboratory settings.

Distinguishing true attitude change from mere response change is a critical issue in the study of persuasion. Hendrick and Seyfried (1974) approached this problem by assuming that genuine attitude change would be reflected by increased attraction to others who hold the same opinion (Byrne, 1971). There are several objections to this approach, however. First, as Wells (1976) noted, demand characteristics and evaluation apprehension may just as easily affect the measurement of attraction as the measurement of attitude change. Thus, both attitude change and attraction could be disingenuous. Second, it is possible for attraction to be genuine and yet not reflect true attitude change. For example, we may identify (Kelman, 1961) with others who agree with a message without having fully accepted (internalized) the message ourselves. Third, more direct evidence of the persuasive impact of a message should be obtained before labeling its effects as valid. For example, evidence that the message has been internalized should be evaluated.

To overcome these objections, the present research employed a causal modeling approach based on Kelman's (1961) three-process model. This research was conducted to demonstrate (a) that identification and internalization are independent processes with separate determinants and separate consequences; (b) that, as Kelman has theorized, internalization is a more central response to persuasion than identification is and (c) that attraction toward similar others is a function of both identification and internalization. In addition, the research was planned to encompass both treatment and individual variation in response to a persuasive message, thereby providing a sensitive test of the similarity–attraction hypothesis (cf. Hendrick & Bukoff, 1976).

The Causal Model

Causal modeling has its roots in the biometric work of Wright (1921) and was developed to uncover causal mediators in cor-

The author thanks Harry S. Upshaw, two anonymous reviewers, and the action editor for helpful comments on earlier versions of this article.

Requests for reprints should be sent to Daniel Romer, Department of Psychology, University of Illinois at Chicago Circle, Box 4348, Chicago, Illinois 60680.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3712-2171$00.75

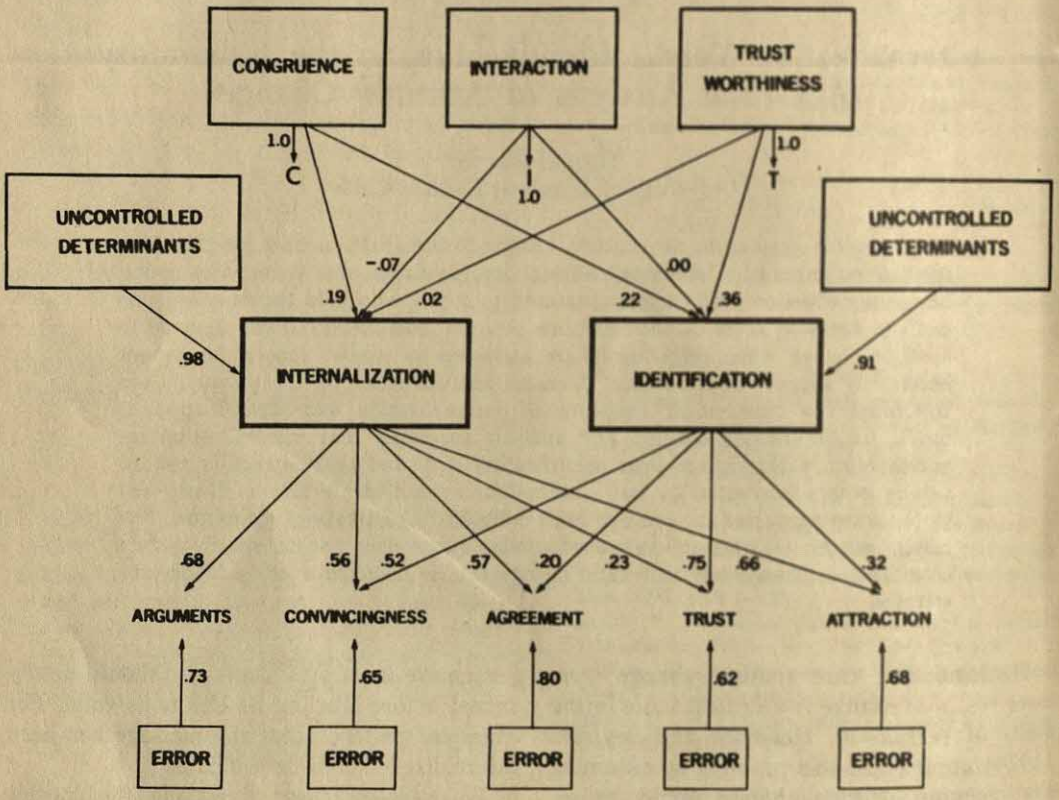


Figure 1. The causal model with observed values for causal paths.

relational research. More recent developments (Goldberger & Duncan, 1973) have extended the method to latent variables and multiple indicators. An important advance in these methods is Jöreskog's (1969; 1973) application of maximum likelihood methods to the analysis of causal models. Jöreskog's method involves the specification of a causal model assumed to underlie a data set and the confirmation of the model with maximum-likelihood factor analysis (cf. Kenny, 1976).

Although causal analysis was originally developed for nonexperimental research, it is also useful for separating latent mediating effects in experimental designs (Alwin & Tessler, 1974; Costner, 1971). In the case of attitude research, agreement with a persuasive message may be mediated by several processes. According to Kelman, the most central (and valid) change occurs when the attitude has been internalized by the recipient. It is possible, however, for change to occur as a result of identifying with the communicator or others who hold the same position. Such change reflects attraction toward others that is based on motives other than the true acceptance of the position.[1]

A causal model representing Kelman's theory as it applies to the present research is shown in Figure 1. Causal links (arrows) between variables represent the relationships that are hypothesized by the present theoretical approach (the coefficients are discussed in the results). Each observed independent and dependent variable is shown to be affected by one or more latent variables or factors (enclosed in boxes). In the model, internalization and identification are assumed to be independent processes that underlie various attitude measures (described below). In addition, the model suggests that experimental

[1] Kelman also hypothesizes the existence of a third process (compliance); however, this mediator is not the focus of the present research.

independent variables may affect one or both of the mediators at the same time. According to the model, it is quite possible that effects of internalization and identification are confounded in ordinary experimental research. The use of causal analysis, however, permits the separation of these mediating processes. In order to employ a causal analysis, it was necessary to manipulate and to measure the mediators in such a way that their causal effects could be distinguished.

Independent variables. If internalization and identification are independent, then each must have its own determinants. Furthermore, to demonstrate that these processes do not merely reflect individual differences, it is necessary to show that they are amenable to experimental variation. Thus, independent variables were chosen that would affect the processes and facilitate their separation.

For internalization, the variable that fulfilled these purposes was the extent to which the persuasive message contained arguments that were congruent with recipients' values. The more congruent the arguments, the more readily they can be internalized. It is also possible, however, for this manipulation to affect identification. Communicators who use more value-congruent arguments might also appear attractive over and beyond the extent to which their arguments are internalized. Thus the congruence manipulation is shown as a potential determinant of both internalization and identification in the diagram.

A second variable, communicator trustworthiness, was manipulated to affect identification. Communicators who are trustworthy should be more likely to induce identification than ones who are not. At the same time, communicator trustworthiness might also affect internalization, so this causal link was also tested in the model. Trustworthy communicators may facilitate internalization because their arguments are more credible.

Finally, the product of the independent variables was included because the variables might interact in affecting either internalization or identification.

Dependent variables. It was also critical in the present research to distinguish internalization from identification at the response level. To accomplish this goal, an effect of internalization that would not be produced by identification was measured. This variable measured the extent to which respondents' attitudes are justified by the beliefs in their own value system. The justification could involve (a) beliefs that the advocated position produces benefit (proarguments) or (b) beliefs that it produces harm (counterarguments). Both of these beliefs can be represented in an expectancy–value model of attitude structure (Peak, 1955; Rosenberg, 1956) in which proarguments correspond to positive expectations of the advocated position and counterarguments to negative expectations. Thus, a single argument index is shown affected only by internalization in the diagram.

Two typical indices of attitude change (agreement with and convincingness of the message) were measured and are shown affected by both internalization and identification in the diagram. A single index of communicator trustworthiness was included to validate the corresponding manipulation. It was assumed to reflect internalization as well as identification because a communicator who presents more value-congruent arguments should also be seen as more trustworthy.

Finally, an index of attraction toward a similar other was measured. The other was described as a fellow subject who agreed with the advocated position. It was assumed that to the extent that subjects also agreed with the advocated position, their similarity and attraction toward the fellow subject would be greater. A major implication of the present model is that this process occurs as a consequence of both internalization and identification and therefore is not a pure measure of true attitude change.

Uncontrolled variation. Also shown in the diagram are sources of variation in the dependent variables that were not under experimental control. These include individual differences in internalization and identification that could be sizable. It is desirable to analyze the present model in terms of both treatment and individual variation, since similarity–attraction is so sensitive to individual variation. Furthermore, subjects' value systems

could be expected to differ greatly, producing considerable individual variation in internalization. Finally, a certain amount of error is shown affecting each dependent variable. The errors were assumed to be mutually uncorrelated, an assumption that should be (and was) evaluated.

Method

Subjects and Design

One hundred and sixty undergraduates fulfilling course requirements in introductory psychology were randomly assigned to four conditions of an orthogonal design involving two levels of value congruence and two levels of communicator trustworthiness. Equal numbers of males and females were assigned to each condition.

Experimental Variables

The intent of the manipulations was to produce change in subjects' attitudes toward the admittance of Puerto Rico to the Union. The value-congruence manipulation involved variation in the acceptability of a message that argued in favor of making Puerto Rico the 51st state. The less congruent message, an adapted version of one used by Watts and McGuire (1964), referred to the need to establish military bases in Puerto Rico and to replace those in Cuba, the international propaganda value of admitting a state composed of a minority population, and the economic advantages of eliminating import duties upon Puerto Rican goods. The more congruent message, which was written by the author, claimed that statehood would ensure a higher minimum wage for the underpaid and exploited work force, a tax structure more suited to the needs of the people, and a voice in Congress to affect the decisions that are now made without Puerto Rico's direct representation. The messages were approximately equal in length, and pretesting confirmed that they were perceived as equally well written.

The arguments were presented as a statement given by a witness for a congressional committee hearing on Puerto Rico's present and future status. The communicators were described as either a distinguished professor from the Yale law faculty who was an expert on Latin America or as a director of public relations for a land development corporation located in Puerto Rico. The public relations director was assumed to be less trustworthy because his firm might benefit from Puerto Rico's statehood; furthermore, it seemed reasonable that university students would see a law school professor as trustworthy and would be more inclined to identify with him than with a public relations person.

Procedure

Subjects were assembled in groups of 6 to 12 and were told that the experiment was concerned with their "ability to process and evaluate information." They were asked, in written instructions, to read the message carefully and to complete a questionnaire that followed the message. The questionnaire contained 14-category scales for rating the convincingness of the arguments, agreement with the advocated position, and the trustworthiness of the source.

When everyone had completed the questionnaire, the argument and attraction questionnaires were administered, their order being counterbalanced within experimental conditions. The argument questionnaire was constructed to increase the likelihood that a respondent's ability to support the advocated position would be measured. Previous research by the author (Romer, 1979) has suggested that "thought listing" procedures such as the one used by Brock (1967) do not adequately represent a respondent's repertoire of arguments and that a questionnaire specifically requesting that respondents list as many arguments as they can is a more direct measure of their ability to argue for or against an attitude position. Therefore subjects were not simply asked to list their "thoughts and ideas" about the advocated position. Instead, they were asked to consider that both positive and negative consequences can follow from an event. They were given as an example the devaluation of the dollar, which might lead to greater exports but higher prices for imported goods. Following the example, they were asked to list as many positive and negative consequences of admitting Puerto Rico as they could within a 5-minute period. At the end of the 5 minutes, both the desirability and the likelihood of the consequences listed were rated. Positive desirability was indicated by a scale of 1 to 5 ranging from "only slightly desirable" to "extremely desirable." Undesirability was rated from −1 ("only slightly undesirable") to −5 ("extremely undesirable"). Likelihood judgments were requested on a scale from 1 ("not very likely") to 5 ("extremely likely").

The attraction questionnaire was designed to measure attraction in a relatively unobtrusive manner. Rather than asking subjects how much they would like a person whose attitude was completely similar to theirs (as in Hendrick & Seyfried), the questionnaire asked subjects to imagine another subject in an experiment similar to the one in which they were participating who expressed "strong agreement" with the message they had read. Once they had "formed a picture of this person," they were asked to describe the person in terms of the likelihood that she or he possessed each of 10 traits. The ratings were made on a 10-category scale ranging from "not very likely" to "extremely likely." The traits, taken from Anderson (1968), covered the entire range of likableness and were all high in judged meaningfulness.

At the conclusion of the experiment, subjects were

invited to ask questions about the experiment and its purpose. Although the content and the occurrence of questions varied considerably across sessions, all subjects were promised written, posted feedback about the results of the experiment.

Measurement of Dependent Variables

Three of the dependent variables were simply taken from the category ratings of convincingness, agreement, and trustworthiness. Convincingness was rated in response to the question "How convincing were the arguments in the statement?" ("not very convincing" to "very convincing"). The agreement item asked "To what extent do you agree that Puerto Rico should be admitted as the 51st state?" with "strongly agree" to "strongly disagree" as polar descriptions; trustworthiness was assessed with "How truthworthy do you regard the witness to be with respect to the Puerto Rican issue? ("not very trustworthy" to "very trustworthy").

The argument index was formed by multiplying the desirability rating by the likelihood rating for each consequence that a subject listed and taking the arithmetic sum of these products as the overall index for a subject. In a similar fashion, the attraction index was obtained by multiplying the likelihood ratings by a transformation of the scale values of likableness tabled by Anderson (1968) for each of the traits. The tabled values range from 0 to 6, covering the entire range of likableness. For the purpose of this index, however, a value of 3 was assumed to reflect neutrality, with values greater than 3 deviating in a positive direction and values less than 3 in a negative direction. These deviation scores defined the trait values that were multiplied by the likelihood judgments in forming the overall index.

Analysis

The causal diagram presented in Figure 1 corresponds to an oblique five-factor model in Jöreskog's system. In this model, observed variables in the diagram (Y) are written as a function of latent factors (F) and error (E):

$$Y = AF + DE,$$

in which A is a matrix of factor loadings and D is a diagonal matrix of error coefficients. This model predicts the correlation matrix of independent and dependent variables (C) from

$$C = ABA' + D^2,$$

in which B is a matrix of interfactor correlations. The parameters of the A and B matrices as defined by the causal model are shown in Table 1.

As is evident, many parameters were predefined as zero. These restrictions are implied by the causal model and indicate that the variable or factor is predicted not to correlate with the relevant factor. Parameters that were meant to be estimated are left blank. These correspond to the causal links that were hypothesized in the diagram. Two of the restrictions are notable: (a) The loading of the argument index on the identification factor was set to zero, reflecting the assumption that identification is independent of the extent to which respondents can argue in favor of the advocated position, and (b) the correlation between internalization and identification was set to zero, reflecting the assumption that identification is independent of internalization. All of the unrestricted correlations in B refer to potential effects of the independent variable factors upon internalization and identification. The unrestricted loadings in A refer to links between factors and the various variables.

An important aspect of the present model concerns the fact that the parameters estimated under it are uniquely defined. In ordinary factor analysis, the solution is arbitrary because the factors can be rotated without violating the assumptions of the model. In the present model, the factors are defined (by the zero loadings in the A matrix) such that no other estimates of the free parameters will fit the model as well as the ones that are obtained.[2] The

Table 1
Parameters Defined by Causal Model

Variable/factor	Factor 1	2	3	4	5
Matrix *A*					
Congruence[a]	—	0	0	0	0
Trustworthiness[a]	0	—	0	0	0
Interaction[a]	0	0	—	0	0
Convincingness	0	0	0	—	—
Agreement	0	0	0	—	—
Trust	0	0	0	—	—
Arguments	0	0	0	—	0
Attraction	0	0	0	—	—
Matrix *B*					
1. Congruence	1.0				
2. Trustworthiness	0.0	1.0			
3. Interaction	0.0	0.0	1.0		
4. Internalization	—	—	—	1.0	
5. Identification	—	—	—	0.0	1.0

Note. The far left column lists variables for Matrix A and factors for Matrix B.
[a] Independent variable coded 1, −1.

[2] According to Jöreskog (1969) the criteria for a unique solution are that there are at least k^2 (where k = the number of factors) restrictions in the A and B matrices and that the rows of the A matrix can be permuted such that restrictions on the factors lie in the upper triangle of the matrix. Both of these criteria are met by the present model.

Table 2
Correlation Matrix and Means and Standard Deviations of Independent and Dependent Variables

Variable	1	2	3	4	5	6	7	*M*	*SD*
1. Value congruence								0.00	1.00
2. Communicator trustworthiness	000							0.00	1.00
3. Interaction	000	000						0.00	1.00
4. Convincingness	247	187	−035					8.42	3.73
5. Agreement	082	139	−024	426				8.67	3.47
6. Trust	193	285	−016	516	272			7.72	3.54
7. Arguments	135	020	−046	368	397	155		15.37	46.07
8. Attraction	216	095	−058	535	429	397	466	16.93	22.56

solution, therefore, is the best that can be obtained, given the restrictions that define the factors.

There are a number of computer programs presently available to solve this causal system. The one used in this analysis was Jöreskog's analysis of covariance structures (Jöreskog, Gruvaeus, & Van Thillo, 1970). This program, as well as others, employs a large sample chi-square test of fit based on the maximum-likelihood principle. The test of fit measures the extent to which the predicted correlation matrix matches the observed matrix. The program produces a solution that minimizes the discrepancy between these matrices; however, a particular model may or may not fit the observed matrix. If a model produces a good fit, it can then be considered a candidate for explaining the data set.

Results

The correlation matrix for both independent and dependent variables is presented in Table 2. Rather than focusing upon specific correlations in the matrix, the present analysis attempts to account for the pattern of correlations in the entire matrix. Before turning to the analysis, however, it is worth asking whether the present matrix fully accounts for the effects that were observed. In particular, the order of measuring the argument and attraction indices was counterbalanced within experimental conditions, and equal numbers of males and females were studied in all cells of the design. However, neither order of measurement nor sex of subject produced any effects upon the variables in the matrix. It was assumed, therefore, that the data set was adequately described by the correlations in the matrix.

The results of the covariance structure analysis are shown in Figure 1. The first question one can ask is whether the correlation matrix is adequately reproduced by the model. The answer appears to be yes. The chi-square test of fit between the predicted and obtained correlation matrix was nonsignificant, $\chi^2(10) = 3.55$, $p = .96$, indicating a close correspondence between predicted and obtained values. Inspection of the differences between these values revealed that no residual was greater than .07 in absolute value ($M = .006$, $SD = .021$). Thus, the model does a good job of predicting the data.

As is evident in the figure, factors corresponding to both internalization and identification mediated variation in the dependent variables. All of the dependent variables loaded positively upon internalization, indicating that respondents who had positive attitudes also thought the communicator was trustworthy, had arguments to support their position, and were attracted toward others who agreed with the advocated position. A factor corresponding to identification, however, was orthogonal to internalization and mediated variation that was independent of arguments that supported the advocated position. It contained loadings from all of the other dependent variables, indicating that the more respondents agreed with the advocated position, the more they trusted the communicator and the more attracted they were to others who agreed with the advocated position. Thus, as expected, attraction toward others was affected by both internalization and identification.

An important assumption of the causal model is the theoretical independence of internalization and identification. As they are

defined, both involve variation in expressed attitudes and attraction toward those who hold the same attitudes, but only internalization is based on arguments in support of the attitude position. Although the goodness-of-fit test indicates that this assumption is plausible, it is also possible to test the model without making the assumption. In this test, the model is fitted without the B-matrix restriction that the factors are uncorrelated. The obtained solution, however, was virtually identical to the one in Figure 1, and the factors were still orthogonal ($r = -.02$). It appears, therefore, that relaxing the assumption does not improve the fit of the model and that the assumption is valid.

In evaluating the effects of the independent variables upon internalization and identification, a special procedure was employed. The obtained correlation (shown in the figure) was set to zero, and the remaining parameters of the model were fixed to their obtained value. The model was then tested to see if the goodness of fit of the entire model was significantly poorer than the fit of the original model. If it is, then the correlation can be considered to be different from zero.

As expected, the value congruence manipulation affected internalization, $\chi^2(1) = 4.38$, $p < .05$. The message that contained more congruent arguments was more readily internalized. There were no other effects upon internalization, however. Apparently, the trustworthiness of the communicator did not affect the credibility of his persuasive arguments.

The trustworthiness manipulation did affect identification, however, $\chi^2(1) = 14.03$, $p < .05$. The more trustworthy communicator apparently induced greater change by virtue of his attractiveness. In addition, the value congruence of the message affected identification, $\chi^2(1) = 5.42$, $p < .05$. Thus the congruence manipulation produced change mediated by both internalization and identification. The interaction of the independent variables did not affect identification, however.

As is evident in Figure 1, uncontrolled sources of variation in both internalization and identification were sizable. This indicates that most of the variation in the mediators was due to individual differences. Nevertheless, the success of the model suggests that both treatment and individual variation can be encompassed by the same causal model.

The estimated errors shown in the Figure indicate that from 40% to 64% of the variation in the dependent variables was unexplained.[3] The assumption that these errors of measurement are uncorrelated does not seem unreasonable in view of the very close fit of the model. That is, there does not appear to be any reliable variation in the dependent variables that was not explained by the internalization and identification factors. Furthermore, the finding that both factors were affected by the independent variables suggests that the factors are not solely composed of error. Although the error variation in the variables was sizable, all of the remaining error-free variation (36% to 60%) was predicted by the model.

Rival Models

Even though a particular model does well in predicting data, it is always possible that an alternative model is superior. One advantage of causal analysis, however, is that the assumptions underlying a model are made explicit so that different models can be tested. For example, it is possible that all of the reliable variation in the dependent variables is accounted for by a single internalization factor. Therefore, a model containing only internalization was fitted to the data. This model was identical to the favored model with the exception that the identification factor was not included. This model did not fare well, however, by the goodness-of-fit test, $\chi^2(17) = 27.67$, $p < .05$. Furthermore, the two-factor model was a significantly better predictor of the correlation matrix than the internalization model was, $\chi^2(7) = 24.12$, $p < .05$. It appears, therefore, that a model involving both internalization and identification is a better explanation of the data than a model involving only internalization is.

---

[3] The estimated proportion of variance due to error in each dependent variable is equal to the square of the corresponding error coefficient.

The possibility exists, nevertheless, that a different two-factor model also accounts for the data. To evaluate this possibility, alternative models that contained different restrictions on the second dependent variable factor were constructed. These included (a) restrictions on both attitude measures so that the second factor only contained loadings from trust, arguments, and attraction; (b) a restriction on trust so that all of the remaining variables could load on the factor; and (c) a restriction on attraction, with the remaining dependent variables free. Although each of these models provided a good fit to the data, the solutions that were obtained were uninterpretable. Two of them (a and c) implied that the variables on the second factor were inversely correlated but positively affected by the independent variables; the third implied that congruence had a negative impact on all the variables of the second factor. Thus although these models could fit the data mathematically, they did so at the expense of plausibility and interpretability. This is not surprising, since there were no obvious theoretical models that would support these solutions. Nor could the second factor in these solutions be attributed to correlated measurement error: The factors in each model were correlated with the independent variable factors suggesting that if these models were valid, the factors were not entirely composed of error. Thus, the tests of these alternative models suggest that the internalization–identification model is the best explanation of the data.

Discussion

There were three things to be demonstrated in the present research. The first was that internalization and identification independently mediate attitude change. This, of course, was the rationale underlying the two-factor model. This model did significantly better in predicting the data than did a model containing only internalization as a mediator. Furthermore, the separate effects of the independent variables upon internalization and identification indicate that each factor reflected valid variation in the dependent variables. In fact, the value-congruence manipulation affected both mediators, but trustworthiness only affected identification. Finally, a test of the model without the orthogonality assumption still affirmed the independence of the factors. Thus, two orthogonal factors appeared to mediate attitude change in the present laboratory setting.

The second goal of the research was to show that internalization is a more central process than identification is in mediating attitude change. This was demonstrated by restricting the argument index from loading on the identification factor. Since subjects were asked to generate as many positive and negative arguments as they could (in 5 minutes), the argument index has face validity as a measure of a respondent's ability to support the advocated position. Therefore only the internalization mediator reflected variation in the perceived consequences and justification of the advocated position. Although the argument index was only one of many possible measures of internalization, it did display validity by correlating with the other measures. Furthermore, the finding that trustworthiness affected identification but not internalization indicates that there were effects that were mediated by processes other than internalization. Therefore, even if a more sensitive measure of internalization had been used, the present results suggest that a separate factor would still be needed to explain variation that is unrelated to internalization.

The third goal of the research was to show that attraction toward similar others is a function of both internalization and identification. This hypothesis was clearly supported. The attraction index loaded on each factor, indicating that variation in attraction was affected by both internalization and identification. Apparently, we express acceptance of an advocated position not only because we can argue in favor of it but also because we identify with others who hold the same opinion. Therefore, Hendrick and Seyfried (1974) were justified in measuring attraction as an indicator of true attitude change. However, the present results show that without evidence about whether or not

internalization has occurred, it is not possible to know whether attraction reflects internalization or identification.

Wells' (1976) charge that attitude change and attraction may both be caused by experimental artifacts (demand characteristics and evaluation apprehension) seems an unlikely explanation for the present results. Since attitudes and attraction loaded on each of *two* factors, one would have to postulate a complex pattern of compliance motives to account for both factors. In the case of internalization, subjects would be assumed to have role played greater acceptance of the more congruent message. However, since all subjects received a persuasive message of comparable length and writing quality, demands for expressing acceptance of the advocated position should have been equal in both congruence conditions. Furthermore, the instructions for generating arguments requested subjects to list as many arguments as they could. If anything, this instruction should have induced demands to be prolific rather than to be favorable toward the advocated position. Finally, if the internalization factor merely reflects compliance, it is unclear why the factor would not also be affected by source trustworthiness.

Whether the identification factor reflects experimental artifacts is less certain. However, attraction was measured less obtrusively than in Hendrick and Seyfried (1974), and, even if compliance contributed to the identification factor, the conclusion that more than one form of attitude change mediated the results would still be tenable. Thus although it is impossible to rule out completely a compliance explanation, it seems unlikely that compliance can account for the entire pattern of the present results.

One implication of the present approach is that, if subjects' attitudes were retested at a time when the message and source were less salient, attitudes based on internalization should be more stable than those based on identification. Kelman (1958) used this assumption to show that compliance, identification, and internalization could be distinguished. The present results are consistent, therefore, with research on the persistence of attitude change in which variation due to communicator trustworthiness decays at a different rate from "message-only" variation once "dissociation" of the source has occurred (Gruder, Cook, Hennigan, Flay, & Halamaj, 1978; Hovland & Weiss, 1951). The basis of this differential decay could be that trustworthiness is mediated by identification more than by internalization.

One advantage of the present causal modeling approach is that both individual and treatment variation could be tested in the same analysis. It was possible, therefore, to observe the similarity–attraction relationship at both the individual and treatment levels. Although the variation in individual attitudes and attraction was far greater than the treatment variation, both appeared to be subsumed by the same causal processes.

The present results indicate that laboratory-produced attitude change may be valid. This is cause for optimism on at least two grounds. First, although the study of attitude change was once a central issue in social psychology, interest in the topic has waned. Researchers doubt the validity of studying attitude change in the laboratory, and the results that have been obtained are regarded with suspicion. Although there are probably many reasons for the reduced interest, the present results suggest that suspicion of the laboratory is not always justified. Second, if genuine attitude change can be distinguished from other forms of change, the study of attitude change can proceed despite the belief that laboratory-induced persuasion is suspect. For example, research can be conducted further to determine how and when a message is internalized. Thus even if the many variables that have been studied only produce surface change, this realization need not impede progress toward determining the factors that cause long-lasting and meaningful change in peoples' attitudes.

References

Alwin, D. F., & Tessler, R. C. Causal models, unobserved variables, and experimental data. *American Journal of Sociology*, 1974, *80*, 58–86.

Anderson, N. H. Likableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology*, 1968, *9*, 272–279.

Brock, T. C. Communication discrepancy and intent to persuade as determinants of counterargument production. *Journal of Experimental Social Psychology*, 1967, *3*, 296–309.

Byrne, D. *The attraction paradigm*. New York: Academic Press, 1971.

Costner, H. L. Utilizing causal models to discover flaws in experiments. *Sociometry*, 1971, *34*, 398–410.

Goldberger, A. S., & Duncan, O. D. (Eds.). *Structural equation models in the social sciences*. New York: Seminar Press, 1973.

Gruder, C. L., Cook, T. D., Hennigan, K. M., Flay, B. R., & Halamaj, J. Empirical tests of the absolute sleeper effect predicted from the discounting cue hypothesis. *Journal of Personality and Social Psychology*, 1978, *36*, 1061–1074.

Hendrick, C., & Bukoff, A. Assessing the reassessment of the validity of laboratory-produced attitude change. *Journal of Personality and Social Psychology*, 1976, *34*, 1068–1075.

Hendrick, C., & Seyfried, B. A. Assessing the validity of laboratory-produced attitude change. *Journal of Personality and Social Psychology*, 1974, *29*, 865–870.

Hovland, C. I., & Weiss, W. The influence of source credibility on communication effectiveness. *Public Opinion Quarterly*, 1951, *15*, 635–650.

Jöreskog, K. G. A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, 1969, *34*, 183–202.

Jöreskog, K. G. A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences*. New York: Seminar Press, 1973.

Jöreskog, K. G., Gruvaeus, G. T., & Van Thillo, M. ACOVS: A general computer program for analysis of covariance structures. Research Memorandum 70–15. Princeton, N.J.: Educational Testing Service, 1970.

Kelman, H. C. Compliance, identification and internalization: Three processes of attitude change. *Journal of Conflict Resolution*, 1958, *2*, 51–60.

Kelman, H. C. Processes of social influence. *Public Opinion Quarterly*, 1961, *25*, 57–78.

Kenny, D. A. An empirical application of confirmatory factor analysis to the multitrait-multimethod matrix. *Journal of Experimental Social Psychology*, 1976, *12*, 247–252.

Peak, H. Attitude and motivation. In M. R. Jones (Ed.), *Nebraska Symposium on Motivation* (Vol. 3). Lincoln: University of Nebraska Press, 1955.

Romer, D. Distraction, counterarguing and the internalization of attitude change. *European Journal of Social Psychology*, 1979, *9*, 1–17.

Rosenberg, M. J. Cognitive structure and attitudinal affect. *Journal of Abnormal and Social Psychology*, 1956, *53*, 367–372.

Watts, W. A., & McGuire, W. J. Persistence of induced opinion change and retention of the inducing message contents. *Journal of Abnormal and Social Psychology*, 1964, *68*, 233–241.

Wells, G. L. Reassessing the validity of laboratory-produced attitude change. *Journal of Personality and Social Psychology*, 1976, *34*, 1062–1067.

Wright, S. Correlation and causation. *Journal of Agricultural Research*, 1921, *20*, 557–585.

Received October 17, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 12, 2181-2199

Attitudes and Cognitive Response: An Electrophysiological Approach

John T. Cacioppo
University of Iowa

Richard E. Petty
University of Missouri—Columbia

Two experiments employed electrophysiological procedures for assessing the covert information-processing activity of message recipients. In Experiment 1, 24 subjects expected to hear discrepant communications and were requested to "collect their thoughts" following each forewarning. As discrepancy increased, anticipatory counterargumentation increased, whereas production of favorable thoughts and agreement decreased. In addition, following forewarnings oral muscle, cardiac, and respiratory activity increased, whereas nonoral muscle activity remained constant and quiescent. In Experiment 2, 60 subjects anticipated and heard a proattitudinal, a counterattitudinal, or a neutral communication. They evaluated more positively and generated more favorable thoughts and fewer counterarguments to the proattitudinal than to the counterattitudinal advocacy, but rated similarly the neutral and proattitudinal advocacies. As in Experiment 1, incipient oral muscle activity increased following the forewarning of an involving *counterattitudinal* advocacy; it also increased for all conditions during the message. Patterns of subtle facial muscle changes reflected the affective nature of the cognitive responding before and during the message. These results provide evidence that electrophysiological assessments offer objective, concurrent, and independent measures of cognitive response in persuasion, and support the notion of recipients as active information processors when topic involvement is high.

Recent work in attitude change has emphasized the manner in which persons process the information contained in persuasive messages. Investigators have studied how such variables as source credibility (Cook, 1969; Gillig & Greenwald, 1974), distraction (Petty, Wells, & Brock, 1976), message repetition (Cacioppo & Petty, 1979), message comprehensibility (Eagly, 1974), forewarning of persuasive intent (Petty & Cacioppo, 1979a) and of topic and position (Petty & Cacioppo, 1977), number of arguments employed (Calder, Insko, & Yandell, 1974), issue involvement (Petty & Cacioppo, 1979b), heart rate (Cacioppo, 1979), group discussion (Burnstein & Vinokur, 1977), and so forth affect the profile of cognitions (e.g., counterarguments, favorable thoughts, neutral thoughts), and attitude change. Theoretical interest in the influence on persuasion of a person's idiosyncratic cognitive responses to an advocacy is certainly not new (cf. Hovland, Lumsdaine, & Sheffield, 1949), but the present level of research activity in the area marks a shift in emphasis toward this approach (cf. Petty, Ostrom, & Brock, in press). The reinvigorated interest in these covert thought processes stems in part from the apparent inability of classical learning theories to provide parsimonious accounts of observed attitude changes (Greenwald, 1968; Petty, 1977),

The authors thank Anthony G. Greenwald, Curt A. Sandman, and the action editor for comments concerning this research; Charles W. Snyder for technical advice and assistance; Keith Richenbacher, Debra Horowski, and June Chen for assistance in Experiment 1; and Leo Quintanar, Kevin Smith, and J. P. Holbrook for assistance in Experiment 2. This material is based on work supported in part by National Science Foundation Grant BNS-7818667.

Requests for reprints should be sent to John T. Cacioppo, Department of Psychology, University of Iowa, Iowa City, Iowa 52242, or to Richard E. Petty, Department of Psychology, University of Missouri, Columbia, Missouri 65211.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3712-2181$00.75

and in part from improvements in measurement procedures (Brock, 1967; Cacioppo, Harkins, & Petty, in press).

We now know that the level of cognitive responding to an advocacy can vary from complete inattention or simple structural (e.g., orthographic) analyses of the message to extensive semantic examination and cognitive elaboration. The latter, but not the former, allows recipients to determine the implications of an event for themselves and for significant others (Cacioppo, Glass, & Merluzzi, 1979; Craik & Tulving, 1975; Janis & Mann, 1977; Rogers, Kuiper, & Kirker, 1977) and appears to be an important determinant of attitude change (cf. Petty et al., in press). On the other hand, the cognitive response (information-processing) approach to persuasion has been questioned in recent years (Baron, Baron, & Miller, 1973; Langer, Blank, & Chanowitz, 1978; Miller & Baron, 1973; Romer, 1979). The self-report measures of cognitive response have been criticized for placing post hoc demands on subjects to report rational(izing) processes for their attitude change. These critics have called for direct, concomitant measures of cognitive activity in persuasion.

Perhaps the most commonly used measure of cognitive response in persuasion is the "listed thoughts" procedure developed by Brock (1967) and Greenwald (1968). Inspection of the existing data regarding the reactivity and reliability of the thought-listing procedure is encouraging. The same pattern of attitudinal responses has been reported in groups that list and do not list their thoughts (Petty & Cacioppo, 1977), and the order of attitude and cognitive response assessment has not affected either measure (Calder et al., 1974; Petty et al., 1976). Third, split–half and test–retest reliability of attitude and cognitive response measures are satisfyingly high (Cullen, 1968). Nevertheless, it can still be argued that requesting subjects to list their thoughts produces responses that would not and do not occur naturally.

Hence, we believed it worthwhile to develop a concurrent measure of cognitive and affective response that could be obtained without the subject's knowledge of its purpose and that would not require the subject to emit a voluntary response. An electrophysiological approach seemed promising:

> It is possible that a hypothetical construct, hypothesized to be a covert response (r), may be directly measured as an electromyogram that occurs after the external stimulus, but before an overt response—(indeed) the covert electromyographic response may be an antecedent or determiner of the following overt response. (McGuigan, Culver, & Kendler, 1971, p. 146)

Thus, by studying electrophysiologically the microprocesses operating in a persuasion setting, much may be learned (i.e., confirmed or disconfirmed) about the presence and nature of the postulated cognitive-response processes.

Fortunately, psychophysiological research has yielded a few biocognitive relationships that make the present endeavor simpler and more assuredly informative. McGuigan and his colleagues, for instance, have amassed substantial data illustrating the sensitivity and reliability of the electromyogram (EMG), particularly of the speech muscles, as an index of covert linguistic activity (for reviews, see Garrity, 1977; McGuigan, 1970, 1973b, 1978; Sokolov, 1972). Also, cardiac activity has been shown to accompany covert information processing, with a decelerated heart rate characteristic of sensory acuity or reception, and an accelerated heart rate characteristic of cognitive elaboration (e.g., Cacioppo & Sandman, 1978; Lacey, Kagan, Lacey, & Moss, 1963; Schwartz & Higgins, 1971). A brief overview of these areas of research is provided next.[1]

Electromyographic Specificity and Cognitive Responding

A conception held generally about the EMG activity of the speech muscles (when this

[1] This discussion is certainly not exhaustive of the psychophysiological responses indicative of mental activity. A few notable exclusions are the pupillary (Goldwater, 1972; Kahneman, 1973) and averaged evoked (cortical) responses (Posner, 1975). The interested reader may wish to consult McGuigan and Schoonover (1973) or Greenfield and Sternbach (1972) for additional information about these measures.

tivity reflects subvocal speech) is that the sultant afferentiation traveling from these uscles to the language centers of the cortex an aid to stimulus comprehension, elabora- on, and abstraction (Hardyck & Petrinovich, 969; McGuigan, 1973a, 1978; Sokolov, 967, 1972). Specifically, the afferentiation om the speech muscles appears to serve as nonmeaningful but redundant phonetically ded representation of the immediately pre- ding efferent language commands, which riginate in the cortex and result from (some- mes preliminary) processing of external and emorial stimuli. This redundancy serves to rengthen, in a manner similar to that of gnitive rehearsal, the links between the ex- rnal event and the internal representations. ltimately, as the links between the external imulus and the resulting (overt) response come established (e.g., automated), less dundancy is needed and less speech EMG tivity is in evidence.

Consistent with these notions, Sokolov 1969) reported that subjects performing a mplex matrix manipulation exhibited bursts covert oral EMG activity at first. Repeated ials with the task, however, resulted in a turn of the EMG activity to basal levels. milarly, nonproficient readers display reater oral EMG activity while reading than proficient readers—but both groups dis- ay elevated oral EMG activity when read- g difficult material (McGuigan, 1967; Mc- uigan & Bailey, 1969b; McGuigan, Keller, Stanton, 1964); greater oral EMG activity present when persons read silently than hen they process nonlinguistic materials .g., music); and the silent processing of oth linguistic and nonlinguistic stimuli leads increased oral EMG activity compared to aseline measures (McGuigan & Bailey, 969a).

Further, these changes in oral EMG do not flect simple changes in the general somatic autonomic arousal of the organism: A con- omitant increase in EMG activity is not ob- rved in irrelevant (e.g., nonlinguistic) uscle groups (cf. McGuigan, 1970; Sokolov, 969); nor has the galvanic skin response een related to oral EMG during the silent erformance of language tasks (Sokolov, 1969, 1972); and quickened breathing is observed during silent reading, presumably reflecting subvocalization (McGuigan et al., 1964; McGuigan & Rodier, 1968). Finally, the EMG responses of the speech muscles are highly specific. Collectively, these results suggest that oral EMG activity *can* reflect information processing in persuasion.[2]

Cardiac and Cognitive Activity

Contemporary investigations of changes in heart rate and covert information processing have included: (a) monitoring heart rate and performance during the anticipation and performance of a wide variety of cognitive (e.g., sentence generation) and sensory (e.g., viewing flashing lights) tasks (e.g., Lacey, 1959; Lorens & Darrow, 1962); (b) the measurement of the differences in heart rate change during task performance by individuals who differ dispositionally in their mode of information processing for that task (e.g., Blatt, 1961); (c) the manipulation of attributes of the task (e.g., processing requirements) while monitoring heart rate and accompanying physiological responses (e.g., Blaylock, 1972; Tursky, Schwartz, & Crider, 1970); (d) monitoring heart rate and general somatic activity during the anticipation and performance of tasks under varying levels of incentive or motivation (e.g., Elliott, 1974; Obrist, Webb, Sutterer, & Howard, 1970); (e) monitoring sensory thresholds, reaction time, or cognitive and attitudinal responses following endogenous changes in heart rate (e.g., Cacioppo,

[2] Festinger and Maccoby (1964) were among the first persuasion researchers to discuss the subvocalizations of a recipient in response to a persuasive appeal. Keating and Brock (1974) pursued the notion that a recipient's subvocalizations may contain rich evidence regarding what was happening to the stimulus between the time the subject heard it and responded overtly to it. However, these efforts were largely unsuccessful empirically, due to the insensitive procedures employed; no direct electrophysiological measurement of speech EMG activity in a persuasion setting has ever been reported. Moreover, previous discussions of subvocal activity in persuasion have focused *solely* upon the *counterarguing* process, whereas we focus here on more general cognitive-response processes.

Sandman, & Walker, 1978; Edwards & Alsip, 1969); and (f) observing the cognitive and attitudinal effects of exogenously manipulated heart rate (Cacioppo, 1979). The cardiac response generally covaries with the cognitive requirements or difficulty of the task, particularly when the task requirements are substantial. Although considerable controversy still exists in psychophysiology regarding the neurophysiological and/or biological processes controlling the heart rate responses observed in the aforementioned studies, the empirical link between the cardiac response and cognitive activity seems well established. Whether the cardiac acceleration observed during the performance of cognitive tasks is initiated by a metabolic control center (cf. Obrist et al., 1970) or by a modulating negative feedback system (cf. Lacey, 1967) is unimportant here. Central to the present study is the finding that the cardiac response reflects considerable changes in cognitive activity, even when these changes are spontaneous and self-induced (Schwartz, 1971; Schwartz & Higgins, 1971).

Methodological Considerations

Several methodological safeguards were instituted here to assure that the electrophysiological measures we observed reflected covert information processing. First, a measure of general (nonlinguistic) somatic activity was obtained in addition to the measures of oral EMG and cardiac activity. To the extent that the activation is specific to the speech muscle fibers and heart rate when processing an external event, we are more confident that we have tapped covert processing rather than irrelevant (i.e., unreliable) movements, postural shifts, and so forth. Second, the level of heart rate and EMG activity observed while anticipating and processing the stimulus was compared with prestimulus measures to assess whether or not electrophysiological "response" actually occurred (McGuigan, 1970). And third, Miller and Baron (1973) have suggested that observations of oral EMG (and cardiac) activity may not differentiate the cognitive elaboration of the advocacy from the covert rehearsal of the arguments constituting the persuasive message. One of our aims was to determine the *existence* of cognitive-response processes in persuasion; hence, we obtained recordings of the electrophysiological measures while subjects were *anticipating* highly involving and counterattitudinal advocacies as well as while subjects were processing them. A good deal of evidence documenting that anticipating a discrepant and involving message evokes issue-relevant cognitive responding has accrued using the thought-listing procedure (Cacioppo, Petty, & Snyder, 1979; Petty & Cacioppo, 1977; Cialdini & Petty, in press).

Experiment 1

The aim of Experiment 1 was primarily methodological. We sought to test the utility of an electrophysiological assessment of cognitive response. To do so, we forewarned individuals that they would hear a (counterattitudinal) message and we asked them to "collect their thoughts" about the issue. Immediately preceding our forewarning, we collected basal measures of physiological response, and we continued to record these meaures while the individuals were anticipating the discrepant message (i.e., when they presumably were following instructions and were generating cognitive responses concerning the advocacy). We expected oral EMG and cardiac activity, but not general somatic activity, to be heightened during the "collect thoughts" interval, compared to basal levels.

Confirmation of these expectations would illustrate the applicability and utility of electrophysiological techniques for studying cognitive response in persuasion. Obtaining these results would *not*, of course, provide evidence that cognitive responses in persuasion are generated naturally, since we asked the individuals to collect their thoughts following the forewarning. A second experiment addressed this latter issue.

A second aim of Experiment 1 was to explore the effects of affect- and nonaffect-laden cognitive responding on electrophysiological response patterns. Neither oral EMG (cf. Garrity, 1977; McGuigan, 1978) nor cardiac activity (Cacioppo & Sandman, 1978; Harris, Katkin, Lick, & Habberfield, 1976) appear to

distinguish the types of affective response evoked, though differentiation between affect- and non-affect-laden thought sequences using heart rate has been observed (Schwartz, 1971). The discrepancy of the impending advocacy was manipulated in Experiment 1 to obtain gradations of affect-laden processing to assess its physiological effects. Additionally, the thought-listing procedure was used to explore the relationships among the cognitive (e.g., favorable thoughts, counterarguments, neutral/irrelevant thoughts) and electrophysiological (e.g., oral EMG, heart rate) responses.

Method

Subjects and Design

Three replications of eight male undergraduates were conducted. Subjects were tested individually, with assignment to replication determined randomly. A $3 \times 3 \times 2 \times 2$ mixed design was employed in which the three replications served as a between-subjects factor, and levels of communication discrepancy (low, moderate, and high), two different topics within each level of discrepancy, and interval during which electrophysiological measures were recorded (prewarning baseline and postwarning "collect thoughts" interval) served as within-subjects factors. (The interval factor was relevant only to the analyses of the electrophysiological measures.)

Materials

A separate audiotape was prepared for each of the three replications. Each tape contained the experimental instructions and six announcements regarding the source of, topic of, and position to be advanced in an upcoming message. (In fact, however, the messages were never presented.) Each level of discrepancy for each topic appeared in one replication. The tapes (i.e., replications) differed in the order of the topics, which was determined randomly for each replication. The experimental stimuli are displayed in Table 1.

Procedure

When subjects arrived at the laboratory, they were placed in a sound-attenuated room and were seated in a comfortable chair. Electrodes were attached for measuring oral (orbicularis oris—lips; digastricus—chin; platysma—throat) EMG activity, nonspeech (trapezius—back) EMG activity, heart rate, breathing rate, and cephalic pulse amplitude. A 5-minute adaptation period preceded the experimental trials.

Subjects were told that in about 40 minutes they would hear several different messages having direct consequences for undergraduates, and that before the presentation of the messages we should like to obtain their comments on and evaluations of the position (i.e., advocacy) to be advanced in each message. The subjects were asked to sit quietly and collect their thoughts for the minute following each announcement; then, at the experimenter's signal, subjects were asked to list everything about which they had been thinking (subjects were given 3 minutes to do so), to rate their agreement with the upcoming advocacy, and to complete several ancillary measures (i.e., felt effort, involvement, distraction, and responsibility). The nature of these forms is described in detail in Petty and Cacioppo (1977). The forewarning, collect thoughts, and thought-listing intervals were repeated six times to cover six different topics, each separated by a variable intertrial interval (ITI) ranging from 90 to 120 seconds. Each ITI was initiated when the experimenter requested the subjects to "please sit quietly for the next minute or so." The final 60 seconds of each ITI served as the baseline measure for the subsequent 60-second "collect thoughts" interval.

Chin muscle activity. Two Grass E5S cup electrodes filled with Grass EKG Sol were placed on the midline of the chin; the first was placed 1.8 cm above the point of the chin and the second was

Table 1
Experimental Stimuli

Source	Topic	Discrepant position		
		Low	Moderate	High
University board of regents	Increasing student tuition by . . .	$5.00	$40.50	$90.00
State medical association	Increasing drinking age to . . .	19 years	21 years	25 years
Member of state legislature	Increasing gasoline sales tax by . . .	2¢	5¢	10¢
University faculty committee	Graduate students teaching all . . .	freshmen	freshmen & sophomores	freshmen & juniors
President of the university	Extending finals by . . .	2 days	5 days	9 days
County municipal courts	Increasing traffic fines by . . .	25%	100%	300%

placed 1.8 cm below the point of the chin. Following Cacioppo et al. (1978), chin muscle activity was calculated for the 60-sec intervals of interest using the following formula:

$$\text{Activity index} = \sum_{i=1}^{n} \frac{(l_i \times s_i)}{l_i}$$

where s_i is the scale value of a particular amplitude of EMG activity (larger amplitude EMG activity was assigned larger scale values; scale values ranged from 0, for deflections 2 mm or less, to 5, for deflections exceeding 2 cm); l_i is the total horizontal length (i.e., time) of a particular scale value of EMG activity measured in millimeters; and n is the number of distinct instances of a particular amplitude of EMG activity. The data were quantified in this manner for the 60-sec baseline and collect thoughts intervals.

Lip muscle activity. Two Grass E5S cup electrodes filled with Grass EKG Sol were placed 1 mm below the bottom lip; each was placed .5 cm in from the ends of the mouth. Lip muscle activity was calculated using the activity index described above.

Throat muscle activity. Two Grass E5S cup electrodes filled with Grass EKG Sol were placed off midline of throat; the first was placed approximately 1.0 cm to the right of and level with the midpoint of the throat, and the second was placed approximately 1.0 cm to the left and 1.0 cm above the midpoint. Throat muscle activity was calculated using the activity index described above.

Back muscle activity. Subjects were asked to place their fingertips on their collarbone while the electrodes were secured. Two Grass E5S cup electrodes filled with Grass EKG Sol were placed over the trapezius muscle group. The first was placed 4.0 cm outward from the midline of the line passing between the first thoracic and seventh cervical vertebrae, and the second was placed halfway between the spine and the head of humerus (near the point of the shoulder). Back muscle activity was calculated using the activity index described above.

Heart rate. Grass E5S cup electrodes filled with EKG Sol were placed over the lower left rib cage and the right collar bone. The signal was amplified by a Grass wide-band AC preamplifier. Heart rate was calculated by counting the number of beats that occurred in the 1-minute intervals of interest.

Breathing rate. The respirometer was a sliding piston (consisting of a photocell and a small light) mounted on an elastic band and placed around the subject's chest (Shmavonian, Miller, & Cohen, 1968). Breathing rate was calculated by counting the number of cycles (to the nearest half cycle) occurring in the intervals of interest.

Cephalic pulse amplitude.[3] A photoplethysmograph was placed over the supraorbital notch (above the eyebrow), providing a relative measure of blood volume in the supraorbital artery. The photoplethysmograph was comprised of three light-emitting diodes (LED), radiating 100 μW of light output at 660 nanometers. The LEDs were spaced at 120° on a radius of 6.3 mm around a high-speed photoconductor, and resistance changes were recorded with a Grass oscillograph. Cephalic pulse amplitude was calculated for the intervals of interest using the following formula:

$$\text{Amplitude index} = \frac{\text{pulse amplitude in mm} \times \text{total mV/cm calibration}}{\text{standard sensitivity in mV/cm}}$$

where total mV/cm was the sensitivity of the preamp settings for a given subject, and the standard sensitivity was 1 mV/cm.

Data reduction. Physiological processes were monitored during the 1 minute preceding and following each of the six forewarnings. Persons scoring the electrophysiological data were unaware of the experimental hypotheses and of the treatments with which the data were associated.

The Grass Model 7 polygraph used in the experiment was equipped with three preamplifiers capable of measuring electromyographic activity. Since four EMG measures were of interest in the experiment, a random procedure was used to determine which three of the four EMG measures would be recorded for each subject within a replication. However, all electrode placements were prepared on each subject, and the subject was unaware of the dummy electrode placements. Difference scores relative to the prewarning levels were computed for each electrophysiological measure.

The subjects classified their cognitive response in a manner described by Petty and Cacioppo (1977): After listing their thoughts, subjects were instructed to place a plus (+) next to those thoughts that were in favor of the advocacy, a minus (−) next to those thoughts opposed to the advocacy, and a zero (0) next to those thoughts that were either neutral toward or irrelevant to the advocacy. Frequency counts served as measures of cognitive response.

---

[3] A photoplethysmograph was placed over the supraorbital artery just above the eye. The amount of light reflected back onto the photoplethysmograph is inversely proportional to the amount of blood between the photoplethysmograph and the supraorbital notch (over which it was placed). Brain blood perfusion was assumed to be indirectly related to the relative volume changes of the pulse wave that traveled to the brain via the supraorbital artery. Pulse rate was not considered to be important because of the venous return mechanism, which drains the blood from a given body area at approximately the same rate as the blood is delivered to that area (Wallace & Wallace, 1968). The validity of this measure, however, has not yet been firmly established. The plethysmograph used in these studies was developed by Robert Isenhart. More information about the plethysmograph is provided in Sandman, McCanne, Kaiser, and Diamond (1977).

Results and Discussion

Measurements of the seven electrophysiological dependent measures (chin, lip, throat, and back EMG activity, heart rate, breathing rate, and cephalic pulse amplitude) were obtained the minute preceding and following each forewarning. Parametric analyses of the questionnaire data were conducted while nonparametric analyses of the electrophysiological data were conducted, since the latter data were not distributed normally (cf. Schwartz, Fair, Salt, Mandel, & Klerman, 1976a, 1976b).

Cognitive Response, Agreement, and Ancillary Measures

We expected topic-relevant cognitive responding and agreement to be affected in a particular manner by discrepancy, but we had no particular expectations regarding neutral thought production and the ancillary measures. Hence, we set the experimentwise error rate at .10 and distributed this protection unequally across the tests (i.e., .05 for counterargumentation, favorable thought production, and agreement; .05 for the remaining five measures). We used Bonferroni-adjusted critical values for all tests, conducted two-tailed tests throughout, allocated 99% of the alpha associated with the contrasts for the first set of variables into the tail corresponding to the predicted direction of the effect (saving 1% as acknowledgment that opposite rather than predicted results sometimes obtain), and allocated 50% of the alpha associated with the tests for the second set of variables into each tail.[4] Further, the Huynh and Feldt (1970) test for the homogeneity of treatment–difference variance (HOTDV) was conducted for each dependent measure (cf. Harris, 1975, pp. 125–127).[5] These tests revealed a violation of the HOTDV assumption only for the measure of neutral thoughts, $\chi^2(2) = 8.11$. Neutral thought production was unaffected by discrepancy even with a positively biased F ratio.

The means for all cognitive response, agreement, and ancillary measures are summarized in Table 2. The analyses indicated that increasing the discrepancy between the subjects' initial positions and the advocated position led to more anticipatory counterargumentation, $F(2, 42) = 8.80$, $p < .01$; production of fewer anticipatory favorable thoughts, $F(2, 42) = 3.70$, $p < .017$; and greater felt effort in preparing cognitively for the message, $F(2, 42) = 5.33$, $p < .01$. One additional test approached significance: Agreement tended to decrease as discrepancy increased, $F(2, 42) = 3.57$, $p < .019$. No other effect or interaction was statistically significant. This pattern of results is similar to that obtained in prior research on cognitive response and agreement as a function of communication discrepancy (Brock, 1967; Cacioppo, 1977).

Table 2
Mean Responses to Thought Listing and Questionnaire Measures for Low, Moderate, and High Levels of Communication Discrepancy: Experiment 1

Measure	Mean		
	Low	Moderate	High
Primary			
Agreement	4.08	3.96	2.67
Counterarguments	2.04	2.73	3.40
Favorable thoughts	1.33	1.00	.63
Neutral thoughts	1.04	.77	.58
Ancillary			
Effort	7.35	8.42	8.42
Involvement	7.50	8.60	8.56
Distraction	3.98	3.40	2.90
Responsibility	6.81	7.96	7.42

Note. Entries for cognitive response measures indicate the mean frequency obtained in thought listings. Entries for questionnaire items are mean response to 11-point scale items in which higher numbers indicated more agreement, effort, involvement, distraction, and responsibility. Twenty-four subjects received two forewarnings at each level of discrepancy.

Electrophysiological Measures

We hypothesized that the activity of the oral muscles and heart rate would increase

[4] Required for statistical significance by these adjustments were ps $< .017$ for counterarguing, favorable thoughts, and agreement, and ps $< .01$ for the remaining measures.

[5] We are indebted to Richard Harris for providing a computer program with which to test the HOTDV assumption.

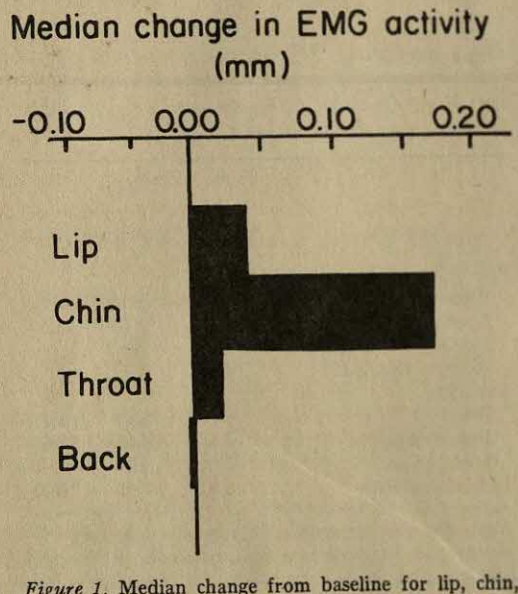


Figure 1. Median change from baseline for lip, chin, throat, and back electromyographic (EMG) activity following a forewarning about an impending counterattitudinal advocacy.

during the collect thoughts interval (relative to basal levels), since we expected cognitive response processes to be reflected electrophysiologically. Strong support for the hypothesis was obtained (see Figure 1). Oral EMG activity was elevated significantly after forewarnings of involving counterattitudinal communications (by the Wilcoxon Test: lip, $p < .001$; chin, $p < .001$; throat, $p < .10$). Also evident in Figure 1, general somatic activity, as measured by back EMG activity, was not altered by the anticipation of the counterattitudinal message ($p > .25$). The Wilcoxon Test for changes from baseline revealed that heart rate ($Mdn = 2.00$ bpm, $p < .02$) and breathing rate ($Mdn = 1.0$ cycle/min, $p < .01$) increased following the forewarning as well, whereas cephalic pulse amplitude was left unchanged ($Mdn = 0.00$). No other comparisons were significant statistically.[6]

Correlational Analyses

The affective intensity of the cognitive responses, which was varied by increasing communication discrepancy, did not affect the electrophysiological activity monitored in this study. Canonical correlations between the cognitive responses (i.e., counterarguments, favorable thoughts, neutral/irrelevant thoughts) and the relevant electrophysiological measures (i.e., lip, chin, throat, and cardiac activity) were computed within each level of discrepancy to determine whether the predicted relationship between cognitive and electrophysiological activity existed and to explore whether any association existed between the affective nature of the covert processing and electrophysiological activity. The correlations were respectable ($r = .47$ for low discrepancy conditions; $r = .44$ for moderate discrepancy conditions; and $r = .64$ for high discrepancy conditions). When a canonical correlation was calculated, collapsing across the levels of discrepancy, a coefficient of .42 was obtained.

Furthermore, the electrophysiological *specificity* obtained in this research is in striking contrast to the massive and diffuse arousal associated with the fight-or-flight reaction of extreme emotional states (Cannon, 1927) and misattribution phenomena (Schachter, 1964; see also, Rhodewalt & Comer, 1979).

The calculation of within-cells correlations among cognitive responses and agreement revealed that anticipatory counterargumentation correlated negatively with agreement ($r = -.67$, $p < .01$), favorable thoughts ($r = -.67$, $p < .01$), and neutral thoughts ($r = -.45$, $p < .01$). Favorable thoughts correlated positively with agreement ($r = .72$, $p < .01$).[7]

---

[6] Subjects were asked to count to five aloud and to move and tense slightly in their chair before the completion of the experiment to assess the validity of the EMG electrode placements. Trapezius EMG activity increased during body tensing and movements. Lip and chin EMG activity increased during overt oral behavior (counting aloud), whereas the throat EMG placement proved to be a relatively insensitive measure.

[7] Analysis of covariance procedures were employed to explore some of the possible causal sequences of cognitive responding and attitude change. It should be noted, however, that these procedures do not *prove* that a particular causal model is operating. These analyses were conducted here to assess if the reduced agreement found with increasing discrepancy possibly resulted from counterargumentation. Pre-

Within-cell correlations further revealed that unfavorable thoughts were positively correlated with ratings of effort ($r = .42$, $p < .05$), involvement ($r = .43$, $p < .05$), and responsibility ($r = .29$, $p < .05$), and correlated negatively with ratings of distraction ($r = -.29$, $p < .05$). None of the electrophysiological measures was correlated significantly with any ancillary measure.

In sum, electrophysiological and thought-listing measures indicated that topic-relevant cognitive responding accompanied the anticipation of an involving counterattitudinal communication. Inspection of Figure 1 supports the notion that cognitive response processes in persuasion can be measured concurrently and without the subject's doing anything overt in particular (e.g., listing thoughts). Of course, no evidence was provided in Experiment 1 concerning the natural existence or elicitation of cognitive response in persuasion, since the subjects were aware that they were to list their thoughts; in fact, they had been instructed to collect their thoughts following each forewarning. We were also unable to distinguish electrophysiologically the affective natures of the recipients' covert preparations for the advocacy. We conducted a second experiment to address these issues.

Experiment 2

The first issue, that of providing electrophysiological evidence concerning the cognitive responses elicited naturally in persuasion, was addressed easily in the design of Experiment 2. Rather than asking the subjects to collect or list thoughts, we simply monitored the electrophysiological activity displayed during the anticipation and presentation of a single advocacy. Subjects had no notion that they would subsequently be asked to list their thoughts.

The second issue, developing an electrophysiological measure capable of differentiating the affective nature of cognitive responding (should it exist in persuasion), proved to be more difficult. Previous studies of attitudes and bodily reactions have employed three basic research strategies: Procedures have been employed (a) to tap the physiological processes indicating covert information processing (e.g., Experiment 1; Cacioppo, 1979); (b) to measure an evaluative reaction by monitoring a classically conditioned physiological response (e.g., Tognacci & Cook, 1975); and (c) to assess the naturally occurring physiological indicators of affective states (e.g., Cooper, 1959; Hess, 1965). Studies of attitudes applying the third research strategy have often used measures of pupillary response or electrodermal activity (galvanic skin responses). These studies have provided some evidence that attitudes, if extreme, may be measurable. Even in these instances, however, the polarity of the attitude (i.e., positive or negative) has not been distinguishable (Cacioppo & Sandman, in press; Mueller, 1970).

Recent work on the neuromuscular substrates of emotion and depression offered us a potential solution. Darwin (1965/1872) first documented the specificity and reliability of facial muscle patterning in the expression of emotions (cf. Cacioppo & Petty, in press-a). More recently, Schwartz and his colleagues (Schwartz, 1975; Schwartz et al., 1976a, 1976b) have found that generating

---

...iously we have found that topic-relevant thinking ...ppears to mediate the subsequent agreement with ...e advocacy (Cacioppo & Petty, 1979; Petty & ...acioppo, 1977), rather than vice versa. We expected ...e same results from the present study. (See Insko ...t al., 1974, for an extended rationale for the use of ...nalyses of covariance [ANCOVA] with attitude and ...gnitive response measures.)

A series of analyses of covariance were performed ...h which agreement served as the criterion and each ...gnitive response (one per analysis) served as the ...variate. A second series of covariance analyses was ...erformed in which agreement served as the covariate ...d each cognitive response served as the criterion. ...e results of these analyses revealed only that the ...variate, counterarguments, reduced the overall F ...tio for the discrepancy factor for agreement: the ...F statistic for agreement was reduced from a signif... ...ast 3.57 to a nonsignificant 1.70 ($df = 2, 41$). The ...alysis of covariance with counterarguments as the ...iterion and agreement as the covariate did not ...minate the significant F ratio for counterarguments, ...COVA $F(2, 41) = 5.53$. These analyses are consistent ...th the notion that increasing discrepancy increased ...unterargumentation, which then reduced agree-...ent.

imagery or attempting to experience through fantasy the emotional states of happiness, sadness, and anger leads to distinctive patterns of EMG activation of the face. Moreover, these patterns go unnoticed by subjects as well as observers (see also Izard, 1971). We reasoned that by monitoring these facial (i.e., corrugator, zygomatic, depressor anguli-oris, and mentalis) muscles during the anticipation and presentation of advocacies, we would be able to distinguish favorable from unfavorable (i.e., counterargument) cognitive responses emitted by a recipient. In Experiment 1 and in previous research, we have found oral EMG to distinguish the extent rather than the affectivity of covert processing (e.g., Cacioppo & Petty, in press-b, in press-c). Hence, we considered the measure of mentalis EMG activity, which taps the electrical activity of the muscle fibers between and including the lower lip and chin, as a measure of the extent rather than emotionality of processing (see also McGuigan, 1978). Again, heart rate was recorded.

Method

Subjects and Design

Sixty male undergraduates were led to believe they were evaluating the sound quality of taped radio editorials that had been produced by the students in a sound-engineering course. Electrodes were placed on each subject's body, and subjects were tested individually in a darkened, sound-attenuated room "to reduce external distractions from the task." Forty-eight subjects were forewarned about and heard either a proattitudinal or a counterattitudinal advocacy on one of two topics (alcoholic beverages or visitation hours). Twelve additional subjects were forewarned only that they would hear a taped communication, and they heard a message about an obscure news event. Subjects in this group served in an external control (neutral advocacy) condition. The assignment of subjects to condition again was determined randomly.[8]

Materials

The topics of alcoholic beverages and visitation hours were selected because initial pilot testing revealed existing university regulations regarding them to be highly involving and counterattitudinal. Forewarnings and messages were constructed that advocated the adoption of either stricter (counterattitudinal) or more lenient (proattitudinal) regulations regarding these issues. The neutral message concerned a small archeological find and was obtained from a past issue of a national news magazine.

Procedure

When subjects arrived at the laboratory, they were told that their task was to evaluate the sound quality of a taped radio editorial, that electrodes would be attached, and that during the study we would be recording the involuntary bodily responses that accompany listening to a communication.[9] Subjects were instructed to refrain from unnecessary movements, to breathe normally, and to keep their eyes closed throughout the study. After adapting to the laboratory, subjects again heard these instructions and were told that the study would begin shortly. At this point, a computer-controlled procedure—which involved (a) a 60-sec prewarning (baseline) interval, (b) a 15-sec forewarning, (c) a 60-sec postwarning–premessage interval, and (d) a 120-sec message—was initiated.

After listening to the tape, the subjects read the following:

> Because your own opinion about the position advocated on the tape may influence the way you rate the quality of the tape, we would like to obtain a measure of how you feel about the views proposed by the speaker on each scale below.

The subjects responded to four 9-point semantic differentials; their responses were summed to obtain a measure of their attitude toward the advocacy. In the same manner as in Experiment 1, subjects were instructed to list everything about which they had thought during the message (subjects were given 3 minutes). Afterwards, subjects rated their listed thoughts as favorable (+), unfavorable (−), or neutral/irrelevant (0) toward the message. Subjects then rated on 11-point Likert-type scales their felt involvement, effort, and distraction, the personal relevance of the message, the sound quality of the tape, and the speaker's rate of delivery and enthusiasm.

Heart rate. Grass E5S cup electrodes filled with Grass EC3 paste were placed over the lower left rib and the right collar bone. The signal was amplified

---

[8] An additional factor included in the design was whether subjects were informed that the communication had implications locally or not. Manipulation checks revealed that our manipulation of this factor failed here, so this factor is not discussed further.

[9] Electroencephalographic measures were obtained also, the results of which are to be reported elsewhere, since they were collected to address a different issue. Suffice it to say that enough electrodes were attached to subjects to make the cover story concerning the measurement of involuntary processes seem entirely plausible to them.

by a Narco Biosystem Physiograph AC preamplifier. The output was displayed on a Narco Biosystem Physiograph 6 and was transmitted on-line to a PDP-8I laboratory computer for analysis.

Facial muscle activity. Grass E5S cup electrodes filled with Grass EC3 paste were placed adjacent to each other in pairs with interelectrode resistance reduced to less than 10,000 ohms. The four muscles over which the pairs of electrodes were placed were the corrugator (just above the eyebrow), zygomatic (upper cheek), depressor anguli oris (lower cheek), and mentalis (between lip and chin) on the left side of the face (cf. Schwartz et al., 1976a). Since surface electrodes were used, recordings of EMG activity were obtained from these and surrounding muscle groups.

Each EMG measure was amplified by a Narco Biosystem Physiograph AC preamplifier, individually rectified and summed by an EMG integrator with a time constant of .2 sec. The average integrated EMG was displayed on the physiograph with a full-scale pen deflection of 40 mm (1 mm = 7.5 μV). The integrated EMG also was transmitted on-line to the PDP-8I laboratory computer, sampled 10 times per second, and recorded. The computer was programmed to eliminate from its recordings any obvious movement artifact or overt (e.g., visually detectable) facial expression.

Data reduction. The cognitive data were scored in the same manner as in Experiment 1. Physiological responses were monitored during the minute preceding the forewarning through the completion of the 120-sec message. The data for each measure were averaged for each subject and each interval. Difference scores relative to the prewarning (basal) levels were calculated for each measure and interval, and nonparametric analyses were employed, since the data were not distributed normally. The presentation here of the results for facial EMG activity is similar in format to the presentation by Schwartz et al. (1976a, 1976b) to facilitate comparisons.

Results and Discussion

Our purposes in this study were to determine if cognitive responses were generated naturally in persuasion settings and, if so, whether or not the *affective nature* of these responses could be assessed electrophysiologically. A multivariate analysis of variance of the 11 questionnaire measures was conducted first to determine the general effects of the experimental factors. As expected, the effect of position was highly significant, based on Wilks' lambda, $F(11, 30) = 3.63$, $p < .01$, and the effect of topic was not significant, $F(11, 30) = 1.74$, $p > .11$. Hence, all analyses reported below are collapsed across the topic factor.

Table 3
Mean Cognitive and Attitudinal Responses as a Function of the Affectivity of the Advocacy: Experiment 2

Measure	Pro-attitudinal	Counter-attitudinal	Neutral
Evaluation of the taped advocacy	27.04	16.75[a]	25.00
Counterarguments	2.00	3.54	2.17
Favorable thoughts	2.34[a]	1.13	1.25
Neutral/irrelevant thoughts	2.04	1.54[a]	3.33
Total thoughts	6.38	6.23	6.74

[a] The mean differs from the corresponding neutral mean at the .05 level by Dunnett's test.

Cognitive Response, Attitude, and Ancillary Measures

Subjects anticipated and listened to a proattitudinal advocacy, a counterattitudinal advocacy, or a neutral communication, and rated their evaluation of and thoughts about the taped presentation. As is evident from an inspection of the means in Table 3 for these measures, the proattitudinal advocacy was evaluated more positively, $F(1, 40) = 30.37$, $p < .001$, and elicited more favorable thoughts, $F(1, 40) = 6.58$, $p < .02$, and fewer counterarguments, $F(1, 40) = 6.50$, $p < .02$, than did the counterattitudinal advocacy.

The Dunnett test for comparisons involving an external control mean (Kirk, 1968) was employed to determine the relative effects of the proattitudinal and counterattitudinal communications relative to the neutral communication on cognitive and attitudinal responding. These comparisons revealed that the neutral communication differed from the counterattitudinal communication in evaluation and in the number of neutral/irrelevant thoughts elicited (ps $< .05$); in addition, these communications differed marginally in the number of counterarguments elicited ($p < .10$). On the other hand, the neutral and proattitudinal communications were evaluated and were thought about similarly (see Table 3). Evidently the subjects enjoyed hearing our "neutral" message about an archeological dig.

Analyses of the ancillary measures and ratings of tape quality failed to produce any significant effects.

Electrophysiological Measures

All tests of the significance of changes from prewarning baselines for the electrophysiological measures were conducted using two-tailed Mann-Whitney tests.

Does cognitive responding occur naturally? The analyses of covert oral (mentalis) EMG activity, which was the most sensitive measure of covert processing in Experiment 1, indicated that it was elevated during the postwarning–premessage interval for the *counterattitudinal* condition ($Mdn = .24$ μV, $p < .03$). This finding replicates that of Experiment 1, but here was obtained without explicit requests for subjects to collect their thoughts. Interestingly, oral EMG activity was *not* altered significantly during this interval in the proattitudinal or neutral conditions ($ps > .15$). Still, the *presentation* of a message, whether it was proattitudinal, counterattitudinal, or neutral, led to increased oral EMG activity ($Mdn = .43$ μV, 1.21 μV, and 1.26 μV, respectively, $ps < .03$). Also, as in Experiment 1, the affective nature of the covert processing was not distinguishable by oral EMG; the mentalis activity did not differ as a function of position. The analyses of heart rate indicated that this measure was less sensitive than oral EMG activity, as differences from basal levels were evident only during the presentation of the communications (counterattitudinal $Mdn = .73$ bpm; proattitudinal $Mdn = .36$ bpm; neutral $Mdn = 2.68$ bpm; $ps < .05$). Finally, changes in heart rate during the communication were greater for the neutral than for the counterattitudinal ($p < .03$) and the proattitudinal ($p < .04$) conditions. The cause of this difference is not immediately apparent, but the news story about the archeological find may have required more thought to comprehend or elaborate, producing greater accelerations of heart rate.

Is the emotional tone of this cognitive activity distinguishable? We next sought to determine if the affective nature of the cognitive responses was distinguishable by the pattern of *facial* EMG activity. Schwartz and his associates have demonstrated that pleasant states (e.g., happiness) lead to less corrugator and more zygomatic and depressor EMG activity than do unpleasant states (e.g., sadness, anger), with corrugator EMG activity providing the most discriminating measure (Schwartz, Fair, Salt, Mandel, Mieske, & Klerman, 1978). Figure 2 displays the median change from baseline for these measures as a function of position and interval in the present study.

Several findings are evident immediately upon inspection of Figure 2. First, only corrugator activity was altered during the announcement of the *forewarning* (upper panel), with greater and equal activation relative to baseline appearing in all conditions ($ps < .01$). Less evident in the upper panel of Figure 2 is the marginally significant tendency for zygomatic activity to discriminate between the counterattitudinal and neutral forewarnings ($p < .06$).

During the *postwarning–premessage* interval (middle panel of Figure 2), corrugator activity remained elevated from basal levels ($ps < .02$), though a significant decrease from the forewarning level was displayed in the proattitudinal condition ($p < .01$). Between-group comparisons yielded a nonsignificant difference in corrugator activity between the proattitudinal and counterattitudinal conditions ($p < .11$), with the direction of the difference that which would be expected from Schwartz and his colleagues' research. That is, corrugator activity was higher when anticipating the counter than proattitudinal advocacy. Finally, the activity in the zygomatic muscle region in the neutral condition was enhanced relative to basal levels ($p < .03$) and distinguished this condition from the group anticipating a counterattitudinal advocacy ($p < .01$). This too is consistent with the previous studies of emotional fantasy and imagery (cf. Schwartz, 1975). No other differences were statistically significant.

The *presentation* of the messages (lower panel of Figure 2) resulted in elevated corrugator EMG activity relative to basal ($ps < .01$) and postwarning–premessage levels (ps

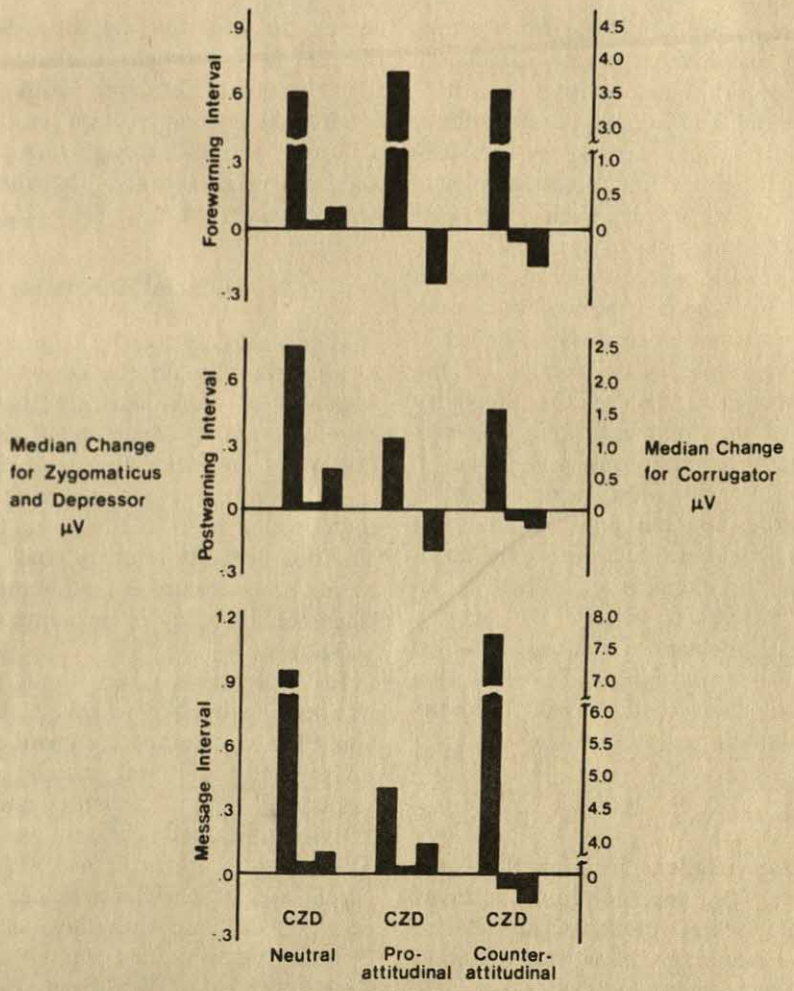


Figure 2. Median change from baseline for corrugatory (C), zygomaticus (Z), and depressor (D) electromyographic activity during the forewarning (top panel), postwarning-premessage (middle panel), and message (bottom panel) intervals. (The data are displayed separately for subjects who anticipated and heard the neutral [$n = 12$], proattitudinal [$n = 24$], and counterattitudinal [$n = 24$] advocacies.)

< .01); in addition, it was elevated marginally during this interval compared to forewarning levels (ps < .06). The zygomatic activity continued to differentiate the affectivity of the covert processing: Zygomatic EMG activity during the counterattitudinal message was lower than that displayed during baseline ($p < .05$), the proattitudinal message ($p < .01$), and the neutral message ($p <$ .05). Furthermore, the zygomatic activity during the proattitudinal message was significantly greater than its basal level and marginally greater than its forewarning ($p <$ .08) and postwarning–premessage levels ($p <$ .08). Similarly, depressor EMG activity was enhanced marginally during the proattitudinal message compared to its postwarning–premessage level ($p < .06$) and compared to the counterattitudinal message ($p < .10$), effects also similar to those found by Schwartz et al. (1976a, 1976b). No other comparisons approached statistical significance.

In sum, oral EMG activity increased from baseline after the forewarning of an impending and involving *counterattitudinal* communication, even though we did not request subjects to collect their thoughts and subjects were unaware that they would be asked to list

their thoughts, heard only one forewarning and message, and had no reason to suspect the purpose of our measurements. Furthermore, anticipating and hearing proattitudinal, counterattitudinal, and neutral communications led to distinctive and predictable patterns of facial EMG activity during the stimulus sequence. Specifically, less corrugator and more zygomatic activity was observed in the proattitudinal and neutral conditions than in the counterattitudinal condition. The activity of the depressor muscle region, for the most part, was unaffected by the affectivity or interval of the communication sequence. The differences that *were* observed, however, were as predicted: Depressor activity tended to be greater during the proattitudinal and neutral communications than during the counterattitudinal communication. Finally, although the patterns of facial EMG activity were similar for the neutral and proattitudinal conditions, there was a high degree of similarity between these conditions in the cognitive and evaluative responses as well.

Correlational Analyses

Canonical correlations between the cognitive responses (i.e., counterarguments, favorable thoughts, neutral thoughts) and electrophysiological scores (i.e., heart rate, mentalis, corrugator, zygomatic, and depression EMG activity) were calculated, once using the physiological responses for the postwarning–premessage interval, and once using the responses for the message interval. Since thought listings were obtained immediately following the message interval, the canonical correlation between thoughts and the bodily responses from this interval should show the strongest association; but physiological responses from both intervals should correlate somewhat with the listed thoughts. Correlations were .30 for the postwarning–premessage interval and .46 for the message interval, $\chi^2(15) = 7.20$ and 16.20, respectively, *ns*. Hence, the covariation of cognitive and electrophysiological response was weak by these indices, but the latter index was stronger than the former, as expected.

Calculations of within-cell correlations among the cognitive response data revealed that, as in Experiment 1, counterargumentation correlated negatively with the attitude toward the communication ($r = -.30$, $p < .05$) and favorable thoughts ($r = -.41$, $p < .05$), whereas favorable thoughts correlated positively with attitude ($r = .45$, $p < .05$).

General Discussion

Theory and research in persuasion have focused recently on the covert idiosyncratic responses of individuals. In this article, we have reported two experiments describing the theory for and development of electrophysiological procedures for assessing this covert cognitive activity. Moreover, the evidence obtained suggests strongly that cognitive response processes are evoked naturally, at least when the advocacy is involving and counterattitudinal.

In Experiment 1, we found that subjects who had been asked to collect their thoughts about an upcoming discrepant message exhibited increased oral muscle, cardiac, and respiratory activity, whereas nonoral somatic activity remained constant and quiescent. These results are in accord with research in cardiovascular psychophysiology (e.g., Lacey et al., 1963) and with the literature on the electromyographic concomitants of thought (Jacobsen, 1973; McGuigan, 1978; Sokolov, 1972). Further, these results demonstrate that cognitive responses in persuasion settings are measurable concurrently and reliably, without asking the subject to respond overtly during the measurement and without the subject's awareness of the purpose or focus of the (electrophysiological) measurement instruments.

A second experiment was conducted to provide answers to two important questions: (a) Do subjects engage normally in active, covert processing when anticipating and hearing persuasive appeals? (b) If so, can the emotional tone of this cognitive activity be assessed electrophysiologically? Using an electromyographic technique to measure subtle responses of the facial muscles, we found that the forewarning of a counterattitudinal, but not a proattitudinal or neutral, communication

led to elevated oral EMG activity. Furthermore, the anticipation and presentation of counterattitudinal, proattitudinal, and neutral communications led to active though covert processing activity, the affective nature of which was revealed in the concomitant patterning of facial EMG activity.[10]

Cognitive Encoding or Elaboration?

It might be argued that the increased oral muscle activity indexed silent rehearsals rather than cognitive elaborations of the message arguments (cf. Miller & Baron, 1973). Listening to prose does cause a slight increase in oral muscle activity (McGuigan & Bailey, 1969a). Hence, comprehending and rehearsing the message arguments probably contributed to the elevated oral muscle activity exhibited *during the message presentation*. However, the *anticipation* of a counterattitudinal advocacy led to elevated speech muscle activity when there were no message arguments to rehearse. One might argue that subjects were rehearsing the forewarning and potential message arguments, stopping to generate counterarguments only when asked by the experimenter at the end of the study to "list everything about which you thought." Besides the absence of parsimony, this explanation fails to account for the subtle facial expressions indicating affect-laden cognitive responding that was displayed while subjects awaited the message. More persuasively, this explanation cannot account for the failure of subjects who were expecting a proattitudinal message to display significant increases in oral muscle activity during the postwarning–premessage interval.

Cognitive, Attitudinal, and Electrophysiological Response

According to the cognitive response analysis, a forewarning of an upcoming and involving discrepant communication elicits preparatory cognitive activity (e.g., anticipatory counterargumentation) in an effort to rally one's cognitive defenses (Cialdini et al., 1976; McGuire & Papageorgis, 1962; Petty & Cacioppo, 1977). Interestingly, there are no studies in the literature that report the effects of anticipating a proattitudinal message that collected thought listings. We believe that anticipating a counterattitudinal as compared to a proattitudinal message generally results in deeper processing (i.e., more extensive cognitive preparation—Cacioppo & Petty, 1979). This difference may be attributable to the relative importance in defending from attack one's attitudes and beliefs (e.g., avoiding cognitive inconsistency), or to the relative thought that was devoted previously to (or scripts developed for) proattitudinal rather than counterattitudinal positions. The present study does not consider which of these interpretations is most plausible.

It should be noted that the observation of greater oral EMG activity being obtained when anticipating a counterattitudinal rather than proattitudinal or neutral communication does *not* imply that oral EMG activity is a measure of "counterargumentation." Indeed, the results of both experiments suggest it is not. As mentioned above, we believe that the cognitive preparation for a counterattitudinal rather than a proattitudinal or neutral message elicits more extensive processing, which may result in the generation of counterarguments, favorable thoughts, *and/or* neutral thoughts. Accordingly, the activation of the oral muscles reflects this difference in the extent of cognitive elaboration rather than the affectivity of the processing. Evidence that oral EMG reflects this depth of processing has been found in a study in which subjects viewed a word and identified whether it was printed in uppercase letters (shallow processing) or was self-descriptive (deep processing —Cacioppo & Petty, in press-b, in press-c). We found that deeper processing resulted in elevated speech EMG activity.

Finally, a word might be said about the

---

[10] Love (1972) attempted to detect subtle changes in facial expression by videotaping the shoulders and face of subjects as they listened to an advocacy. Raters then scored the nonverbal cues emitted by these subjects. This measure proved to be insensitive to the experimental manipulations. The electrophysiological approach illustrated here has the advantage of being sensitive to subtle changes in responding by recipients.

sensitivity and specificity of the electrophysiological measures employed. The procedures we employed for scoring these responses eliminated obvious movement artifacts and facial expressions (the number of the edits did not differ across the experimental conditions); hence, we intentionally confined ourselves to the study of covert electrophysiological responses. Nevertheless, in this research, oral EMG activity has been a more sensitive measure of covert information processing than heart rate (see also McGuigan, 1978). Heart rate and speech muscle activity increased following the forewarning of an upcoming discrepant message, but the change in heart rate in Experiment 2 was not significant statistically; heart rate *did* increase significantly during the presentation of the communications, although again, the electromyographic measures proved more sensitive. Similarly, the various measures of facial EMG activity were not equally sensitive to affect-laden processing. Schwartz and his colleagues (1976a, 1976b, 1978) have found, as we here have found, that corrugator EMG activity best distinguishes subtle affective states. These instances of differing sensitivity, or response discordance, illustrate a point made by the Laceys (e.g., 1959, 1967) regarding the specificity of autonomic and somatic activation. Moreover, this remarkable specificity contrasts sharply with Cannon's (1927) theory of physiological arousal and emotion, and provides unique evidence for cognitive response processes in persuasion.

References

Baron, R. S., Baron, P. H., & Miller, N. The relationship between distraction and persuasion. *Psychological Bulletin*, 1973, *80*, 310–323.

Blatt, S. J. Patterns of cardiac arousal during complex mental activity. *Journal of Abnormal Social Psychology*, 1961, *63*, 272–282.

Blaylock, B. Some antecedents of directional fractionation: Effects on "intake-rejection," verbalization requirements, and threat of shock on heart rate and skin conductance. *Psychophysiology*, 1972, *9*, 40–52.

Brock, T. C. Communication discrepancy and intent to persuade as determinants of counterargument production. *Journal of Experimental Social Psychology*, 1967, *3*, 269–309.

Burnstein, E., & Vinokur, A. Persuasive argumentation and social comparison as determinants of attitude polarization. *Journal of Experimental Social Psychology*, 1977, *13*, 315–332.

Cacioppo, J. T. *Heart rate, cognitive response, and persuasion*. Unpublished doctoral dissertation, The Ohio State University, 1977.

Cacioppo, J. T. The effects of exogenous changes in heart rate on the facilitation of thought and resistance to persuasion. *Journal of Personality and Social Psychology*, 1979, *37*, 487–496.

Cacioppo, J. T., Glass, C. R., & Merluzzi, T. V. Self-statements and self-evaluations: A cognitive-response analysis of heterosocial anxiety. *Cognitive Therapy and Research*, 1979, *3*, 249–262.

Cacioppo, J. T., Harkins, S. G., & Petty, R. E. The nature of attitudes and cognitive responses and their relationships to behavior. In R. E. Petty, T. M. Ostrom, & T. C. Brock (Eds.), *Cognitive responses in persuasion: A text in attitude change*. Hillsdale, N.J.: Erlbaum, in press.

Cacioppo, J. T., & Petty, R. E. Effects of message repetition and position on cognitive response, recall, and persuasion. *Journal of Personality and Social Psychology*, 1979, *37*, 97–109.

Cacioppo, J. T., & Petty, R. E. Neuromuscular circuits in affect-laden information processing. *The Pavlovian Journal of Biological Science*, in press. (a)

Cacioppo, J. T., & Petty, R. E. Electromyographic specificity during covert information processing. *Psychophysiology*, in press. (b)

Cacioppo, J. T., & Petty, R. E. Lip and nonpreferred forearm EMG activity as a function of orienting task. *Biological Psychology*, in press. (c)

Cacioppo, J. T., Petty, R. E., & Snyder, C. W. Cognitive and affective response as a function of relative hemispheric involvement. *The International Journal of Neurosciences*, 1979, *9*, 81–89.

Cacioppo, J. T., & Sandman, C. A. Physiological differentiation of sensory and cognitive tasks as a function of warning, processing demands, and reported unpleasantness. *Biological Psychology*, 1978, *6*, 181–192.

Cacioppo, J. T., & Sandman, C. A. Psychophysiological functioning, cognitive responding, and attitudes. In R. E. Petty, T. M. Ostrom, & T. C. Brock (Eds.), *Cognitive responses in persuasion*. Hillsdale, N.J.: Erlbaum, in press.

Cacioppo, J. T., Sandman, C. A., & Walker, B. B. The effects of operant heart rate conditioning on cognitive elaboration and attitude change. *Psychophysiology*, 1978, *15*, 330–338.

Calder, B. J., Insko, C. A., & Yandell, B. The relation of cognitive and memorial processes to persuasion in a simulated jury trial. *Journal of Applied Social Psychology*, 1974, *4*, 62–93.

Cannon, W. B. The James–Lange theory of emotions: A critical examination and an alternative theory. *American Journal of Psychology*, 1927, *39*, 106–124.

Cialdini, R. B., Levy, A., Herman, C. P., Kozlowski, L. T., & Petty, R. E. Elastic shifts of opinion: Determinants of direction and durability. *Journal of Personality and Social Psychology*, 1976, *34*, 663–672.

Cialdini, R. B., & Petty, R. E. Anticipatory opinion effects. In R. E. Petty, T. M. Ostrom, & T. C. Brock (Eds.), *Cognitive responses in persuasion*. Hillsdale, N.J.: Erlbaum, in press.

Cook, T. D. Competence, counterarguing, and attitude change. *Journal of Personality*, 1969, *37*, 342–358.

Cooper, J. B. Emotion in prejudice. *Science*, 1959, *130*, 314–318.

Craik, F. I. M., & Tulving, E. Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General*, 1975, *104*, 268–294.

Cullen, D. M. *Attitude measurement by cognitive sampling*. Unpublished doctoral dissertation, Ohio State University, 1968.

Darwin, C. *The expression of the emotion in man and animals*. Chicago: University of Chicago Press, 1965. (Originally published, 1872.)

Eagly, A. H. Comprehensibility of persuasive arguments as a determinant of opinion change. *Journal of Personality and Social Psychology*, 1974, *29*, 758–773.

Edwards, D. C., & Alsip, J. E. Stimulus detection during periods of high and low heart rate. *Psychophysiology*, 1969, *5*, 431–434.

Elliott, R. The motivational significance of heart rate. In P. A. Obrist, A. H. Black, J. Brener, & L. V. DiCara (Eds.), *Cardiovascular psychophysiology*. Chicago: Aldine, 1974.

Festinger, L., & Maccoby, N. On resistance to persuasive communications. *Journal of Abnormal and Social Psychology*, 1964, *68*, 359–366.

Garrity, L. I. Electromyography: A review of the current status of subvocal speech research. *Memory & Cognition*, 1977, *5*, 615–622.

Gillig, P. M., & Greenwald, A. G. Is it time to lay the sleeper effect to rest? *Journal of Personality and Social Psychology*, 1974, *29*, 132–139.

Goldwater, B. C. Psychological significance of pupillary movements. *Psychological Bulletin*, 1972, *77*, 340–355.

Greenfield, N. S., & Sternbach, R. A. (Eds.). *Handbook of psychophysiology*. New York: Holt, Rinehart & Winston, Inc., 1972.

Greenwald, A. G. Cognitive learning, cognitive response to persuasion, and attitude change. In A. G. Greenwald, T. C. Brock, & T. M. Ostrom (Eds.), *Psychological foundation of attitudes*. New York: Academic Press, 1968.

Hardyck, C. D., & Petrinovich, L. F. Treatment of subvocal speech during reading. *Journal of Reading*, 1969, *12*, 361–368, 419–422.

Harris, R. J. *A primer of multivariate statistics*. New York: Academic Press, 1975.

Harris, V. A., Katkin, E. S., Lick, S. R., & Habberfield, T. Paced respiration as a technique for the modification of autonomic response to stress. *Psychophysiology*, 1976, *13*, 386–391.

Hess, E. H. Attitude and pupil size. *Scientific American*, 1965, *212*, 46–54.

Hovland, C. I., Lumsdaine, A. A., & Sheffield, F. D. *Experiments on mass communication*. Princeton, N.J.: Princeton University Press, 1949.

Huynh, H., & Feldt, L. S. Conditions under which mean square ratios in repeated measurements designs have exact F-distributions. *Journal of the American Statistical Association*, 1970, *65*, 1582–1589.

Insko, C. A., Turnbull, W., & Yandell, B. Facilitative and inhibiting effects of distraction on attitude change. *Sociometry*, 1974, *37*, 508–528.

Izard, C. E. *The face of emotion*. New York: Appleton-Century-Crofts, 1971.

Jacobsen, E. Electrophysiology of mental activities and introduction to the psychological process of thinking. In F. J. McGuigan, & R. A. Schoonover (Eds.), *The psychophysiology of thinking: Studies of covert processes*. New York: Academic Press, 1973.

Janis, I. L., & Mann, L. *Decision making: A psychological analysis of conflict, choice, and commitment*. New York: Free Press, 1977.

Kahneman, D. *Attention and effort*. Englewood Cliffs, N.J.: Prentice-Hall, 1973.

Keating, J. P., & Brock, T. C. Acceptance of persuasion and the inhibition of counterargumentation under various distraction tasks. *Journal of Experimental Social Psychology*, 1974, *10*, 301–309.

Kirk, R. E. *Experimental design: Procedures for the behavioral sciences*. Belmont, Calif.: Brooks/Cole, 1968.

Lacey, J. I. Psychophysiological approaches to the evaluation of psychotherapeutic process and outcome. In E. A. Rubinstein & M. B. Parloff (Eds.), *Research in psychotherapy*. Washington, D.C.: American Psychological Association, 1959.

Lacey, J. I. Somatic response patterning and stress: Some revisions of activation theory. In M. H. Appley & R. Trumbull (Eds.), *Psychological stress: Issues in research*. New York: Appleton, 1967.

Lacey, J. I., Kagan, J., Lacey, B., & Moss, H. A. The visceral level: Situational determinants and behavioral correlates of autonomic response patterns. In P. H. Knapp (Ed.), *Expression of the emotions in man*. New York: International Universities Press, 1963.

Langer, E., Blank, A., & Chanowitz, B. The mindlessness of ostensibly thoughtful action: The role of "placebic" information in interpersonal interaction. *Journal of Personality and Social Psychology*, 1978, *36*, 635–642.

Lorens, S. A., Jr., & Darrow, C. W. Eye movements, EEG, GSR, and EKG during mental multiplication. *Electroencephalography and Clinical Neurophysiology*, 1962, *14*, 739–746.

Love, R. *A videotape technique for the measurement*

of cognitive response to persuasion. Unpublished doctoral dissertation, Ohio State University, 1972.

McGuigan, F. J. *Subvocal speech during silent reading*. (Accession No. ED 015115). Bethesda, Md.: ERIC Document Reproduction Service, 1967.

McGuigan, F. J. Covert oral behavior during the silent performance of language tasks. *Psychological Bulletin*, 1970, *74*, 309–326.

McGuigan, F. J. Conditioning of covert behavior—some problems and some hopes. In F. J. McGuigan & D. B. Lumsden (Eds.), *Contemporary approaches to conditioning and learning*. Washington, D.C.: Winston, 1973. (a)

McGuigan, F. J. Electrical measurement of covert processes as an explication of "higher mental events." In F. J. McGuigan & R. A. Schoonover (Eds.), *The psychophysiology of thinking: Studies of covert processes*. New York: Academic Press, 1973. (b)

McGuigan, F. J. *Cognitive psychophysiology*. Englewood Cliffs, N.J.: Prentice-Hall, 1978.

McGuigan, F. J., & Bailey, S. C. Covert response patterns during the processing of language stimuli. *Interamerican Journal of Psychology*, 1969, *3*, 289–299. (a)

McGuigan, F. J., & Bailey, S. C. Longitudinal study of covert oral behavior during silent reading. *Perceptual and Motor Skills*, 1969, *28*, 170. (b)

McGuigan, F. J., Culver, V. I., & Kendler, T. S. Covert behavior as a direct electromyographic measure of mediating responses. *Conditional Reflex*, 1971, *6*, 145–152.

McGuigan, F. J., Keller, B., & Stanton, E. Covert language responses during silent reading. *Journal of Educational Psychology*, 1964, *55*, 339–343.

McGuigan, F. J., & Rodier, W. I., III. Effects of auditory stimulation on covert oral behavior during silent reading. *Journal of Experimental Psychology*, 1968, *76*, 649–655.

McGuigan, F. J., & Schoonover, R. A. (Eds.). *The psychophysiology of thinking: Studies of covert processes*. New York: Academic Press, 1973.

McGuire, W. J., & Papageorgis, D. Effectiveness of forewarning in developing resistance to persuasion. *Public Opinion Quarterly*, 1962, *26*, 24–34.

Miller, N., & Baron, R. S. On measuring counterarguing. *Journal for the Theory of Social Behavior*, 1973, *3*, 101–118.

Mueller, D. J. Physiological techniques of attitude measurement. In G. F. Summers (Ed.), *Attitude measurement*. Chicago: Rand McNally, 1970.

Obrist, P. A., Webb, R. A., Sutterer, J. R., & Howard, J. L. The cardiac–somatic relationship: Some reformulations. *Psychophysiology*, 1970, *6*, 569–587.

Petty, R. E. *A cognitive response analysis of the temporal persistance of attitude changes induced by persuasive communications*. Unpublished doctoral dissertation, Ohio State University, 1977.

Petty, R. E., & Cacioppo, J. T. Forewarning, cognitive responding, and resistance to persuasion. *Journal of Personality and Social Psychology*, 1977, *35*, 645–655.

Petty, R. E., & Cacioppo, J. T. Effects of forewarning of persuasive intent and involvement on cognitive responses and persuasion. *Personality and Social Psychology Bulletin*, 1979, *5*, 173–176. (a)

Petty, R. E., & Cacioppo, J. T. Issue involvement can increase or decrease persuasion by enhancing message-relevant cognitive responses. *Journal of Personality and Social Psychology*, 1979, *37*, 1915–1926. (b)

Petty, R. E., Ostrom, T. M., & Brock, T. C. (Eds.). *Cognitive responses in persuasion*. Hillsdale, N.J.: Erlbaum, in press.

Petty, R. E., Wells, G. L., & Brock, T. C. Distraction can enhance or reduce yielding to propaganda: Thought disruption versus effort justification. *Journal of Personality and Social Psychology*, 1976, *34*, 874–884.

Posner, M. I. Psychobiology of attention. In M. S. Gazzaniga & C. Blakemore (Eds.), *Handbook of psychobiology*. New York: Academic Press, 1975.

Rhodewalt, F., & Comer, R. Induced-compliance attitude change: Once more with feeling. *Journal of Experimental Social Psychology*, 1979, *15*, 35–47.

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology*, 1977, *35*, 677–688.

Romer, D. Distraction, counterarguing, and the internalization of attitude change. *European Journal of Social Psychology*, 1979, *9*, 1–17.

Sandman, C. A., McCanne, T. R., Kaiser, D. N., & Diamond, B. Heart rate and cardiac phase influences on visual perception. *Journal of Comparative and Physiological Psychology*, 1977, *91*, 189–202.

Schachter, S. The interaction of cognitive and physiological determinants of emotional state. In L. Berkowitz (Ed.), *Advances in experimental social psychology*. New York: Academic Press, 1964.

Schwartz, G. E. Cardiac responses to self-induced thoughts. *Psychophysiology*, 1971, *8*, 462–467.

Schwartz, G. E. Biofeedback, self-regulation, and the patterning of physiological processes. *American Scientist*, 1975, *63*, 314–324.

Schwartz, G. E., Fair, P. L., Salt, P., Mandel, M. R., & Klerman, G. L. Facial expression and imagery in depression: An electromyographic study. *Psychosomatic Medicine*, 1976, *38*, 337–347. (a)

Schwartz, G. E., Fair, P. L., Salt, P., Mandel, M. R., & Klerman, G. L. Facial muscle patterning to affective imagery in depressed and nondepressed subjects. *Science*, 1976, *192*, 489–491. (b)

Schwartz, G. E., Fair, P. L., Mandel, M. R., Salt, P., Mieske, M., & Klerman, G. L. Facial electromyography in the assessment of improvement in depression. *Psychosomatic Medicine*, 1978, *40*, 355–360.

Schwartz, G. E., & Higgins, J. D. Cardiac activity preparatory to overt and covert behavior. *Science*, 1971, *173*, 1144–1146.

Shmavonian, B. M., Miller, L. H., & Cohen, S. I. Differences among age and sex groups in electro-

dermal conditioning. *Psychophysiology,* 1968, *5,* 119–131.

Sokolov, A. N. Speech-motor afferentiation and the problem of brain mechanisms. *Soviet Psychology,* 1967, *6,* 3–15.

Sokolov, A. N. Studies of the speech mechanisms of thinking. In M. Cole & I. Maltzman (Eds.), *A handbook of contemporary Soviet psychology.* New York: Basic Books, 1969.

Sokolov, A. N. *Inner speech and thought.* New York: Plenum Press, 1972.

Tognacci, L. N., & Cook, S. W. Conditioned autonomic responses as bidirectional indicators of racial attitude. *Journal of Personality and Social Psychology,* 1975, *31,* 137–144.

Tursky, B., Schwartz, G. E., & Crider, A. Differential patterns of heart rate and skin resistance during a digit-transformation task. *Journal of Experimental Psychology,* 1970, *83,* 451–457.

Wallace, S., & Wallace, J. D. Hemodensitometry in the evaluation of cerebral circulation. *Radiology,* 1968, *90,* 454–465.

Received October 25, 1978 ■

Instructions for Submission of Manuscripts

Papers submitted to the *Journal of Personality and Social Psychology* after April 1, 1979 will be considered under new policies approved by the Publications and Communications Board. Effective with the 1980 volume of the journal, the following policies will apply:

The *Journal of Personality and Social Psychology* publishes original papers in all areas of personality and social psychology. It emphasizes empirical reports but may include specialized theoretical, methodological, and review papers. Beginning in April 1980, the journal will be divided into 3 independently edited sections: ATTITUDES AND SOCIAL COGNITION, INTERPERSONAL RELATIONS AND GROUP PROCESSES, and PERSONALITY PROCESSES AND INDIVIDUAL DIFFERENCES. Authors should submit manuscripts to the appropriate section editor according to the section definitions and manuscript preparation instructions provided on page 468 of the March 1979 issue of the journal.

Section editors' addresses are as follows: ATTITUDES AND SOCIAL COGNITION, Melvin Manis, Editor, Research Center for Group Dynamics, Institute for Social Research, University of Michigan, Ann Arbor, Michigan 48106; INTERPERSONAL RELATIONS AND GROUP PROCESSES, Ivan D. Steiner, Editor, Department of Psychology, University of Massachusetts, Amherst, Massachusetts 01003; PERSONALITY PROCESSES AND INDIVIDUAL DIFFERENCES, Robert Hogan, Editor, Johns Hopkins University, Department of Psychology, 33rd and Charles Streets, Baltimore, Maryland 21218.

Journal of Personality and Social Psychology
1979, Vol. 37, No. 12, 2200-2211

Remembering Schema-Consistent Information: Effects of a Balance Schema on Recognition Memory

Keith P. Sentis and Eugene Burnstein
University of Michigan

The major premise of this article is that schema-consistent (balanced) information is structurally represented in memory in a qualitatively different fashion from schema-inconsistent (imbalanced) material. The former closely resembles overlearned information and "easily chunked" information in that multicomponent stimuli are represented in memory as unitary wholes, whereas the latter is stored as discrete propositions. Subjects studied short scenarios that described either balanced or imbalanced structures involving three relations and then answered questions that required retrieval of either one, two, or all three of the relationships. For imbalanced information, reaction times to questions increased with the number of to-be-retrieved relations, whereas with balanced information, reaction times actually *decreased* as the number of to-be-retrieved relationships increased. Thus, for balanced structures, all three relations were more quickly verified than any two, and any two were more quickly verified than any one. A model is proposed for the operation of the balance schema that explains why the whole is retrieved more easily than its parts.

During at least the past 100 years, there has been a sustained belief that the concept of *knowledge structures* is useful for explaining mental processes. This is evident in the remarkable similarity between the descriptions of cognitive organization contained in the works of several nineteenth-century theorists and those offered by current researchers in cognitive psychology and artificial intelligence. To cite only one example, regarding cognitive representations of frequently repeated episodes, Butler (1877) felt that "the memory of many past performances strikes a sort of fused balance in the mind, which results in a general method of procedure with but little conscious memory of even the latest performances, and with none whatever of by far the greater number of the remoter ones" (p. 158). A recent iteration of Butler's ideas can be found in Schank and Abelson (1977), who note that "as an economy measure in the storage of episodes, when enough of them are alike they are remembered in terms of a standardized generalized episode which we will call a script . . . This economy of storage has a side effect of poor memory for detail" (p. 19).

The research presented here is concerned with knowledge structures pertaining to the social environment, that is, *social schemas*. Social schemas are coherent conceptual frameworks for representing relationships among social stimuli that guide the individual in organizing the social world. These cognitive structures are built up on the basis of experience with social reality and are active in the interpretation and comprehension of novel social events. In particular, we will focus on schemas about social relationships that can

Portions of these data were reported at the American Psychological Association meeting, San Francisco, August 1977. This research was supported in part by Grant MH29140-02 from the National Institute for Mental Health. We are indebted to Roberta Klatzky and Hazel Markus for their valuable advice and commentary throughout all phases of this research. We also thank Michael Siladi for his assistance with the data analysis, as well as Ruth Hamill, Stephen Slade, and Ed Smith for their comments on an earlier version of this article.

Requests for reprints should be sent to Keith Sentis, who is now at Bell Laboratories, Murray Hill, New Jersey 07974.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3712-2200$00.75

be described in terms of structural balance. Heider (1958) postulated that social relations, to the extent that they form a balanced structure, are subject to Gestaltlike organizational principles in that balanced relations are perceived as a cohesive unit, whereas imbalanced relations are not—the latter, in fact, are thought to be disharmonious and cognitively segregated (Heider, 1946).

Naturally, a schema about relations between people should influence the processing of social information. Indeed, one of Heider's important hypotheses was that information consistent with a balance schema will be better remembered than information inconsistent with the schema (1958, p. 184). Prior investigations of this hypothesis have employed a variety of procedures, including paired associates learning (Zajonc & Burnstein, 1965a, 1965b; Zajonc & Sherman, 1967) and conceptual rule learning (Cottrell, 1975), as well as recall (Gerard & Fleischer, 1967) and recognition tasks (Delia & Crockett, 1973; Sherman & Wolosin, 1973). The results, however, have provided only checkered support, with contradictory findings occurring not only across but also within methodologies.

The present inquiry into the nature of the balance schema, especially its effects on memory, uses chronometric procedures that are standard in the study of information processing but are relatively novel in the area of social schemas. Substantively, it is concerned with how the balance schema functions. This issue was most recently explored by Picek, Sherman, and Shiffrin (1975). They had subjects indicate whether the relationships between people in stories they had read earlier were positive or negative or were not presented. Fewer errors were observed on the balanced stories than on the imbalanced ones. Moreover, there was a distinctive effect of the order in which the relationships were presented, with the most errors occurring on the third and succeeding relationships in the imbalanced structures—the point in the narrative at which the social situations became unambiguously imbalanced (or balanced). These findings suggest that once the degree of balance of a social structure could be determined, the balance schema resulted in schema-consistent information being stored in a simpler and more available manner than is schema-inconsistent information. As a result, subjects were able to retrieve the former more easily from memory. The critical question, which remains open, is *how* does it come about that schema-consistent information is stored more efficiently and hence is more easily retrieved than schema-inconsistent information.

Retrieval processes have been extensively investigated using a paradigm in which subjects study several sentences, are later presented with a probe sentence, and are asked to decide whether they have seen it before (e.g., Anderson, 1976; Anderson & Bower, 1973). Some of the discoveries using this procedure are important for our analysis of social schemas, in particular the finding that the more facts learned about a particular concept, the longer it takes to recognize correctly a relevant sentence as having appeared before ("old") or as not having appeared before ("new"). The relationship between the number of sentences learned about a concept and sentence recognition time has been termed the "fanning effect." This name derives from the memory model used in this research, which postulates that the concept becomes a central node in a memory network and that facts about the concept are linked to this central node. These facts are thought to "fan out" from the concept node. As their number increases, the number of network paths to be searched in the recognition process also increases, and therefore the more time-consuming the search process. As a consequence, the more that is known about a concept, the longer it takes to correctly recognize any sentence about that concept.

The counterintuitive implications of the fanning effect are that an increase in knowledge about a topic hinders our ability to answer questions about the topic readily or that an expert will answer questions more slowly than a novice. The fanning effect thus poses a paradox. How can someone with expertise on a certain topic answer questions about the topic more readily than a less knowledgeable person, despite having to search through a much larger data base of

relevant information? One class of explanations for this paradox centers on the frequency with which the facts are retrieved. Hayes-Roth (1977), for instance, has shown that the fanning effect disappeared after extended practice with the same set of facts. More relevant to our purpose, however, is the work by Smith, Adams, and Schorr (1978), who were interested in how the retrieval interference represented by the fanning effect could be overcome when the facts involved have *not* been frequently retrieved. They observed that when subjects learned sets of facts that could be *combined as a meaningful unit* by reference to knowledge about the world, the fanning effect was eliminated. These findings imply that information that can be *integrated* is more easily retrieved than information that cannot be integrated. Recall that a central theme in Heider's analysis of social perception is that balanced social relationships form a Gestalt, that is, they tend to be represented as integrated wholes or units. Imbalanced relations, on the other hand, are postulated to be cognitively segregated. If this is true, it would seem to follow from the research on the fanning effect by Smith et al. that balanced relationships are retrieved with relative ease *because* the balance schema allows for a highly integrated representation in memory of schema-consistent information. This means that the schema serves as a mapping function and generates an integrated memory representation of the individual components of information that are consistent with it. More specifically, such an analysis implies that (a) the balance schema integrates information consistent with the schema, which (b) results in less retrieval interference for balanced relationships than for imbalanced ones; consequently, (c) memory for balanced structures is relatively superior. As mentioned earlier, Smith et al. (1978) showed that integration of facts offset retrieval interference and reduced recognition times. If balanced information is stored in memory in a more integrated fashion than imbalanced information (and hence is more easily retrievable), this should be reflected in lower reaction times (RTs) for questions about the balanced structures than for questions about imbalanced structures.

In essence, we propose that balanced social relationships are represented in memory in a fundamentally different fashion from imbalanced relationships. We investigated the structural differences in the representation of balanced and imbalanced information by using a technique originated by Sternberg (1966, 1969, 1974), whereby subjects memorize a set of items and then respond to a test item by indicating whether or not the latter is a member of the memorized set. If RT increases as a function of size of the memory set, this indicates a *serial comparison* between the probe stimulus and the memory set items. However, if RT does not increase with memory set size, then *parallel comparisons* would be indicated.[1] In the former case, a positive slope for the RT function with respect to set size is a sensitive index of the time it takes for each comparison. Furthermore, the scanning of the memory set in an item-by-item fashion implies that the components of the memory set are represented as discrete units in memory. In contrast, an RT function with no slope would mean that the several items could be processed as rapidly as one. This parallel processing would indicate that the components of the memory set are stored as an integrated representation such that the items are available for comparison all at once (Hayes-Roth, 1977). Variations of Sternberg's technique have been used to investigate the issue of integration of visual images. For example, De Rosa and Tkacz (1976) found no effect of set size on RT in a Sternberg task when the material in the memory set could be integrated into a single representation. Similar results were obtained by Guenther (Note 1) using cartoonlike pictorial stimuli and by Salthouse (1977) using

[1] As discussed by Townsend (1972), a positive RT slope would also be indicative of a limited capacity parallel process. In addition, Anderson (1976) demonstrated that a serial model can make equivalent predictions to any parallel model. However, in many instances, the equivalent serial (or parallel) model must incorporate some rather nonintuitive assumptions. Thus, for our purposes, the original distinction will suffice.

dot pattern prototypes. Finally, Thompson and Klatzky (1978) suggest that Gestaltlike organizational principles are used to encode visual stimuli and that under certain conditions these principles facilitate parallel, unlimited capacity processing.

The present research on the balance schema used a variant of the Sternberg technique to evaluate the hypothesis that information consistent with a balance schema is represented in memory as an integrated unit. In this study, subjects studied two scenarios describing P-O-X triads (i.e., social structures comprised of the affective relationships between two people and their attitudes toward an issue). One scenario described a balanced structure, the other an imbalanced one. Following the study session, subjects answered questions about these scenarios that required them to verify one, two, or all three of the relationships in the structure. Notice that in our version of the Sternberg task, the size of the memory set is constant and the size of the probe (i.e., the size of the retrieval set) is varied. We expected that the slope of RT as a function of the number of structural relationships to be verified should reflect the integrative function of the balance schema. This means that if the balance schema results in subjects forming an integrated representation of the scenarios, then the RT function ought to reflect differences in the degree to which this integration occurred. Specifically, as the degree of integration increases, the slope of RT as a function of the size of the retrieval set should decrease. We predict, therefore, that because schema-consistent information is more readily integrated than schema-inconsistent information, the former should exhibit a flat slope and the latter a positive slope for the function relating RT to the number of relationships to be verified in the question.

Method

Subjects

Subjects were 38 undergraduates enrolled in introductory psychology classes. The 35 males and 3 females received course credit for their participation. Data from six subjects who made more than six errors during the 56 trials were excluded from the analyses.

Materials

Scenarios. Two short descriptions of hypothetical stuations were used, each involving two persons of the same sex and an issue. Each of the scenarios was five sentences long and contained the following information: (a) brief description of both individuals, (b) the interpersonal sentiment relationship between the individuals, (c) the relationship between the persons and the issue, that is, their attitudes toward the issue.

Two versions of each scenario were generated by reversing the relationships between the entities. This reversal of relationships produced one version in a balanced state and the other in an imbalanced one. Two sets of scenarios were formed by varying the order of occurrence of the balanced scenario. Approximately equal numbers of subjects were randomly assigned to each set of scenarios. In both sets, the scenario about admissions quotas preceded the abortion scenario. In the two scenarios that are reproduced below, the reversed relationships that appear in parentheses were substituted for the preceding italicized sections to produce imbalanced versions of the scenarios.

Scenario 1. Mary has been married for three years and has a year-old child. She is very much *opposed to* (in favor of) liberal abortion legislation and works on a local committee that is lobbying *against* (for) such legislation. Susan strongly *opposes* (favors) liberal abortion laws and is one of the most active members of the local committee. Mary and Susan became acquainted through their work together on the committee. They have grown to *like* (dislike) each other very much.

Scenario 2. Bill is a senior at a large university and has majored in Sociology. Bill is applying to several prestigious law schools and strongly *opposes* (favors) the graduate schools' admissions policy that is based on racial quotas. John is also a senior at the same university and he is an Urban Studies major. John is applying to these law schools also and is very much *opposed to* (in favor of) the quota system. Bill and John *are* (are not) friends; in fact, they *like* (dislike) each other very much.

Questions. Fourteen distinct questions were constructed about each scenario (28 in all). Since we were interested in analyzing RT as a function of the complexity of the question, the questions were presented in a schematic format rather than in verbal form, in an attempt to minimize differences in question encoding times. Each schematic question took the form of three words from the scenario (two names and an issue) and from one to three symbols (+ or −), used to represent the relationships between the entities in the scenarios. Typical schematic questions are shown in Figure 1. The spatial arrangement of the words and symbols was constant across questions. Subjects were required to verify the rela-

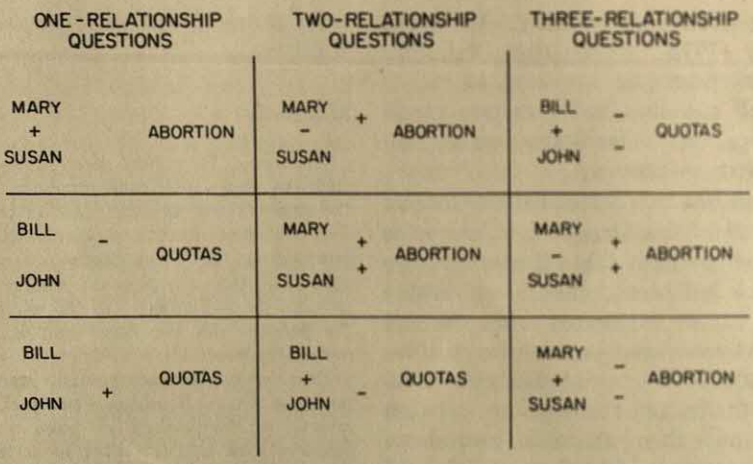


Figure 1. Examples of schematic questions presented to subjects.

tionships represented by the symbol(s) in the schematic question against the corresponding relationship(s) in the scenario by responding *true* or *false* to each question.

The 14 schematic questions about each scenario formed three sets, according to the number of relationships between entities that the subject was required to verify. These sets were:

1. *One-relationship questions*—six questions that required the subject to verify one relationship among the entities in the scenario (see examples in Figure 1, left panel).

2. *Two-relationship questions*—six questions that required verification of two relationships (Figure 1, center panel).

3. *Three-relationship questions*—two questions requiring verification of all three relationships in the scenario (Figure 1, right panel).

In seven of the 14 schematic questions, each relationship in the question accurately reflected those in the scenario. Thus, the correct response to these questions was *true*. In the other seven questions, the relationship(s) were opposite to those in the scenario, thus requiring a *false* response. Note that in order to completely counterbalance the question format across the true–false and balance–imbalance factors (see *Design* section), the following constraint was required. For any given schematic question, either *every* symbol accurately reflected the relationship in the scenario, or *none* did, that is, all were the opposite of those in the scenario.

Of the six one-relationship questions described in Set 1 above, two questions (one true and one false) pertained to the relationship between the people in the scenario (as illustrated at the top of the left panel of Figure 1), and four questions (two true and two false) pertained to the relationship between the people and the impersonal entity in the scenario (middle and bottom of left panel in Figure 1).

Question Set 2 was composed of two questions (one true and one false) about each of the possible subsets of two relationships among the three entities. The three possible subsets of relationships are shown in the center panel of Figure 1.

In Question Set 3, all three relationships were specified in each question.

Four presentation orders (two for each set of scenarios) were prepared, each order consisting of two consecutive random permutations of 28. This resulted in a total of 56 trials that were broken into consecutive blocks in each of which the entire set of 28 questions occurred. Approximately equal numbers of subjects were assigned to each presentation order.

Design

The characteristics of the questions as described above, coupled with the fact that the reversal of relationships changed a given scenario from a balanced to an imbalanced state, meant that the correct response to a particular schematic question was either true or false and concerned either a balanced or an imbalanced structure, depending on the version of the scenario. Thus, collapsing across subjects effectively counterbalanced for scenario effects, question order effects, and their interaction. This resulted in a $3 \times 3 \times 2 \times 2 \times 2$ within-subjects design, the factors being (a) the number of relationships in the schematic question; (b) the particular subset of relationships that the question dealt with; (c) the correct response (true or false); (d) the type of structure (balanced or imbalanced); and (e) the replication block (1 or 2).

Apparatus

Presentation of questions and recording of responses and RTs were controlled by a PDP 11 computer equipped with three ADDS 980 video terminals. Six subjects were run at a time, each proceeding through the trials at his or her own pace.

Each subject sat in an individual soundproofed cubicle with a 12-inch (30.48 cm) monitor and two movable response-button boxes. Subjects were instructed to move the button box labeled TRUE to their dominant hand, to position the boxes comfortably, and to rest their forefingers on the buttons. The schematic questions were centered on the screen of the monitor, which was approximately at eye level.

Procedure

Subjects were seated in their cubicles and were handed an introductory sheet that described the experiment as one concerning "the manner in which individuals learn and use information about the world around them." The sheet also asked the subjects to study "very short descriptions of situations that one might encounter in the course of daily life" and informed them that they would be asked questions about specific details of the descriptions.

After the introductory sheet, the subjects were given two minutes to study each of the scenarios that were typewritten on separate sheets.[2] Following the study period, subjects were given an instruction booklet that described the format of the questions and stressed the importance of answering them by pressing the true or false button as quickly as possible while maintaining perfect accuracy.

The instruction booklet was followed by 12 practice trials (six schematic questions about each scenario). The format of the practice questions was identical to that of the questions used in the actual trials. During the practice trials, the first incorrect response made by a given subject was flagged by an INCORRECT message at the center of the monitor. The second and succeeding errors by that subject were flagged with the message INCORRECT. COMPLETE ACCURACY IS VERY IMPORTANT. SLOW DOWN IF NECESSARY.

Following the practice trials, subjects were asked if they had any questions. During the 56 actual trials that followed, a 2-sec pause was interpolated between a response and the presentation of the next schematic question.

Results

The overall error rate was 4.6% with slightly fewer errors occurring on balanced information (4.2%) than on imbalanced information (5.0%). The analysis of RT includes correct responses only. Data from the practice trials were excluded from the analyses, as were data from trials with aberrant RTs, defined as those more than three standard deviations from the subject's mean (1.3% of the trials).

The principal data are the parameters that describe RT as a function of the number of structural relationships to be verified. Straight lines were fitted to each subject's data, and the parameters (slopes and intercepts) of these RT functions were analyzed in $2 \times 2 \times 2$ within-subjects analyses of variance (ANOVAs), with factors of structure type (balanced/imbalanced), replication block (first/second), and correct answer (true/false). The results reported below are based on data collapsed over replication blocks because there was no effect of replication block on the slopes of the RT functions, $F(1, 31) < 1.0$, and the effect of replication block on the intercepts of the RT functions indicated that subjects merely improved their performance with practice, $F(1, 31) = 8.74$, $p < .006$).

The mean RTs for correct responses appear in Figure 2, together with the means of the parameters of straight lines fitted to each subject's data using a least-squares criterion. In general, the results confirm the predictions regarding the slopes of RT as a function of the number of relationships to be verified. For trials in which the correct response was *true*, the 95% confidence interval for the balanced slope ranges from −95 to −15 msec, whereas the same confidence interval for the imbalanced slope ranges from 84 to 164 msec. For the false trials, the 95% confidence intervals for the balanced and imbalanced slopes were −19 to 61 msec and −32 to 48 msec, respectively.

---

[2] Earlier studies in this program (Sentis, Note 2) indicated that study time is an important factor in this paradigm. When subjects were allowed to allocate their study time differentially between balanced and imbalanced stories, their subsequent recognition performance (in terms of both errors and RT) was markedly superior for imbalanced information. Other studies that employed a single study period for balanced and imbalanced structures also reported superior memory performance on imbalanced information (Gerard & Fleischer, 1967; Spiro & Sherif, 1975). This result appears to be due to the greater distinctiveness of the imbalanced scenarios relative to the balanced ones. The greater salience of imbalanced information apparently commands a disproportionate share of subjects' attention during the study phase and subsequently results in better performance on imbalanced structures during the test phase.

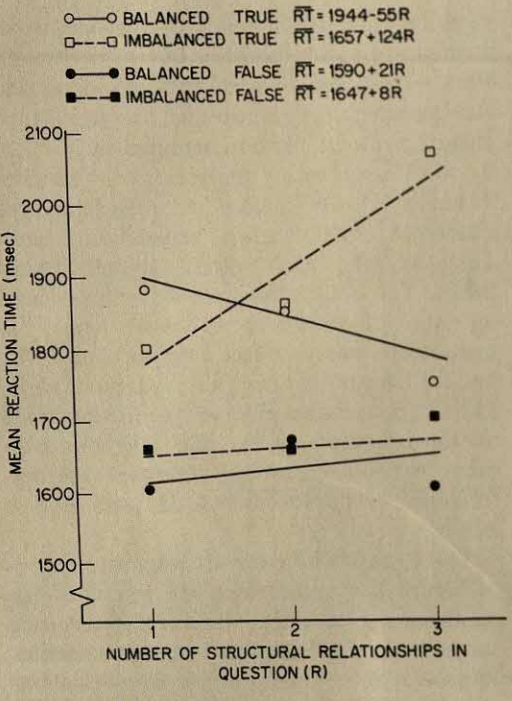


Figure 2. Mean reaction time (RT) for correct responses as a function of number of structural relationships to be verified in the questions. (Least-squares linear functions were applied to individual subject means. Shown are the means of the parameters of the RT functions by type of structure and response.)

Discussion

The results show that the balance schema had a large effect on the processing associated with correct *true* responses, as indicated by the slopes of the *true* RT functions, which differed considerably according to whether the question to be answered concerned a balanced or an imbalanced structure. As can readily be seen in Figure 2, the RT results from the *true* responses confirm the prediction that imbalanced information is not well integrated in memory and thus exhibits a *positive* slope for RT as a function of the number of structural relationships retrieved from memory. In contrast, balanced information yielded not merely a flat slope for the *true* RT function but a reliably *negative* slope. Thus, there is the distinct possibility that the balance schema results in the formation of such highly integrated memory structures that the whole structure is more readily accessed than a subpart. Indeed, it seems that the smaller the subpart, the more difficult the access. In short, not only was parallel processing of the structural relationships indicated, but more intriguingly, the balance schema, by virtue of its integrative function, actually appears to inhibit the processing of subparts of the structure.

In contrast to the *true* responses, the slopes of the RT functions from the correct *false* responses did not differ according to the degree of structural balance. In addition, the RTs to *false* questions were considerably faster than the *true* RTs. These qualitative differences between the *true* and *false* RT functions suggest a two-stage memory search model that is similar to the recognition memory model of Atkinson and Juola (1973). For information stored in long-term store (LTS), Atkinson and Juola propose a recognition model assuming that the subject either makes a fast initial response based on the familiarity of the test item or, if the item is neither extremely familiar or extremely unfamiliar, delays responding until a more extensive search of the memorized information is carried out. As shown in Figure 3, for the present data we assume that the familiarity criterion for a fast *no* is set above the bulk of the "false" distribution and the fast *yes* familiarity criterion is set above most of the "true" responses. The area of extended search between these criteria overlaps most of the *true* distribution, and an account of this extended search process will be given below. Note that the lower tail of the *true* distribution falls below the fast *no* criterion, and thus the incorrect *true* responses (i.e., responses that ought to have been true but were not) should be relatively fast. On the other hand, the upper tail of the *false* distribution lies in the search area, and hence subjects' incorrect *false* responses will occur after a search and should be relatively slow. The RT data for errors are consistent with this reasoning, with the 42 incorrect *true* responses averaging 1,617 msec and the 36 incorrect *false* responses averaging 1,785 msec (although this difference is not statistically significant).

Anecdotal evidence also supports the pro-

posed two-stage model. Subjects often reported that they were sure they had made some errors, despite receiving no feedback during the actual question session. This introspection is in line with the model, which posits that even if the initial familiarity of an item produces a decision to respond quickly, the extended memory search continues and quite likely results in the subject's confirming that he has or has not made the correct response.

Let us now consider the processing that occurs during the extended search initiated for *true* responses. In particular, we wish to address the theoretical significance of the negative slope for the balanced *true* RT function. Of course, our study is not the first to discover that the information about a complete structure maybe more accessible than information about any part. For example, Postman, Bruner, and Walk (1951) found that subjects were able to determine what a defective word ought to be before they noticed which letter in the word was defective. Similarly, Horowitz, Day, Light, and White (1968) showed subjects a word with one letter missing and found that they could report the completed word twice as fast as the missing letter; and Johnson (1975) demonstrated that subjects identify words faster than their constituent letters. Using a paradigm similar to the present studies, Hayes-Roth (1977) observed that after extensive learning of three-part propositions, which presumably resulted in integrated ("unitized") memory representations, two-part subsets of the three-part propositions were verified faster than one-part subsets. In light of these findings, it is reasonable to interpret the negative slope of the balanced RT function for *true* responses as indicating the balance schema results in such a highly integrated cognitive representation of schema-consistent information that subjects found it more difficult to process subsets of the balanced structure than to access it in its entirety. The implication is that there is a cost associated with the integrative function of the balance schema, namely, the integrated memory structure must be decomposed in order to access subsets of it. Because the three-relationship questions require no decomposition, they can be answered most quickly of all. On the other hand, two-relationship questions, since they require that one relationship be removed from the structure prior to responding, are answered more slowly. The one-relationship questions require still another decomposition step and are answered the most slowly.

The present results and the other similar findings discussed above suggest the following processing model for information about social structures varying in their degree of balance. It assumes that structural relationships in the scenarios are transferred to LTS and are maintained there in a particular format (detailed below). The question-answering task requires retrieval of structural relationships into short-term store or working memory, where the decision-making operations needed

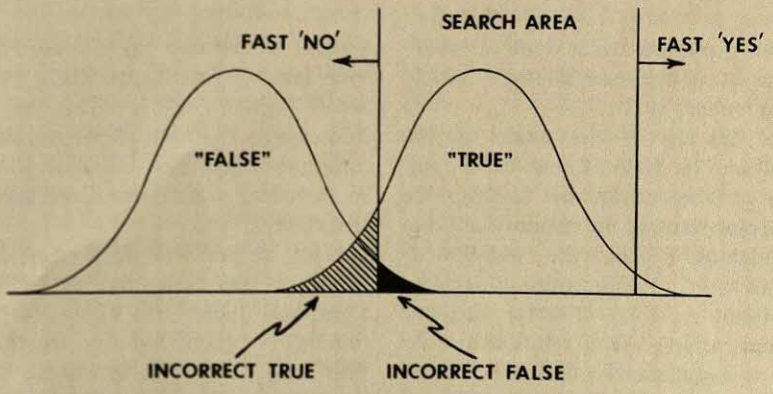


Figure 3. Theoretical distribution of familiarity values for *true* and *false* questions with assumed high and low criteria set to result in the obtained pattern of reaction times.

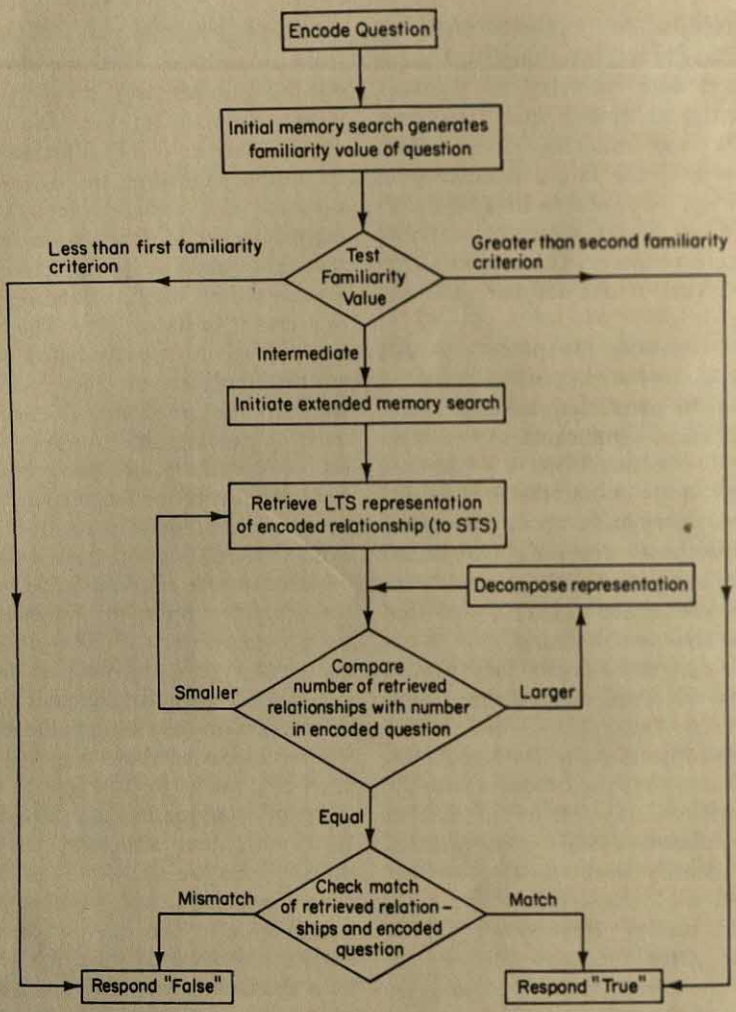


Figure 4. Processing algorithm for answering questions about balanced and imbalanced structures.

to answer the questions occur (for an overview of memory models, see Klatzky, 1975; Shiffrin & Schneider, 1977).

A primary feature of the model is the mode of storage in LTS. Imbalanced relationships are inherently difficult to integrate and are therefore stored in memory as discrete propositions. Thus, the imbalanced P-O-X triads used in the present studies would be stored as three discrete memory representations, one for each relationship. As a result, each imbalanced relationship must be retrieved individually. Balanced information on the other hand, is integrated into unitized nodes that are retrieved in an all-or-none fashion (cf. Hayes-Roth, 1977; Schneider & Shiffrin, 1977). Thus in the present studies, the three balanced relationships that comprise a schema-consistent scenario would be stored as a single unitized memory representation.

With respect to the question-answering phase of this task, the model assumes the algorithm shown in Figure 4. The initial familiarity check results in the relatively "fast and flat" RT functions for the *false* responses that are not affected by the degree of structural balance. Assuming the fast-check

criterion values that are shown in Figure 3, the bulk of the *true* questions generate intermediate familiarity values and would necessitate a more extended memory search.

This extended search incorporates two primary operations—retrieval from LTS and decomposition. The sequencing of these operations depends both on the mode of information storage (discrete propositions vs. unitized nodes) and the complexity of the question to be answered. Consider the sequence of operations that occurs in the algorithm shown in Figure 4 for *true* responses to questions. First, the LTS representation of a relationship in the encoded question is retrieved to short-term store. Next, the number of retrieved relationships is compared to the number in the encoded question. When their numbers are equal, the match of the retrieved relationships with the encoded question is checked and a response is made. Following the first retrieval from LTS, the outcome of the stage that compares the number of relationships retrieved and the number in the question determines the remaining sequence of retrieval or decomposition operations. Imbalanced structures are stored as discrete propositions, and the relationships that comprise them must be retrieved one at a time. Because each relationship to be verified in the question requires additional retrieval operations, the number of such operations covaries with the complexity of the questions pertaining to an imbalanced structure. Thus, the 124-msec slope for the imbalanced *true* RT function represents the time required to search LTS and retrieve a particular structural relationship. In contrast, all of the relationships in a balanced structure are stored in a single unitized node, and consequently a memory search initiated for any single relationship results in retrieval of the entire structure. The storage of balanced structures in unitized nodes permits *true* questions about those scenarios to be answered after only a single retrieval operation, regardless of the number of relationships in the question. However, in order to answer one- and two-relationship *true* questions, the unitized nodes that are retrieved from LTS must be decomposed (cf. Hayes-Roth, 1977). The −55-msec slope for the balanced information reflects the time required for these decomposition operations and comprises the "penalty" in the model.

How reasonable is the model outlined above? First, the RT data from the errors, as well as subjects' introspection regarding these errors, are consistent with the model. Second, the model posits that decomposition operations take place in working memory. These operations consist of transforming the information retrieved from LTS into a format suitable for answering the questions. It seems reasonable to *assume,* therefore, that these relatively simple transformations would be less time-consuming than retrieval from LTS. The relative magnitudes of the slopes of the RT functions representing the decomposition and retrieval operations bear this out, with retrieval taking roughly twice as long as decomposition (124 msec versus 55 msec). Finally, the model predicts equal RTs for *true* responses requiring the same retrieval and decomposition operations. Two conditions in the present experiment required the same set of operations, namely a single retrieval operation with no decomposition operations. The conditions requiring the same set of operations were the one-relationship *true* questions regarding imbalanced structures and the three-relationships *true* questions concerning balanced structures. Given the model, these two conditions should exhibit comparable RTs, and in fact, the predicted values for these conditions (based on the equations in Figure 2) differ by only *2 msec.*

In summary, the present model of the balance schema is a fairly well-specified account of how information concerning social relationships is processed. It accounts for the different signs and magnitudes of the slopes of the RT functions associated with answering questions regarding balanced and imbalanced structures. The model also raises a number of questions. At what point in the encoding sequence is the balance schema activated? How are the retrieval and comparison operations sequenced in working memory? What is the effect of the balance schema on memory for larger social structures? A more complete specification of the model waits on further

research, especially with respect to the temporal organization of retrieval, decomposition, and comparison operations in working memory, as well as the entire encoding process that results in the different modes of storage in LTS.

Our research on the balance schema presents evidence supporting the hypothesis that information consistent with a schema is stored in an integrated fashion. In general, this integrated mode of storage will facilitate retrieval of schema-consistent information and will enhance the processing of such material. This integrative function, however, is a two-edged sword. If the schema results in such a high degree of integration that the relevant information structure becomes unitized, then processing is only facilitated to the extent that there is a match between the amount of information to be retrieved (i.e., the size of the probe) and the size of the unitized structure. Under certain conditions, therefore, the integration function of schemas may be dysfunctional because of the necessity of decomposing the unitized memory representation to access the subparts of the structure.

In certain respects, the integrative function of schemas can be thought of as "chunking," that is, schema-consistent information seems to be processed in larger units than is schema-inconsistent material. Similar results have been obtained in research on the perceptual abilities of chess experts (Chase & Simon, 1973) and masters of the oriental game of Go (Reitman, 1976). This research indicates that experts in these games (individuals with a schema about chess or Go) recognize larger patterns of pieces on the board than do novices (players without a schema). However, differences between the expert and novice do not obtain for randomly placed pieces (i.e., for schema-inconsistent information). Thus, one reason that an expert's play is superior to that of a novice is this ability to process the schema-consistent information in larger units or chunks. In related research, Chiesi, Spilich, and Voss (1979) have shown that the schema, or structured knowledge, possessed by baseball fans allows them to integrate or chunk baseball-related information, thereby enhancing the processing of such information. Moreover, Spilich, Vesonder, Chiesi, and Voss (1979) report that a baseball schema results in the integration of the discrete events in a game into meaningful units or sequences and leads to subsequent recall of larger chunks of game-related information. The findings of Spilich et al. also suggest that a schema's integration or chunking function is hierarchical in nature and results in easier retrieval of higher-order units than of lower-order constituents—a finding completely consistent with our own.

To the extent that the integration function is a general property of social schemas, the present research has a number of implications for social cognition in general and for interpersonal communication in particular. For example, interpersonal communication typically involves the transmission of one chunk at a time, and as a result, efficient communication requires the individuals to use a mutually acceptable chunk size. Consider the case of an expert interacting with a novice. If a novice attempts to query an expert regarding a particular topic, a common source of inefficient communication is the diminutive (low-level) chunks used by the novice to phrase the question. In many cases, the expert does not understand the question and asks the novice for more background information in order to put the question into a meaningful context. The obverse of this problem frequently occurs when an expert does answer a novice's question. The expert's response may often be delivered in large (high-level) chunks incomprehensible to the novice. For example, computer documentation is notoriously difficult to comprehend. It is typically written by a programmer who usually understands not only the particular program to be documented but also the other programs and routines that interface with it, as well as the larger computer system. The crux of the communication difficulty with computer documentation, therefore, seems to be the programmer's failure to decompose the high-level chunks used in an expert's representation of the program into chunks of a size more digestible by novices.

Reference Notes

1. Guenther, R. K. *The representation of picture and prose episodes.* Unpublished doctoral dissertation, University of California, Santa Barbara, 1977.

2. Sentis, K. P. *Effects of social schemas on selective attention.* Unpublished manuscript, University of Michigan, 1977.

References

Anderson, J. R. *Language, memory, and thought.* Hillsdale, N.J.: Erlbaum, 1976.

Anderson, J. R., & Bower, G. H. *Human associative memory.* Washington: Winston, 1973.

Atkinson, R. C., & Juola, J. F. Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.), *Attention and performance IV.* New York: Academic Press, 1973.

Butler, S. *Life and habit.* London: Trubner, 1877.

Chase, W. G., & Simon, H. A. Perception in chess. *Cognitive Psychology,* 1973, *4,* 55–81.

Chiesi, G. T., Spilich, G. J., & Voss, J. F. Acquisition of domain-related information in relation to high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior,* 1979, *18,* 257–273.

Cottrell, N. B. Heider's structural balance principle as a conceptual rule. *Journal of Personality and Social Psychology,* 1975, *31,* 713–720.

Delia, J. G., & Crockett, W. H. Social schemas, cognitive complexity, and the learning of social structures. *Journal of Personality,* 1973, *41,* 414–429.

De Rosa, D. V., & Tkacz, S. Memory scanning of organized visual material. *Journal of Experimental Psychology: Human Learning and Memory,* 1976, *2,* 688–694.

Gerard, H. B., & Fleischer, L. Recall and pleasantness of balanced and imbalanced cognitive structures. *Journal of Personality and Social Psychology,* 1967, *7,* 332–337.

Hayes-Roth, B. Evolution of cognitive structures and processes. *Psychological Review,* 1977, *84,* 260–278.

Heider, F. Attitudes and cognitive organization. *Journal of Psychology,* 1946, *21,* 107–112.

Heider, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

Horowitz, L. M., Day, R. S., Light, L. L., & White, M. A. Availability growth and latent verbal learning. *The Journal of General Psychology,* 1968, *78,* 65–83.

Johnson, N. F. On the function of letters in word recognition: Some data and a preliminary model. *Journal of Verbal Learning and Verbal Behavior,* 1975, *14,* 17–29.

Klatzky, R. L. *Human memory: Structures and processes.* San Francisco: Freeman, 1975.

Picek, J. S., Sherman, S. J., & Shiffrin, R. M. Cognitive organization and coding of social structures. *Journal of Personality and Social Psychology,* 1975, *31,* 758–768.

Postman, L., Bruner, J. S., & Walk, R. D. The perception of error. *British Journal of Psychology,* 1951, *42,* 1–10.

Reitman, J. S. Skilled perception in GO: Deducing memory structures from inter-response times. *Cognitive Psychology,* 1976, *8,* 336–356.

Salthouse, T. A. Number of memory representations in perceptual concepts. *Journal of Experimental Psychology: Human Learning and Memory,* 1977, *3,* 18–28.

Schank, R., & Abelson, R. *Scripts, plans, goals, and understanding.* Hillsdale, N.J.: Erlbaum, 1977.

Schneider, W., & Shiffrin, R. M. Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review,* 1977, *84,* 1–67.

Sherman, S. J., & Wolosin, R. J. Cognitive biases in a recognition task. *Journal of Personality,* 1973, *41,* 395–413.

Shiffrin, R. M., & Schneider, W. Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review,* 1977, *84,* 127–190.

Smith, E. E., Adams, N., & Schorr, D. Fact retrieval and the paradox of interference. *Cognitive Psychology,* 1978, *10,* 438–464.

Spilich, G. J., Vesonder, G. T., Chiesi, H. L., & Voss, J. F. Text processing of domain-related information for individuals with high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior,* 1979, *18,* 275–290.

Spiro, R. J., & Sherif, C. W. Consistency and relativity in selective recall with differing ego-involvement. *British Journal of Social and Clinical Psychology,* 1975, *14,* 351–361.

Sternberg, S. High-speed scanning in human memory. *Science,* 1966, *153,* 652–654.

Sternberg, S. Memory scanning: Mental processes revealed by reaction-time experiments. *American Scientist,* 1969, *57,* 421–457.

Sternberg, S. Memory scanning: New findings and current controversies. *Quarterly Journal of Experimental Psychology,* 1974, *27,* 1–32.

Thompson, A. L., & Klatzky, R. L. Studies of visual synthesis: Integration of fragments into forms. *Journal of Experimental Psychology: Human Perception and Performance,* 1978, *4,* 244–263.

Townsend, J. T. Some results on the identifiability of parallel and serial processes. *British Journal of Mathematical and Statistical Psychology,* 1972, *25,* 168–199.

Zajonc, R. B., & Burnstein, E. The learning of balanced and unbalanced social structures. *Journal of Personality,* 1965, *33,* 153–163. (a)

Zajonc, R. B., & Burnstein, E. Structural balance, reciprocity, and positivity as sources of cognitive bias. *Journal of Personality,* 1965, *33,* 570–583. (b)

Zajonc, R. B., & Sherman, S. J. Structural balance and the induction of relations. *Journal of Personality,* 1967, *35,* 635–650.

Received November 7, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 12, 2212-2220

Locus of Control, Interpersonal Trust, and Assertive Behavior Among Newlyweds

William J. Doherty
Department of Family Practice
College of Medicine
University of Iowa

Robert G. Ryder
Human Development and
Family Relations Program
University of Connecticut

This study related two cognitive personality characteristics—locus of control and interpersonal trust—to assertive behavior in a sample of recently married couples. Assertive behavior was measured by the Inventory of Marital Conflicts, an observational procedure in which couples resolve disagreements about hypothetical marital problems. Hypotheses were formulated in terms of individual locus of control as well as two combinations of locus of control and interpersonal trust—internal – low trust and external – high trust. Results showed that internal husbands were more assertive than external husbands in the marital conflict situation, that external – high trust husbands were least assertive, and that internal – low trust wives were highly assertive. These findings are interpreted in light of previous locus-of-control and trust research, as well as in terms of unconventional marital role behavior.

Contemporary research on the marital relationship has focused on the cultural and situational determinants of spouses' attitudes and behavior. Sociologists have emphasized the influence of marital role prescriptions (e.g., Stryker, 1964), whereas psychologists have studied the processes by which spouses shape each other's behavior through reciprocal rewards and punishments (e.g., Weiss & Margolin, 1977).

In contrast to these major interest areas in marital studies, the personality or individual differences dimension of marriage has received relatively little attention in recent years. Historically, this area was once highly popular with researchers. As reviewed by Tharp (1963) and Barry (1970), two research traditions once dominated personality and marriage studies. The first tradition was concerned with the personality correlates of marital "adjustment." From the 1930s through the 1950s, studies consistently found that neurotic traits in individual spouses (especially husbands) were associated with greater marital instability and lower levels of self-reported marital adjustment (Burgess & Cottrell, 1939; Burchinal, Hawkes, & Gardner, 1957; Terman, 1938).

The second personality and marriage research tradition was concerned with whether spouses tend to be similar or complementary in personality characteristics. Although Winch's (1958) research on complementary needs in mate selection stirred considerable controversy, the bulk of the research has supported the similarity hypothesis for married couples (Tharp, 1963). Clore and Byrne (1977) summarized a review of this issue with the following conclusion: "Marriages, and stable marriages in particular, are composed of people with similar personalities, attitudes, and other characteristics" (p. 548).

In recent years, research on the personality aspects of marriage has nearly come to a standstill. At least two problems with past research in this area may help to explain this decline in interest. First, the major findings added little to the common sense notions that "likes marry likes" and that unhappy individuals are apt to have unhappy marriages. Sec-

Requests for reprints should be sent to William J. Doherty, Department of Family Practice, University of Iowa, Iowa City, Iowa 52242.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3712-2212$00.75

ond, on a methodological level, these studies suffered from a reliance on global, atheoretical personality traits as independent variables (e.g., neuroticism and maladjustment) and on self-reports of overall marital adjustment as criterion variables. In the past decade, each of these categories of variables has become the object of considerable scholarly criticism. The traditional trait approach has been attacked by personality psychologists like Walter Mischel (1968, 1973) on the grounds that global trait ascriptions ignore the situational constraints on behavior. In addition, marriage researchers have criticized marital adjustment research on both conceptual and methodological bases, arguing that the construct is ill defined and value laden and that marital adjustment inventories have been psychometrically weak (Laws, 1971; Ryder, 1967; Spanier, 1972; Spanier and Cole, 1976). Perhaps for these reasons, related to the turmoil in both personality psychology and marital studies, scholars have turned to other areas of interest.

The present study attempts to break new ground in personality and marriage research through two innovations: first, the use of cognitive personality constructs as opposed to traditional trait dimensions, and second, a focus on an interaction dimension of marriage —assertive behavior. As far as can be determined, no previous study has found a personality correlate of observed marital interaction.

Specifically, this study examines the relationship between the locus of control orientations of individual spouses and their assertive behavior in a marital disagreement situation. Spouses' interpersonal trust expectancies are used as a moderator personality dimension between locus of control and marital assertiveness.

There is no general agreement on the meaning of assertiveness. After examining the various usages found in the psychological literature, Rich and Schroeder (1976) defined assertive behavior in the general sense as "skills that (a) are concerned with seeking, maintaining, and enhancing reinforcement and (b) occur in interpersonal situations involving the risk of reinforcement loss or the possibility of punishment" (p. 1083). In a review of the related marriage research, Olson and Cromwell (1975) have urged that researchers adopt a standard definition of assertive behavior in marriage, namely, "attempts to change the behavior of the (spouse)" (p. 6). Combining elements of these two definitions, the present study conceptualizes assertive behaviors in marriage as attempts to modify the partner's behavior in order to maintain or enhance one's interests during a marital conflict. This definition is deliberately broad and does not distinguish assertive behavior from aggressive behavior. The assertiveness construct is operationalized in this study primarily through the coding of spouses' influence-attempt behavior during a contrived marital disagreement situation.

As defined and operationalized by Rotter (1966), locus of control is a generalized expectancy that one's outcomes are contingent more on one's own efforts or stable personality characteristics (internal) or more on outside forces such as luck, fate, or powerful others (external). Interpersonal trust as defined by Rotter (1967) is a generalized expectancy that other persons can be relied upon to live up to their verbal promises.

Although locus of control has been studied only rarely in the context of ongoing relationships, a considerable body of research using college student groups has found that internals tend to be less malleable to social influence than externals, as well as more persuasive and assertive (see reviews by Phares, 1976, and Strickland, 1977). These findings are consistent with the proposition that internals, believing more strongly that they can control their outcomes, are more likely to develop and use the social skills necessary to manipulate their environment. Extending this line of reasoning to the marital relationship, one can speculate that married individuals' locus of control orientations may be associated with assertive and yielding behavior vis-à-vis their marriage partner. Presumably, internality would be positively associated with assertiveness. According to Rotter's theory (Rotter, 1975), however, generalized locus-of-control expectancies would be only moderately predictive of behavior in a specific,

ongoing behavioral context such as marriage. Rotter believes that the more familiar the situation, the less powerful the influence of generalized expectancies.

To enhance the predictive value of locus of control in the study of assertive behavior in marriage, interpersonal trust was used as a moderating personality characteristic. In other words, this study assumed that the relationship between locus of control and assertive behavior would be mediated by the individual's interpersonal trust expectancies. This strategy was adopted for the following reasons:

1. Interpersonal trust itself has implications for understanding the marital relationship. Evidence presented by Rotter (Note 1) suggests that low trust individuals are more suspicious and behave more competitively than do high trust individuals. High trust persons, in contrast, are characterized in some studies as pleasant and conventional.

2. Previous research has found combinations of locus of control and interpersonal trust to be better predictors of certain criterion behaviors than is locus of control alone. In particular, Hochreich (1974, 1975) has distinguished theoretically and empirically between two types of external males based on their trust scores. External – low trust persons ("defensive externals") have been found to behave much like internals in achievement situations; however, they tend to project blame for failure onto the environment. External – high trust individuals ("congruent externals"), on the other hand, were found to behave in a relatively passive, noncompetitive fashion consistent with theoretical assumptions about externality. This distinction has been found only in males.

3. Finally, in research reviewed by Seeman (1976), sociologists and political scientists have distinguished fruitfully between variations of trust and control expectations. One prominent hypothesis from this research tradition holds that internal control expectancies in the political area, when combined with low trust in government institutions, are associated with greater political activism.

In sum, two combinations of locus of control and interpersonal trust seem to have special importance for assertive behavior in marriage. If internals are generally more assertive than externals, then the low trusting or suspicious internal should be especially likely to protect his or her self-interest in an interpersonal conflict. Conversely, the least assertive behavior should characterize the generally unassertive external who also has a high degree of confidence in the honest intentions of other persons.

Based on the foregoing review, this study hypothesizes that individual spouses' locus-of-control orientations, with interpersonal trust as a moderating personality dimension, will be associated with levels of assertive behavior by that spouse in marital conflict situations. Specifically, it is hypothesized: (a) that internality will be associated with greater assertiveness, (b) that the combination of internality and low trust will be associated with the highest levels of assertive behavior, and (c) that the combination of externality and high trust will be associated (at least in husbands) with the lowest levels of assertive behavior.

Method

Subjects

Eighty-six recently married couples were recruited by mail from marriage license records of first marriages. All had been married less than one year (average 6.4 months), and they may be characterized as a white, college-educated, middle-class group. None had children. The sample was part of a larger pool of recently married couples generated by the Couples Research Project at the University of Connecticut.

Procedure

Couples recruited in the first phase of data gathering went through a two-session procedure in a university setting. During the initial session, couples wrote essays about their relationship, then completed an omnibus questionnaire that included two instruments used in this study. Couples were then invited back for a second session of testing. Although virtually every couple agreed to return, in practice only about half did so. During the second session, couples were administered the locus-of-control and interpersonal trust scales as well as the marital interaction measure. The n for this first group of couples was 28.

Couples recruited in the second phase were administered all the instruments (except the essays) consecutively in one session. The n for Phase 2 was 58. A comparison between the Phase 1 and Phase 2 groups revealed no differences in locus of control, interpersonal trust, or demographic variables (except that Phase 2 wives were more likely to have completed college). The two groups were combined for data analysis.

Instrumentation

1. *Rotter's Internal–External Locus of Control (I–E) Scale* was used to measure locus of control expectancies. A subset of items from Rotter's scale similar to Mirels' (1970) Personal Control Factor was derived through factor analysis, but it yielded no improvement in terms of relating to the marital variables. Hence, only the whole scale score was retained. The I–E scale contains 29 items in a forced-choice format, including 6 filler items. A higher score indicates a more external orientation.

2. *Rotter's Interpersonal Trust Scale* (Rotter, 1967) was used to measure interpersonal trust expectancies. The scale consists of 25 Likert-type items plus fillers. A higher score indicates a more trusting orientation.

3. *The Inventory of Marital Conflicts* (IMC), a laboratory interaction procedure, was used to provide both a process measure and outcome measure of assertive behavior (Olson & Ryder, 1970). The IMC poses a nonthreatening disagreement situation in which spouses typically try to influence each other's behavior. The procedure presents couples with 18 vignettes in which married couples are involved in typical marital disagreement situations (over money, sex, household chores, etc.). Each spouse individually reads the vignettes and answers two main questions about each conflict situation: (a) Who is primarily responsible for the problem? and (b) should a proposal for resolving the conflict be accepted or rejected? (After the partners completed the individual part of the IMC procedure, they filled out the I–E and interpersonal trust questionnaires, which took about 20 minutes.) Finally, the partners are brought together and asked to discuss each conflict situation and reach a consensus on two questions: (a) Who is primarily responsible for the problem? and (b) which of two mutually exclusive solutions is best? This discussion is audiotape-recorded with no experimenter present in the room.

Disagreement is built into the IMC procedure rather than being left to chance. For 12 of the 18 vignettes, the partners are given somewhat different versions of the conflict situation. The husband's version slants the responsibility for the problem toward the wife, whereas the wife's version makes the husband look more responsible for the marital conflict. The couples were told in the IMC instructions that in some cases, they would be reading different perspectives on the same problem.

The IMC yields two types of data: (a) win scores (who won more of the disagreements) and (b) interaction data taken from the taped discussions. The coding of the couple interaction was based on the Marital and Family Interaction Coding System (Olson & Ryder, Note 2), which contains 15 content codes for verbal behavior (plus an uncoded statement category). The codes are divided conceptually into three categories: task related, assertive, and affective. Each statement made during the IMC discussion procedure was coded directly from the audiotape by two independent coders, whose frequency ratings for each code were averaged. Since the concern of this study is with assertive or influence behavior in a disagreement situation, coding was done only for discussions of vignettes on which the spouses had initially disagreed on their individual forms. Because of damage to several tapes, the final n for the IMC interaction data was 80.

Reliability of the IMC variables. Two types of reliability estimates were performed on the IMC data. First, interrater reliability was assessed for each code by examining the agreement between raters on the scores of each code for each spouse. A code's score was its frequency divided by the total frequency for all interaction codes for that spouse. (This procedure controls for amount of talking.) Corrected by the Spearman–Brown formula for two independent raters, the interrater reliability coefficients ranged from .49 to .99, averaging .86 across all the codes for both spouses.

The second set of reliability procedures examined the internal consistency of the IMC data. To compute internal consistency reliability estimates for the IMC variables, the items (vignettes) were divided into two halves that were equal in the degree of disagreement elicited in the sample. As expected from previous studies (Olson & Ryder, 1970), the win scores proved highly unreliable (split-half $r = .12$, with Spearman–Brown correction) and were retained for exploratory purposes only. For the interaction codes, split-half reliability coefficients, corrected by the Spearman–Brown formula, ranged from .31 to .92, averaging .59 across all codes for both spouses. Codes with interrater or internal consistency reliability coefficients of under .60 (averaged over both spouses) were dropped from subsequent analyses. Remaining were 9 of 15 originally coded variables: initiation of discussion (who begins discussion of each vignette), laughter, outcome question (e.g., "What do you think?"), read/content (e.g., "My story says the husband had tried to stop smoking"), self-disclosure (the first spouse to reveal an opinion on the vignette), partisan opinion (statements defending one's opinion), outcome disagreement (the first statement of disagreement during the discussion of a vignette), process disagreement (all subsequent disagreement statements), and disapproval of spouse (criticizing the other's viewpoints or character). Average interrater and split-half rs for those 9 codes were .91 and .72, respectively.

Deriving an assertiveness interaction dimension. Four of the reliable IMC interaction codes are con-

Table 1
Unrotated Principal Components Factor Matrix For IMC Codes

Code	Factor 1	Factor 2	Factor 3
	Husbands		
Initiation	.723	.178	.304
Self-disclosure	.770	.195	−.378
Read	.587	.106	.586
Outcome question	−.052	−.471	.485
Partisan opinion	−.600	.010	−.310
Process disagreement	−.545	.559	.178
Outcome disagreement	−.748	−.067	.433
Disapproval of spouse	−.069	.770	.188
	Wives		
Initiation	.839	.142	.068
Self-disclosure	.637	.532	−.416
Read	.772	−.262	.123
Outcome question	.295	.132	.546
Partisan opinion	−.493	.234	−.607
Process disagreement	−.409	.724	.089
Outcome disagreement	−.625	−.335	.356
Disapproval of spouse	−.150	.626	.589

Note. IMC = Inventory of Marital Conflicts.

ceptually related to assertive behavior as defined earlier—partisan opinion, outcome disagreement, process disagreement, and disapproval of spouse. Each of these codes represents an effort either to stand one's ground or to change the other's opinion—and thus to gain an advantage in the disagreement. The other codes, except for laughter, appear primarily to represent behaviors that facilitate discussion of the vignettes, for example, initiating the discussion of a vignette, asking for clarification, and describing the content of the story.

As an empirical aid to the selection of the assertiveness dimension, a principal components factor analysis was performed on the reliable IMC codes, with laughter excluded. Separate analyses were conducted for husbands and wives. Unities were used as diagonal elements and no rotation was performed. Inspection of the factor loadings in Table 1 reveals three assertiveness-related codes—partisan opinion, process disagreement, and outcome disagreement—loading in the same direction on Factor 1. Three task-related codes loaded in the opposite direction from the assertiveness codes on this factor. This pattern suggests a negative relationship between task-oriented and assertiveness behaviors. The fourth code that had been related conceptually to assertiveness—disapproval of spouse—did not load on Factor 1. However, a decision was made to retain this code as part of an assertiveness dimension for two reasons: first, it represents a theoretically interesting aspect of marital assertiveness, and second, it correlated with process disagreement .41 for wives and .25 for husbands, which suggests that disapproval of spouse has a fair degree of commonality with one other assertiveness code.

For each spouse an assertiveness score was derived by summing the z scores for partisan opinion, process disagreement, outcome disagreement, and disapproval of spouse. Husband and wife assertiveness scores correlated .45 with each other, a finding that suggests a moderate degree of reciprocity between spouses on these assertive behaviors.

Data Analysis

The primary statistical procedure was two-way analysis of variance of locus of control (internal-external) and interpersonal trust (low-high) on marital assertiveness. Separate analyses were performed on the IMC observed interaction dimension and the win score variable. Of central interest were the main effects for locus-of-control and planned t test comparisons focusing on the external – high trust group and the internal – low trust group.

Extreme groups were created on locus of control and trust by dropping out the middle third of each distribution. Mean scores for the remaining husbands were as follows: Internal = 6.6; external = 15.4; low trust = 60.6, high trust = 76.7. The corresponding means for wives were: Internal = 8.0, external = 16.4; low trust = 59.9, high trust = 79.2. Four cells were created for husbands and wives separately in a 2 × 2 design—internal – low trust, internal – high trust, external – low trust, external – high trust.

Exploratory analyses of variances were also conducted on assertiveness difference sources (husband minus wife), and an examination was made of the relationship between couple locus-of-control patterns (husband I–E × wife I–E) and marital assertiveness.

Results

Descriptive Data

A significant difference between husbands' and wives' locus-of-control scores was found, with husbands more internal (M = 11.0) than wives (M = 12.3). (Studies typically report no sex differences on the Rotter I–E scale in college student samples—see Strickland, 1977; Doherty, in press, also found no sex differences in a national probability sample.)

Spouses' locus-of-control scores correlated at .18, which was nonsignificant. Husbands and wives did not differ in average interpersonal trust scores ($M = 68.5$ for both spouses); the correlation between husband and wife trust scores was $r = .07$, which was not significant.

The relationship between locus of control and trust was different for husbands and wives, with husbands' scores on these two variables correlating at $-.18$ and wives' scores correlating at $-.38$, $p < .001$, indicating a positive association between internality and trust. This pattern contrasts with Hamsher, Geller, & Rotter's (1968) report that I–E and trust typically correlate higher for males than for females in college samples.

Two other husband–wife comparisons are worth noting. First, there were no significant husband–wife differences on the individual variables that comprised the marital assertiveness dimension—process opinion, outcome disagreement, process disagreement, and disapproval of spouse. Second, the win score variable also showed no husband–wife mean differences, with husbands winning an average of only 1.1% more of the disagreements in the IMC procedure.

Locus of Control, Interpersonal Trust, and Marital Assertiveness

Table 2 shows the cell means and sizes for the two-way analyses of variance (Locus of Control × Interpersonal Trust) on the IMC marital assertiveness variables. Results for the observed assertive behavior variable will be presented first, followed by results for the win scores and the supplementary analyses. It should be noted that some of the cells in Table 2 have small *n*s, especially the internal – low trust wife group.

Observed assertive behavior. For husbands, a significant main effect was found for locus of control, $F(1, 38) = 7.93$, $p < .008$, indicating that internal husbands had higher assertiveness scores in the IMC marital conflict situation than did external husbands. Table 2 shows that the external – high trust husband group had the lowest average assertiveness score. A planned t test comparison between the external – high trust group and the other three groups combined showed a significant difference, $t(38) = 3.09$, $p < .004$. Consistent with the congruent versus defensive external distinction, the external – high trust group was significantly less assertive

Table 2
Means and Standard Deviations of Marital Variables By Locus of Control and Interpersonal Trust Groups

Group	IMC assertiveness			IMC wins[a]		
	n	*M*	*SD*	*n*	*M*	*SD*
Husbands						
Internal–low trust	10	.232	2.196	10	59.9	19.5
Internal–high trust	13	.706	3.110	13	53.6	10.7
External–low trust	11	−.599	1.970	12	57.8	16.4
External–high trust	8	−2.867	2.076	9	30.1	13.8
Wives						
Internal–low trust	6	3.012	3.221	6	46.5	13.8
Internal–high trust	14	.512	1.140	16	57.6	12.5
External–low trust	14	−.158	2.620	16	51.2	17.1
External–high trust	9	.611	2.735	10	45.4	19.9

Note. IMC = Inventory of Marital Conflicts.
[a] Win points for husband relative to wife. Higher score represents more husband wins, lower score represents more wife wins.

than the external – low trust group, which was indistinguishable from the internal groups.

For wives, the analysis of variance yielded a nonsignificant effect for locus of control, $F(1, 39) = 2.90$, $p < .097$. However, the interaction effect was significant, $F(1, 39) = 4.37$, $p < .043$. Inspection of the cell means suggests that the extremely high assertiveness scores of the internal – low trust wife group were primarily responsible for the significant interaction effect. A planned t test comparison between the internal – low trust wife group and the other groups combined indicated that the difference was significant, $t(39) = 2.53$, $p < .016$. There were no assertiveness differences among the other three groups.

Win scores. Although win scores were found unreliable across vignettes in the IMC, they were retained for exploratory purposes. For husbands, the analysis of variance results showed a pattern similar to observed interaction findings. The main effect for locus of control was significant, $F(1, 40) = 7.60$, $p < .009$, with internals winning more disagreements than externals. However, in this analysis, the very low win scores of the external – high trust husband group created a significant interaction effect, $F(1, 40) = 5.34$, $p < .026$. The other three husband groups averaged more IMC wins than their wives, whereas the external – high trust group won an average of only 30% of their disagreements. The difference was highly significant, $t(40) = 4.76$, $p < .001$. There were no significant effects for wives on the win score variable.

Supplementary analyses. Results of analyses of variance employing mixed husband–wife locus of control groups (Husband I–E × Wife I–E) yielded no significant effects for either observed assertiveness or win scores. The other supplementary analysis examined the relationship between individual spouse locus of control and trust scores and the couple's relative assertiveness (husband minus wife). For husbands, the results showed a significant main effect for locus of control, $F(1, 38) = 5.09$, $p < .030$, indicating that internal husbands when compared to external husbands were more assertive relative to their wives. Of the four Husband I–E × Trust groups, only the internal – low trusters had average assertiveness scores higher than their wives. This difference between the internal – low trust group and the other husband groups combined was significant, $t(38) = 2.04$, $p < .049$. There were no significant effects for wives.

Discussion

The main findings of this study may be summarized as follows: (a) For husbands, internals behaved more assertively in the marital conflict than externals did; external – high trusters were least assertive; (b) for wives, internal – low trusters behaved most assertively in the marital conflict. Husband–wife locus-of-control combinations were not associated with differential levels of assertive behavior.

These findings may be interpreted in terms of previous theory and research on locus of control and interpersonal trust and in terms of unconventional role behavior. To begin with, the tendency for internal husbands to behave more assertively than external husbands is consistent with locus-of-control theory. Since they believe they can control events, internals are thought to pursue valued outcomes more vigorously and persistently than externals do, although relatively little research has examined this hypothesis in social interaction situations. Presumably, supporting one's interests in a marital conflict constitutes a generally valued goal that internal husbands seek more assertively than do external husbands. No explanation is readily apparent for the failure to find a main locus-of-control effect for wives, but one may speculate that succeeding in a more-or-less public marital conflict may be more important to husbands than wives.

The tendency for external – high trust husbands to engage in unassertive marital interactions is likewise consistent with previous research on congruent externality (Hochreich, 1974, 1975). According to this research, reviewed earlier, external – high trust males tend to behave in a passive, noncompetitive manner consistent with theoretical assumptions about externality. Believing that an

interpersonally trustworthy environment controls their lives, these men may have little motivation to win or "make points" in marital disagreements.

Although there is previous empirical research on external – high trust males, the internal – low trust combination has not been highlighted in the literature. However, it seems intuitively reasonable that more vigorous assertive behavior should follow from the belief that one's environment is controllable but socially untrustworthy. In the marital context, the present data suggest that the combination of internality and low trust is associated with greater assertive behavior by wives in marital conflict situations. A supplementary analysis found internal – low trust husbands to be more assertive relative to their wives than were the other husband groups.

The second set of interpretations of these findings takes the position that external – high trust husbands and internal – low trust wives represent personality patterns opposite to the conventional stereotypes of husband and wife role behavior. The external – high trust husband does not match the cultural norm of the active male mastering his environment and representing with appropriate caution the family's interests in the world. Being at the opposite end of the conventional masculinity continuum from the "macho" male, the external – high trust husband is fairly unassertive and yielding in arguments with his wife.

Nonnormative in the opposite direction are internal – low trust wives. Instead of assuming a relatively passive and trusting stance toward the world, these women combine qualities of a strong belief in personal control of their lives and a sense of skepticism about other people's intentions. Women with this orientation are apt to relate to their husbands in an assertive manner, arguing vigorously and persistently for their viewpoints.

In sum, a picture emerges of the internal – low trust wife that is exactly opposite that of the external – high trust husband. Whereas he tends to be passive and uninvolved in marital disagreements, she tends to be active and heavily involved. Unfortunately, because of small cell sizes in the present study, couples with an external – high trust husband and an internal – low trust wife could not be compared with other husband–wife matches.

Conclusion

Findings have been reported here for a sample of recently married couples without children. Whether the same results would hold for couples married longer is problematical. In addition to sampling issues, a further note of caution pertains to the small amount of variance accounted for in most of the statistical analyses. Accounting for such low levels of variance, however, is not surprising in light of the notorious difficulty involved in predicting observed behavior from paper-and-pencil personality instruments.

Within these limits, this study suggests that there may still be life in personality and marriage research. For the first time, cognitive personality variables have been related to observed marital interaction behavior. In addition to breaking new ground in personality and marriage research, the present study has identified two interesting types of married persons, external – high trust husbands and internal – low trust wives. Following a strategy suggested by Rotter (1975), it may be useful in the future to employ a measure of locus of control that is specific to marital outcomes. In this way more powerful predictions may be possible concerning some interesting marital behavior patterns and attitudes, and some new understandings may be reached about the role of individual differences in marriage.

Reference Notes

1. Rotter, J. B. *Interpersonal trust, trustworthiness, and gullibility*. Paper presented at meeting of the Eastern Psychological Association, Boston, April 1977.

2. Olson, D. H., & Ryder, R. G. *Marital and family interaction coding system*. Unpublished manuscript, University of Minnesota, 1977.

References

Barry, W. Marriage research and conflict: An integrative review. *Psychological Bulletin*, 1970, *73*, 41–54.

Burchinal, L. G., Hawkes, G. R., & Gardner, B. Personality characteristics and marital satisfaction. *Social Forces*, 1957, *35*, 218–222.

Burgess, E. W., & Cottrell, L. *Promoting success or failure in marriage*. New York: Prentice-Hall, 1939.

Clore, G. L., & Byrne, D. The process of personality interaction. In R. B. Cattell & R. M. Dreger (Eds.), *Handbook of modern personality theory*. New York: Wiley, 1977.

Doherty, W. J. Divorce and belief in internal versus external control of one's life: Data from a national probability sample. *Journal of Divorce*, in press.

Hamsher, J. H., Geller, J. D., & Rotter, J. B. Interpersonal trust, internal–external control, and the Warren Commission Report. *Journal of Personality and Social Psychology*, 1968, *9*, 210–215.

Hochreich, D. J. Defensive externality and attribution of responsibility. *Journal of Personality*, 1974, *42*, 543–557.

Hochreich, D. J. Defensive externality and blame projection following failure. *Journal of Personality and Social Psychology*, 1975, *32*, 540–546.

Laws, J. L. A feminist review of marital adjustment. *Journal of Marriage and the Family*, 1971, *33*, 483–516.

Mirels, H. L. Dimensions of internal versus external control. *Journal of Consulting and Clinical Psychology*, 1970, *34*, 226–228.

Mischel, W. *Personality and assessment*. New York: Wiley, 1968.

Mischel, W. Toward a cognitive social learning reconceptualization of personality. *Psychological Review*, 1973, *80*, 252–283.

Olson, D. H., & Cromwell, R. E. Power in Families. In D. H. Olson & R. E. Cromwell (Eds.), *Power in Families*. Beverly Hills, Calif.: Sage Publications, 1975.

Olson, D. H., & Ryder, R. G. Inventory of Marital Conflicts (IMC): An experimental interaction procedure. *Journal of Marriage and the Family*, 1970, *32*, 443–448.

Phares, E. J. *Locus of control in personality*. Morristown, N.J.: General Learning Press, 1976.

Rich, A. R., & Schroeder, H. E. Research Issues in Assertiveness Training. *Psychological Bulletin*, 1976, *83*, 1081–1096.

Rotter, J. B. Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs*, 1966, *80*(1, Whole No. 609).

Rotter, J. B. A new scale for the measurement of interpersonal trust. *Journal of Personality*, 1967, *35*, 651–665.

Rotter, J. B. Some problems and misconceptions related to the construct of internal versus external control of reinforcement. *Journal of Consulting and Clinical Psychology*, 1975, *43*, 56–67.

Ryder, R. G. Compatibility in marriage. *Psychological Reports*, 1967, *20*, 807–813.

Seeman, M. Empirical alienation studies: An overview. In R. F. Geyer & D. R. Schwietzer (Eds.), *Theories of alienation*. Leiden, The Netherlands: Nijhoff, 1976.

Spanier, G. B. Further evidence of methodological weaknesses in the Locke–Wallace Marital Adjustment Scale and other measures of adjustment. *Journal of Marriage and the Family*, 1972, *34*, 403–404.

Spanier, G. B., & Cole, C. L. Toward clarification and investigation of marital adjustment. *International Journal of Sociology of the Family*, 1976, *6*, 121–146.

Strickland, B. R. Internal–external control of reinforcement. In T. Blass (Ed.), *Personality variables in social behavior*. New York: Wiley, 1977.

Stryker, S. The interactional and situational approaches. In H. T. Christensen (Ed.), *Handbook of marriage and the family*. Chicago: Rand McNally, 1964.

Terman, L. *Psychological factors in marital happiness*. New York: McGraw-Hill, 1938.

Tharp, R. G. Psychological patterning in marriage. *Psychological Bulletin*, 1963, *60*, 97–117.

Weiss, R. L., & Margolin, G. Assessment of marital conflict and accord. In A. Ciminero, K. Calhoun, & H. Adams (Eds.), *Handbook of behavioral assessment*. New York: Wiley, 1977.

Winch, R. F. *Mate selection: A study of complementary needs*. New York: Harper, 1958.

Received November 14, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 12, 2221-2239

An Examination of Self-Perception Mediation of the Foot-in-the-Door Effect

William DeJong
Brandeis University

In 1966, Freedman and Fraser demonstrated that an individual is more likely to comply with a large request for help if that person has previously agreed to an initial small request—a phenomenon they called the "foot-in-the-door" effect. In the present survey, studies that have sought to replicate the foot-in-the-door effect are reviewed. The adequacy of a self-perception explanation for the foot-in-the-door effect is assessed by examining (a) the importance of the size of the initial request; (b) the effect of noncompliance with the initial request; (c) the impact of salient external justifications for the initial act of compliance; (d) the impact of social labels on subsequent levels of compliance; and (e) attempts at actually measuring changes in self-perception. Alternative explanations of the foot-in-the-door effect are considered and rejected, and directions for future research are outlined.

In 1966 Freedman and Fraser tested the notion that once an individual has complied with a small, sometimes trivial request, that person will be more likely to comply with a larger and more substantial demand made in the future—an effect they christened the "foot-in-the-door" phenomenon. During the last 12 years, social psychologists have continued to be fascinated by those results. How is it that the simple act of assenting to a small request for help can dramatically increase, even double, the probability of a person's agreeing to give help in the future?

Because the foot-in-the-door paradigm has become an important vehicle by which to study the link between self-concept and behavior, it is important to take a step back and ask some basic questions about the progress of this research: (a) Is the foot-in-the-door effect a reliable one? Under how wide a range of situations has it been demonstrated? (b) Self-perception theory (Bem, 1972) is most often used to explain the effect. What hypotheses can be derived from this explanation, and what evidence has been gathered for each of them? (c) How else can the phenomenon be explained? What is the evidence for and against the alternative propositions that have been offered? (d) What directions can future research take? It was in the hope of answering these questions that the present article was designed. First, I begin with a review of the Freedman and Fraser (1966) experiment.

The Freedman and Fraser Study

The procedure of their experiment was straightforward. A male experimenter first contacted several suburban housewives in their homes, identifying himself as a member of either the Community Committee for Traffic Safety or the Keep California Beautiful Committee. The initial requests he made of these women were small and innocuous, and almost all of the subjects agreed to them. Half were asked to display a small sign in the front window of their homes. The other sub-

Preparation of this article was supported in part by Public Health Service Postdoctoral Fellowship 1/F32/MH07294-01.

My deepest appreciation is extended to Teresa Amabile, Nancy Cantor, David Funder, Albert Hastorf, Mark Lepper, and David Rosenhan for their comments on an earlier draft of this article.

Requests for reprints should be sent to William DeJong, Department of Psychology, Brandeis University, Waltham, Massachusetts 02154.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3712-2221$00.75

jects were asked to sign a petition advocating certain legislation. These requests were concerned with one of two issues, driving safety or making California more beautiful.

Two weeks later, the housewives were again contacted. A different experimenter, who claimed to be a representative from the Citizens for Safe Driving, asked the women if they would be willing to have a large, ugly billboard reading "Drive Carefully" installed in their front yard for a period of one week. Thus emerged a simple two-way factorial design. For half of the subjects this second request was like the first in terms of the action required of them (i.e., displaying a sign). For those who had previously signed a petition, the action now required of them was quite different. As a cross-dimension, the issue of concern was the same as before for half of the subjects (i.e., driving safety). Those originally approached by the Keep California Beautiful Committee were now dealing with a different issue.

Freedman and Fraser demonstrated a foot-in-the-door effect of remarkable strength and generality. Each of the four experimental conditions produced significantly greater compliance with the final request than did a control condition in which subjects were not approached with a first request: Only 20% of the subjects in that control condition were willing to have the billboard installed, compared with 55% of the experimental subjects. To be sure, those subjects for whom the second request was similar to the first along the two dimensions of issue of concern and mode of action were somewhat more helpful than subjects in the other three experimental conditions, but that difference did not approach statistical significance. This increased compliance effect did not seem to depend much on the type of request that had been made initially. Even the seemingly trivial action of agreeing to sign a petition about the need to "keep California beautiful" more than doubled the probability that a subject would agree to display a billboard that exhorted motorists to drive more safely. The exciting feature of these results is that the effect cannot be explained merely in terms of either the subject's involvement with a particular experimenter or her increased commitment to a certain issue or mode of action. As Freedman and Fraser recognized, a more complicated (and psychologically interesting) explanation is required.

Attempts at Replication

Since Freedman and Fraser's (1966) original report, several replications of the foot-in-the-door effect have been attempted. Before this research is reviewed, the criteria that any study must meet before it can be considered a valid replication attempt should be listed:

1. Obviously, a proper control group, consisting of subjects not receiving an initial request, must be included as part of the experimental design. Three studies must be excluded for this reason: Harris, Liguori, and Joniak (1973); Harris and Samerotte (1975); and Schmidt (1973). The first two, in all fairness, were not explicitly designed to test for a foot-in-the-door effect.

2. The data analysis must include all subjects assigned to the experimental condition and not just those who agree to the first request. Any analysis that excludes those not complying with the first request can be criticized on grounds of differential subject self-selection. Studies by Arbuthnot et al. (1976–1977) and Beaman, Svanum, Manlove, and Hampton (1974) must be excluded from consideration because of this problem.

3. High compliance with the initial request must be obtained. Harris, Liguori, and Stack (1973, Study 3), for example, failed to obtain a significant foot-in-the-door effect, but interpretation of their findings is made problematic by the fact that only 42% of their experimental subjects agreed to the first request (see also, Dutton & Lennox, 1974). Ideally, of course, 100% of the subjects would consent to the first request, but that goal is rarely obtained. Some studies have reported a significant foot-in-the-door effect with a rate of initial compliance as low as 67% (e.g., Freedman & Fraser, 1966, Study 1).

It should be noted that in testing for significant foot-in-the-door effects, each experimental group should be compared individually with the control group. Some studies (e.g.,

Cann, Sherman, & Elkes, 1975; Harris, 1972, Study 1) combined experimental groups in their analyses, and the data from such studies have been reanalyzed when possible.

Table 1 summarizes known replications of the foot-in-the-door phenomenon that had been attempted by the time of this review (November 1978). The table lists the two requests made of the subjects, the rate of compliance with the second request for both the experimental groups and the no-initial-request control group, and the conclusions drawn from appropriate statistical comparisons.

A glance at the table shows that a significant foot-in-the-door effect was not always found. It is unwarranted, however, to declare that the effect cannot be reliably obtained. It must be noted that only a few studies (Cann, 1976; Cialdini & Ascani, 1976; Cialdini, Cacioppo, Basset, & Miller, 1978; DeJong & Funder, 1977, Study 1; Harris & Samerotte, 1976; Miller & Suls, 1977; Tipton & Browning, 1972; DeJong, Note 2) showed the percentage of compliance with the second request for one or more experimental groups to be lower than or equal to that for the control group. Thus, almost all of the failures to replicate were in the predicted direction, but did not reach traditional levels of statistical significance. And as noted previously, two failures to replicate (Dutton & Lennox, 1974; Harris, Liguori, & Stack, 1973, Study 3) are difficult to interpret because of the very low levels of compliance with the first request in the experimental group.

The subsequent discussion of many of these studies reveals other potential explanations for failures to replicate. They can be mentioned here briefly. First, the initial request must be large enough to cause people to think about the implications of their own behavior; if the request is too small, the effect will probably not be obtained (Seligman, Bush, & Kirsch, 1976). Thus, for example, it may be that a request to take a small card is too small to increase compliance with a high-cost request like donating blood (Cialdini & Ascani, 1976). Moreover, some evidence suggests that if the initial request is too large, the probability of obtaining the effect will be reduced (Baron, 1973; Miller & Suls, 1977). Second, people must feel that their initial compliance resulted from the exercise of free choice, not because of pressure to comply (Uranowitz, 1975; Zuckerman, Lazzaro, & Waldgeir, in press). Fish and Kaplan's (1974) findings are difficult to interpret for this reason. Subjects in their experiment were first asked to write an essay on various ways of fighting poverty, but they were asked to do this as part of a class about the poverty program. It is possible, then, that the subjects may have felt pressured into writing the essay because they saw it as a class assignment.

A key issue of concern is the generality of the foot-in-the-door phenomenon. Under how wide a range of situations has it been demonstrated? As Table 1 reveals, most of the requests that have been made of subjects involved little time or effort. But some did involve major commitments on the part of the subjects. For example, Lowman (1973) used the foot-in-the-door technique to increase household participation in a complicated trash recycling program. In one study, Freedman and Fraser (1966, Study 1) induced housewives to allow a survey team of five or six men to come into their homes to classify the household products they used.

The length of the delay between the two requests has also been varied. The foot-in-the-door effect has been demonstrated with delays between the two requests of up to two weeks (e.g., Freedman & Fraser, 1966, Study 2; Harris, 1972, Study 2), but most successful replications have involved much smaller delays of a day or two. The effect of delay has never been systematically explored, save Cann et al.'s (1975, Study 1) comparison between delay (7–10 days) and no-delay procedures (see Table 1). The length of the delay itself may be less important than whether the occasion of the second request somehow reminds people of their earlier behavior.

The degree of similarity between the two requests is an important dimension as well. Unfortunately, most investigators have employed requests that share many common features, so that the robustness of the foot-in-the-door phenomenon has rarely been put to

Table 1
Replications of the Foot-in-the-Door Effect

Study	First request	Second request	Results[a]	Conclusion
Baer, Goldman, & Juhnke (1977)	(1) Give the time to experimenter (2) Same as (1); later misinformation given to different experimenter	Correct misinformation given to experimenter by another elevator passenger	(1) 70 (2) 35 (C) 33	FITD No FITD
Baron (1973)	(1) Accept leaflet on the dangers of pollution (2) Sign antipollution petition, get two friends to sign, and mail in	Agree to put 3 foot × 5 foot antipollution sign in front yard	(1) 50[b] (2) ? (C) 20	FITD No FITD
Cann (1976)	(1) Agree to receive a questionnaire on recycling in the mail and fill it out	Volunteer time for a neighborhood cleanup project	(1) 26 (C) 26	No FITD
Cann, Sherman, & Elkes (1975, Study 1)	(1) Answer three questions on driving habits; no delay between the two requests (2) Same as (1); 7–10-day delay between requests	Agree to accept 15 pamphlets on traffic safety and distribute to neighbors	(1) 78 (2) 70 (C) 50	FITD No FITD
Cann, Sherman, & Elkes (1975, Study 2)	(1) Answer three questions on driving habits; no delay between the two requests	Agree to accept 15 pamphlets on traffic safety and distribute to neighbors	(1) 72 (C) 45	No FITD
Cialdini & Ascani (1976)	(1) Take and display small card advertising blood drive	Agree to donate blood the next day	(1) 32 (C) 32	No FITD
Cialdini, Cacioppo, Basset, & Miller (1978, Study 2)	(1) Agree to display a United Way window poster	Agree to pick up a United Way poster packet at dormitory lobby	(1) 70[c] (C) 70	No FITD
Crano & Sivacek (Note 1, Study 1)	(1) Answer 10 questions about beverages	Agree to answer 30 questions on driving habits	(1) 66 (C) 31	FITD
Crano & Sivacek (Note 1, Study 2)	(1) Answer 10 questions on household products	Agree to answer 45 questions on the mass media	(1) 56 (C) 32	No FITD
DeJong (Note 2)	(1) Agree to sign a petition for pro-disabled legislation; learn they were one of many to sign (2) Same as (1); learn they were the first to sign	Notify a second experimenter that he dropped a quarter	(1) 64 (2) 28 (C) 32	FITD No FITD
DeJong & Funder (1977, Study 1)	(1) Answer 15 questions on the quality of life in the local community	Agree to answer 50 questions on highway safety	(1) 46 (C) 56	No FITD

Table 1 (*continued*)

Study	First request	Second request	Results[a]	Conclusion
DeJong & Funder (1977, Study 2)	(1) Answer 15 questions on quality of life in the local community; receive letter acknowledging participation	Agree to answer 50 questions on highway safety	(1) 66 (C) 56	No FITD
DeJong & Musilli (Note 3)	(1) Agree to participate in a 5-minute survey on parking facilities for compact cars; experimenter appeared physically normal (2) Agree to participate in a 5-minute survey on parking facilities for disabled drivers; experimenter appeared physically normal	Agree to participate in a 30-minute telephone survey on highway laws and driving hazards	(1) 55 (2) 53 (C) 40	FITD No FITD
Dutton & Lennox (1974)	(1) Give money to a white panhandler	Agree to donate time to various activities as part of an interracial Brotherhood Week	(1) 54.8[d] (C) 46.2	No FITD
Fish & Kaplan (1974)	(1) Write a short essay on ways of fighting poverty	Volunteer time and services to a welfare agency	(1) 36 (C) 33	No FITD
Freedman & Fraser (1966, Study 1)	(1) Answer eight questions on household soaps (2) Agree to be in survey on households soaps	Agree to allow six-man survey team to enter home and spend 2 hours classifying all household products	(1) 53 (2) 33 (C) 22	FITD No FITD
Harris (1972, Study 1)	(1) Give directions (2) Give the time	Give the experimenter a dime	(1) 39 (2) 44 (C) 11	No FITD FITD
Harris (1972, Study 2)	(1) Write a letter to a minority high school student, indicating willingness to answer questions about the university and student life	Sign class list to volunteer time to a university publicity campaign	(1) 18 (C) 9	FITD
Harris, Liguori, & Stack (1973, Study 3)	(1) Allow name to be sent to local congressman as supporter of organization's programs	Agree to donate money or cookies to a fund-raising baked-cookie sale	(1) 30 (C) 25	No FITD
Harris & Samerotte (1976, Study 1)	(1) Watch experimenter's possessions; a theft attempt is later thwarted by the subject (2) Same as (1); the second request is made by a different experimenter	Give money to experimenter to permit the purchase of food	(1) 20[e] (2) 40 (C) 35	No FITD No FITD

(*table continued*)

Table 1 (*continued*)

Study	First request	Second request	Results*	Conclusion
Harris & Samerotte (1976, Study 2)	(1) Watch experimenter's possessions; a theft attempt is later thwarted by the subjects (2) Same as (1); the second request is made by a different experimenter (3) Watch experimenter's possessions; no theft attempt is made (4) Same as (3); the second request is made by a different experimenter	Give money to experimenter to permit photocopying of an article	(1) 14 (2) 29 (3) 50 (4) 21 (C) 29	No FITD No FITD No FITD No FITD
Lowman (1973)	(1) Answer four questions on recycling and container use	Agree to participate in a glass and metal trash recycling program	(1) 85 (C) 68	FITD
Miller & Suls (1977)	(1) Give directions that are difficult to explain (2) Give directions that are simple to explain	Help a male experimenter pick up dropped groceries	(1) 12 (2) 64 (C) 32	No FITD FITD
Pliner, Hart, Kohl, & Saari (1974)	(1) Wear pin to advertise a fund drive (2) Wear pin and persuade member of family to do so	Contribute money to the fund	(1) 74 (2) 81 (C) 46	FITD FITD
Reingen & Kernan (1977)	(1) Agree to participate in a 5-question survey on household products	Agree to participate in a 20-question survey on household products	(1) 75 (C) 58	No FITD
Seligman, Bush, & Kirsch (1976)	(1) Answer 5 questions on the energy crisis and inflation (2) Answer 20 questions (3) Answer 30 questions (4) Answer 45 questions	Agree to answer 55 more questions for the same survey	(1) 38 (2) 35 (3) 74 (4) 74 (C) 31	No FITD No FITD FITD FITD
Seligman, Miller, Goldberg, Gelberd, Clark, & Bush (1976)	(1) Listen to a 2-minute, pro-McGovern speech; agree to display a small campaign sign (2) Agree to display a small campaign sign only (3) Listen to a 2-minute speech on fire prevention, agree to display a small fire prevention sign (4) Agree to display a small fire prevention sign only	Agree to display a McGovern poster in front window	(1) 38 (2) 23 (3) 30 (4) 31 (C) 16	FITD No FITD No FITD No FITD
Snyder & Cunningham (1975)	(1) Agree to answer 8 questions on household paper products or on traffic safety	Agree to answer 30 questions for the other organization	(1) 52 (C) 33	No FITD

Table 1 (*continued*)

Study	First request	Second request	Results[a]	Conclusion
Tipton & Browning (1972)	(1) Help an elderly woman pick up dropped groceries	Help a young woman in a wheelchair up over a curb	(1) 0[f] (C) 36	No FITD
Uranowitz (1975)	(1) Watch experimenter's shopping bags while he retrieves a dollar bill (2) Watch experimenter's shopping bags while he retrieves his wallet	Notify a second experimenter that she dropped her package	(1) 80 (2) 45 (C) 35	FITD No FITD
Zuckerman, Lazzaro, & Waldgeir (in press)	(1) Agree to participate in a 5-minute survey on traffic safety	Agree to participate in a 20-minute survey on household products	(1) 64 (C) 45	No FITD

Note. The first condition in a study is indicated by (1), the second condition by (2), and so on; results are described accordingly; (C) = the no-initial-request control group; FITD = the foot-in-the-door effect.

[a] Results are reported as the percentage of subjects complying with the second request. A chi-square test (corrected for continuity) was executed for comparisons between individual experimental groups and the control group. If $p < .10$, it is concluded that the foot-in-the-door effect was successfully replicated. If a chi-square could not be executed based on the information provided by the authors, the conclusions based on the authors' own statistical comparisons are reported.

[b] Data reported are for subjects run by the male experimenters. Percentage compliance for subjects run by the female experimenters is not reported, but no FITD effect was found. For Condition 2 (male experimenters), $p > .10$ by chi-square, according to Baron (1973).

[c] Data reported are for verbal compliance with the request. Actual compliance was as follows: (1) 10%; (C) 20%.

[d] Data are reported in terms of average summed difficulty level of volunteered activities; data on percentage compliance are not reported. In a previous experimental session, all subjects run in this study were given false physiological feedback, which led them to believe they might harbor racial prejudice.

[e] Data have been collapsed across the variable of type of food to be purchased by the experimenter.

[f] Only those who helped the elderly woman were included as subjects in the experimental condition. Despite the bias toward confirmation this procedure created, no foot-in-the-door effect was found.

a severe test. For example, most studies that ask the subjects to participate in a survey as the initial request also seek their participation in a second survey. As noted previously, Freedman and Fraser (1966, Study 2) actually manipulated the degree of similarity between the two requests in terms of the issue involved and the mode of action required of subjects, but they found those factors to have little impact on compliance with the second request. Seligman, Miller, Goldberg, Gelberd, Clark, and Bush (1976) also failed to find generally higher compliance when the two requests were similar.

Self-Perception Explanation

Freedman and Fraser (1966) were surprised to learn that the size of the foot-in-the-door effect they obtained did not depend on the nature of the first request. As noted before, possible explanations for the effect centering around subjects' commitment to a particular experimenter, organization, issue, or mode of action were effectively ruled out. To account for their unexpected results, they offered the following explanation:

> What may occur is a change in the person's feelings about getting involved or about taking action. Once he has agreed to a request, his attitude may change. He may become, in his own eyes, the kind of person who does this sort of thing, who agrees to requests made by strangers, who takes action on things he believes in, who cooperates with good causes. (Freedman & Fraser, 1966, p. 201)

Thus, the self-perception mediation of the foot-in-the-door phenomenon involves a two-stage process. First, people must observe their own behavior and the situational context in

which it occurs and, from those data, make an inference about their own dispositions and attitudes. It has been argued that subjects comply with the initial request in the apparent absence of any external inducements and then decide on the basis of that evidence that they are the kind of people who cooperate with good causes or help out other people.[1] Second, this changed self-perception is thought to increase the probability of their performance of similar kinds of action in the future. As presently articulated, self-perception theory does not adequately explain the processes that mediate this last step (cf. Bem, 1972).

The self-perception explanation of the foot-in-the-door phenomenon has led to the formulation of several testable hypotheses. The evidence for each of these is now considered.

Size of the Initial Request

A straightforward prediction derived from self-perception theory is that the larger the size of the first request to which people agree, the greater the probability of their compliance to the second larger request. In other words, the greater the costs of compliance with the initial request, the more likely it should be for people to find dispositional meaning in their behavior. To test this hypothesis, Seligman, Bush, and Kirsch (1976) first asked subjects to answer either 5, 20, 30, or 45 short questions for a survey on reactions to the energy crisis and inflation. There were no differences among those four conditions with regard to compliance with the first request for help. Later, subjects were called back by a different experimenter representing the same survey group and were asked to answer 55 more questions. Unexpectedly, subjects in the 5- and 20-question groups were not more likely to answer the 55 questions than were control subjects contacted for the first time. However, the two larger initial requests were effective in inducing significantly higher rates of compliance with the second request.

Seligman, Bush, and Kirsch recognized the possibility that sufficiently large initial requests might discourage subsequent compliance, but evidence for that was not found in their study. It should be noted that subjects in all cases were told the survey would require only "a couple of minutes," thus making even the larger requests seem relatively small. A study by Miller and Suls (1977) suggests the possibility of a curvilinear relationship between size of the initial request and subsequent compliance. Subjects in their experiment were asked to give street directions to a first experimenter, the directions being either difficult or simple to explain. Compared with control group subjects, those who gave the simple directions were more likely to help a second experimenter pick up his dropped groceries a few moments later. Those who gave the difficult directions were less likely to help. Existence of a curvilinear relationship, of course, could not be explained by self-perception theory alone.

Agreement to the Request Versus Actual Performance

It could also be predicted from self-perception theory that compliance with the second request would be more likely if subjects actually performed what was asked of them in the initial request than if they agreed to comply but were never called upon to carry out their promise. The only study that directly addressed this issue is an experiment reported by Freedman and Fraser (1966, Study 1). Housewives in the so-called agree-only condition were asked to participate in a consumer survey about household products; those willing to help were told the survey would be conducted at a later time. Subjects in the performance condition actually participated the survey. In a later call, the same experimenter asked subjects if they would permit

[1] Self-perception theory has not articulated well the exact nature of the self-perception changes that are said to occur. It is not clear, in the case of most foot-in-the-door studies, whether subjects classify their behavior as *compliance* or as *helping*. This uncertainty is due to the fact that most experimenters use compliance with a request for help as their dependent measure (see Table 1). Experiments that instead require subjects to initiate help giving (e.g., Miller & Suls, 1977; Uranowitz, 1975) suggest most subjects may view their assent to the request as helpfulness.

urvey team to come into their homes to list ne household products they used. Whereas 22% of the control group subjects agreed to hat request, over half the subjects in the performance condition did so. In contrast, a mere 33% of the agree-only condition subjects complied with that request. A clear heoretical interpretation of these results is made difficult by the fact that subjects in the agree-only condition did not learn the exact size of the request to which they had agreed. Their compliance to the second request may have been lower only because they guessed the initial request to be quite large (cf. Miller & Suls, 1977). Parenthetically, it must be noted that other experimenters have shown that actual performance of the initial request is not requisite in demonstrating the foot-in-the-door effect (e.g., Snyder & Cunningham, 1975; Zuckerman et al., in press).

Effect of Noncompliance With the Initial Request

Self-perception theory led Snyder and Cunningham (1975) to predict that subjects induced not to comply with an initial request would be less likely to comply with a second request. In their experiment, one group of subjects was first asked to participate in a 50-question telephone survey, a request sufficiently large to guarantee almost universal noncompliance. Two days later, a different-sexed experimenter representing a second organization asked subjects to answer 30 questions. Consistent with their prediction, subjects initially approached with the large request were less compliant with the second request than were subjects in the control group. Similar results were reported by Cann et al. (1975) and Reingen and Kernan (1977). These studies support the idea that induction of initial noncompliance leads people to perceive themselves as the kind of people who do not comply with such requests. One might predict that the smaller the request that people refuse, the lower the probability of their compliance to a second request; this proposition remains to be tested. An added complication arises when the timing of the second request is considered. In two studies, Cann et al. (1975) reported that noncompliance with a large initial request actually led to greater compliance when the second request was made immediately after the first, rather than after a delay period of several days. For example, one study showed that 90% of the subjects agreed to an immediate second request after having refused the experimenter's initial large request, compared with 50% in the control group. Only 29% compliance was obtained when that second request was made 7–10 days later.

These findings replicate an effect first demonstrated by Cialdini et al. (1975), which they dubbed the "door-in-the-face" technique. Quite simply, these authors claimed that if an experimenter first approaches subjects with an unreasonable request that is sure to be refused, but then immediately asks a smaller favor, subjects will view the experimenter's action as a concession. The subjects, in turn, will feel normative pressure to reciprocate that concession, and they will respond to that pressure with compliance with the second request. Replications of this effect abound (Cialdini & Ascani, 1976; Miller, 1974; Miller, Seligman, Clark, & Bush, 1976; Reingen, 1977).

In one experiment, Cialdini et al. (1975) demonstrated that the subjects' perception of the encounter as a kind of bargaining situation probably underlay the effect. In the rejection–moderation condition, both requests for help were made by the same experimenter. In the so-called two-requester condition, subjects' noncompliance with the first request was followed immediately by another request put to them by a second experimenter. This second experimenter was seemingly unrelated to the first, but subjects knew this person had overheard their conversation with the first experimenter. Whereas a control procedure produced a 31% compliance rate, 55% of the rejection–moderation condition subjects consented to the second request. But when that same second request was made by a different experimenter, only 10% of the subjects in the two-requester condition complied, a result consistent with those reported by Snyder and Cunningham (1975). The authors argued that subjects in the latter condition could

not interpret the second request as a concession, simply because it was made by a different experimenter.[2]

Why do subjects in the door-in-the-face procedure who do not comply with the initial large request not come to see themselves as the kind of people who refuse such requests, as Snyder and Cunningham's (1975) results might lead us to expect? Even-Chen, Yinon, and Bizman (1978) argued that the initial requests used by Cialdini et al. (1975) and others were too large for subjects to draw dispositional inferences from their failure to comply. Implicit in their argument is the suggestion that subjects must believe the request is one to which someone (though not they) might consent in order for a self-perception change to occur. In their study, when subjects refused a large, but not reasonable, first request, they were less likely to agree to a second smaller request. Only those who refused an extremely large request showed a subsequent door-in-the-face effect. These results are suggestive, but Even-Chen et al.'s argument cannot account for the results obtained by Cialdini et al. In the earlier experiment, whether a door-in-the-face effect or a result consistent with self-perception theory was found depended solely on who made the second request.

In general, research on the effect of initial noncompliance has been supportive of the self-perception hypothesis. Only under a special set of circumstances, that is, when the first request is extremely large and the second request can be viewed as a concession on the part of the experimenter, does an initial refusal not lead to subsequent refusals.

Effect of Consensus Information

The self-perception analysis of the foot-in-the-door effect leads to the prediction that if subjects were informed they were not unique in their compliance with the initial request, they would be induced to see their behavior as determined by situational pressures and not to view it as having any implications for their own traits or attitudes (cf. Cook, Pallak, & Sogin, 1976; Cooper, Jones, & Tuller, 1972). In one recent experiment designed to test this prediction (DeJong, Note 2), shoppers were approached individually by a physically able male experimenter and shown a card on which was written the resolution of a petition formulated by a group called the Upper Valley Foundation to Help the Handicapped. After reading the resolution, 87% of the subjects agreed to sign the petition. At that point, some subjects were shown a petition with many signatures and one blank space at the bottom (consensus condition). As they signed, the experimenter said, "As you can see, almost everyone we've asked has signed our petition." The other subjects were shown a completely blank petition, and as they signed, the experimenter noted that they were the only ones so far to do so (nonconsensus condition). When the subjects walked on, a second male experimenter dropped a quarter as he walked some 15 to 20 feet ahead of them. Thirty-two percent of the subjects in the no-initial-request control group notified the second experimenter of his loss, whereas 64% of those told that many people had signed the petition did so. Unexpectedly, only 28% of those told they had been unique in their agreement to sign helped the second experimenter.

Consensus information is more psychologically complex than attribution theorists have admitted (cf. McArthur, 1973). In addition to pointing to the power of situational constraints on behavior (Kelley, 1967), consensus information can help define the behavioral norms that are operative in a specific context. Subjects in the consensus condition may have helped the second experimenter more often because they were reminded that helping others was appropriate behavior. In addition,

---

[2] Pendleton and Batson (1979) offered a self-presentation explanation of the door-in-the-face technique, suggesting that subjects are motivated to comply with the second request to avoid being perceived as unhelpful. Although the results for Cialdini et al.'s (1975) two-requester control group seem to counter this alternative, Pendleton and Batson argued that the phrasing of the second experimenter's request ("But maybe you could help *me*") may have induced reactance, thus producing a low rate of compliance in that condition. Further research is needed to resolve this controversy.

consensus information can reaffirm the reasonableness of one's own behavior. Subjects in the nonconsensus condition may have been shocked to learn of their unique status and may literally have been "lost in thought" when the second experimenter dropped his quarter. To date, no foot-in-the-door study involving the introduction of consensus information has yielded results consistent with self-perception theory.

Presence of External Justification

Theoretical analyses of attributional processes (Kelley, 1967) propose that people assign dispositional meaning to behavior after a careful assessment of the possible explanatory power of controlling influences in the environment. Thus, a self-perception account of the foot-in-the-door phenomenon stresses that the amount of external pressure used to induce compliance with the initial request is paramount. If those external pressures provide people with an adequate explanation of their behavior, they will not be led to infer anything about their own traits or attitudes. Under those circumstances, an increased probability of compliance with subsequent requests for help would not be expected. Indeed, Lepper (1973) even suggested the possibility of an "overjustification" effect; that is, if people comply with an initial request under conditions of strong external justification, they may come to infer that they are intrinsically motivated and that they are the kind of people who agree to such requests only when external pressures are present. As a result of this kind of changed self-perception, people may actually come to be less likely to comply with future requests when those kinds of pressures are absent.

The impact of external justifications for the initial act of compliance on subsequent helping was demonstrated by Uranowitz (1975). Female shoppers were asked by an experimenter to watch his grocery bags while he ran back to retrieve a lost article. In the high-justification condition, he claimed he had lost a wallet containing a lot of money; in the low-justification condition he said that he had dropped only a dollar bill. The assumption, of course, was that subjects in the high-justification condition would believe that circumstances demanded their compliance and that anyone faced with such a request would agree to it. When the first experimenter returned with the lost article, the subject proceeded on her way. A female experimenter later dropped a package in the subject's path, and the subject's response was noted. The results were striking. Eighty percent of the subjects in the low-justification condition helped the second experimenter, whereas only 45% of the high-justification and 35% of the control subjects did so.

Zuckerman et al. (in press) investigated this same problem using an experimental procedure modeled after that used by Snyder and Cunningham (1975). Housewives were first asked to take part in a 5-minute telephone survey; some were promised a monetary payment in exchange for their cooperation, and others were not. If a subject agreed to participate, she was told that the interview would be conducted at a later time. Subjects promised the monetary payment were told they would receive a check after the interview. Two or three days later, the subjects were called by a second experimenter representing a different service organization and were asked to consent to a 20-minute interview. Forty-five percent of the control group subjects agreed to that request, whereas 64% of the subjects promised no money consented to the interview. In contrast, only 33% of those promised a monetary payment in return for their initial compliance agreed to the second request. Similar results were reported by Reingen and Kernan (1977).

DeJong and Funder (1977), however, found the monetary payment to have the opposite effect. The day after their participation in a 15-question survey, subjects in the payment condition received the $2 that had been promised them in exchange for their help. Subjects were later asked to take part in a 50-question survey being conducted by a second organization. Fifty-six percent of control subjects never before contacted agreed to the second request, whereas 78% of the subjects in the payment condition did so. These findings were unexpected. It was thought that the

actual receipt of money by subjects in the payment condition would increase the salience of the external justification for their compliance with the first request. A follow-up study showed that this finding was not due to subjects' expectation that they would be paid for helping with the second survey. When the second caller informed subjects they could not be paid for their help, the pattern of results was virtually unchanged. One possible explanation for this unexpected set of results is that a letter that accompanied the payment may have labeled subjects as "doers" (cf. Kraut, 1973). That certification may have had greater implications for subjects' self-perceptions than did the small monetary payment.

Finally, in a study conducted by DeJong and Musilli (Note 3), subjects were approached at home with a small initial request made by a female experimenter who appeared to be either handicapped or physically normal. It was hypothesized that subjects approached by the disabled experimenter would feel psychological pressure to comply with her request and would, in turn, attribute their compliance to the fact of her disability. The experimenter asked half the subjects to answer questions for a survey on special parking facilities for disabled drivers for a group called Friends of the Handicapped. The others were told that the experimenter represented Friends of the Environment and that the questions concerned special parking for compact cars. Two days later, subjects were called on the telephone by a second experimenter and were asked to participate in a 30-minute survey on traffic safety for a different organization. The hypothesis was partially supported. Fifty-six percent of the subjects approached by the handicapped experimenter for Friends of the Handicapped complied with the second request, whereas 40% of control subjects did so. However, only 41% of subjects approached by the handicapped experimenter for Friends of the Environment complied with the second request.

In sum, evidence is generally consistent with the proposition that high external justification for initial acts of compliance can dampen the probability of subsequent compliance. Only one study found no such effect at all (DeJong & Funder, 1977), and there is a plausible explanation for that failure. Little evidence for an overjustification effect exists; only Reingen and Kernan (1977) and Zuckerman et al. (in press) reported results for a high-justification condition to be actually below those for a control group, and these differences were small. One possible reason for not finding overjustification effects is that subjects in these studies were not preselected for their high level of initial intrinsic or altruistic motivation.

There may be circumstances under which high external justification for initial compliance actually leads to higher levels of subsequent compliance. Such an outcome might be anticipated if people believe that behavior that is truly intrinsically motivated is more highly regarded than that which is thought to arise in response to external pressure (cf. Nemeth, 1970). Thus, when subjects' self-esteem (or their good name) is threatened by the perception that they have complied in response to the dictates of external contingencies, they may be motivated to be more helpful in the future when such pressures are absent. A report by Upton (1973) is consistent with this notion. He found that previous blood donors classified as intrinsically motivated did not respond well to an appeal for donations that offered money in exchange for a pint of blood. When no such bribe was offered, the percentage of such subjects willing to donate was almost 30 percentage points higher. It is conceivable that subjects offered the bribe wished to avoid the perception that they were motivated to help others only as a means of garnering rewards for themselves.

Effect of Social Labels

Sociological theories of deviance stress that society encourages those whom it labels as deviant to learn and accept a deviant role identity. Once such people come to share this definition of themselves, the changed self-image is believed to sustain their deviant behavior (see Schur, 1971). Although these theories focus almost exclusively on labels of deviance and generally ignore positively val-

enced labels, their basic propositions can still easily be translated into the language of self-perception theory. Though actual behavior (and the context in which it occurs) may provide the clearest evidence about one's traits or attitudes, self-perception theory recognizes that a self-definition or label provided by others may serve as an important source of information about one's dispositions (cf. Miller, Brickman, & Bolen, 1975) or at least may signal to people that their behavior and its implications for their self-image should be reflected on, thereby energizing a self-perception analysis.

This proposition was put to the test by Kraut (1973). First, a male experimenter went to subjects' homes and solicited contributions for the Heart Association. If subjects made a donation, they were randomly assigned to be in a labeled or a nonlabeled condition. Subjects receiving the charitable label were simply told, "You are a generous person. I wish more of the people I met were as charitable as you." Subjects who did not contribute were also assigned to a labeled or a nonlabeled condition. Subjects receiving the label were frankly told that they were "uncharitable." One or two weeks later, a second experimenter contacted the subjects and asked for a donation to a fund-raising drive for multiple sclerosis. (Unfortunately, a control group consisting of subjects receiving only the second solicitation was not run.) As Kraut predicted, 62% of those given the positive label donated the second time, whereas only 47% of the nonlabeled subjects did so. Kraut recognized that this effect could be explained in terms of a social reinforcement model if the charitable label constituted a reward for the subjects. He also found that those who did not donate the first time gave somewhat less when they had been branded as uncharitable by the first experimenter, but this difference was not significant.

The impact of labels on self-perception and subsequent behavior was also examined in a study conducted by Paulhus, Shaffer, and Downing (1977). Prior to their actual donations, blood donors were asked to read communications that were designed to label their motives for giving blood. Half the subjects read a communication that emphasized altruistic motives, and as a cross-dimension, half the subjects were told about the personal benefits they would receive (e.g., free blood in case of emergency). After donating, subjects filled out a questionnaire that included an item on their plans for giving again within the next year. The results showed a main effect for salience of altruistic motives, such that subjects led to feel they had acted altruistically reported a greater likelihood of future donation.

McArthur, Kiesler, and Cook (1969) labeled some subjects as doers (people who know what needs to be done and then take the appropriate action) on the basis of bogus test results. This positive label increased the proportion of subjects willing to help distribute antipollution leaflets only when subjects had been told that their doer personality qualified them to be paid for their participation in a second experiment. Apparently, subjects' belief that being a doer was a saleable quality strengthened the impact of the label, either because it made the feedback more convincing or simply because the subjects thought about it more.

In contrast, Steele (1975) expected labels to have effects different from those predicted by self-perception theory. He argued that a negative label (name-calling) that impugned subjects' character would motivate them to take action that would restore their self-esteem, thus increasing the probability that they would comply with a later request for help. Because positive labels would not do this, Steele expected them to have little impact on help giving. In this study, an experimenter claiming to be a pollster contacted subjects by telephone. In the relevant-negative-name condition, subjects were lambasted for being "apathetic about the welfare of others," whereas subjects in the irrelevant-negative-name condition were criticized for their lack of concern for driving safety. Recipients of the relevant-positive-name message were praised for their desire to help their fellow man. (No irrelevant-positive-name condition was run.) Two days later, subjects were called by a second experimenter representing a food cooperative in a lower income

neighborhood. To help the cooperative achieve its goal of "aiding the less fortunate," subjects were asked to compile a detailed list of the quantities and brands of foods and household items used in their homes. Compared with a control group never before contacted, subjects receiving the positive label were more likely to help, though not significantly so (cf. Kraut, 1973). On the other hand, virtually all subjects in both negative-name conditions promised to help out, and most actually did so. This effect was replicated in a second study.

The name-calling procedure employed by Steele (1975) differs from the labeling procedure used by other researchers in one important way. The labels used by others were based either on recent behavioral evidence (Kraut, 1973) or on the results of phony psychological tests (McArthur et al., 1969). In contrast, no such evidence substantiated the claim made by Steele's experimenter. This suggests that whether the negative label induces a change in self-perception may depend on the degree to which subjects' future behavior can belie the label. If refutation of the label is still possible, it may not lead to a change in self-definition, but may lead to a vigorous effort to defend self-esteem or to assuage guilt.

This possibility was explored in a recent study by Gurwitz and Topol (1978). Students at a large suburban university were first contacted by telephone and accused of not taking advantage of the opportunities afforded by a nearby city. This accusation was directed at subjects as members of a group (students at the university) or as individuals. Before the accusation was made, half the subjects were asked how many times in the past few months they had gone into the city, the modal answer being zero. The other subjects were not led in this way to provide evidence in support of the accusation. Later that evening, a second experimenter had subjects fill out a questionnaire concerning their interest in a number of student-organized activities in the city.

When subjects had not been led to provide evidence in support of the accusation, the degree to which they later belied the label depended on whether the accusation had been made about them as individuals or as members of a group. When they were seen as part of a group, they indicated more interest in the activities, thus belying the label. In contrast, when the subjects had been led to provide such evidence prior to the accusation, they were more likely to disconfirm the label when it had been directed at them as individuals. Similar results were found in a laboratory study conducted by Gurwitz and Topol (1978) in which subjects were labeled as being low in self-confidence. Why subjects' responses to the label depended so greatly on whether it was applied to them individually or as members of a group is not clear.

In sum, although the effect of positive labels has been fairly consistent across several studies, the impact of negative labels on subsequent behavior presents a more complicated picture.

Measurement of Altered Self-Perceptions

Perhaps the most convincing evidence in support of the self-perception explanation of the foot-in-the-door effect would come from studies that actually measured the changes in self-perception that are thought to occur. Unfortunately, this kind of direct evidence is hard to come by. A long catalog of excuses for this failure can be offered, most of them pointing to possible inadequacies in design and measurement (see Bem, 1972; Lepper, 1973). But the major difficulty may be that the self-perception changes that follow a person's initial compliance with a small request have never been clearly specified and may be much more complicated than has been acknowledged.

First, it may be unreasonable to expect the subjects' general perception of themselves to be altered by the kinds of brief experiences involved in these foot-in-the-door studies. In fact, particular subjects may only come to infer something about their attitudes or motives for a fairly limited set of situations (e.g., telephone solicitations from strangers). This notion is consistent with recent statements of self-concept theory (Gergen, 1971;

McGuire & Padawer-Singer, 1976), which emphasize that a person may harbor a variety of self-definitions that differ in salience across various situation contexts.

Second, measurement of these supposed changes in self-perception is made difficult by the fact that individuals are likely to code their behavior along different dimensions or into different groupings (Bem & Allen, 1974). Which behaviors constitute actions that are psychologically similar to the initial act of compliance is a function of each person's view of the world. To give one example, some persons might code their behavior as *compliance* or as *participating in a survey,* whereas others might code it as *helping someone in need.*

For these two reasons, investigators who have hoped to show a general change in subjects' self-definitions seem to have had little chance of success. Still, two experiments have successfully demonstrated that extrinsic incentives for help giving can lead helpers to describe themselves as less altruistically motivated. During the course of an experiment on "first impressions," Batson, Coke, Jasnoski, and Hanson (1978) asked male undergraduates to help an experimenter code data. Payment for this help was offered before subjects agreed to help (payment condition), after the subjects' agreement (payment after), or payment was not mentioned at all (no payment). The request for help was always made in the presence of a male confederate who never volunteered his services. A control group received no such request. Subjects were then asked to rate themselves and the confederate on several dimensions, including helpfulness and cooperativeness. As expected, subjects who agreed to aid the experimenter in exchange for money rated themselves as less altruistic than the confederate, whereas subjects in the other three conditions rated themselves as more altruistic.

In an experiment conducted by Smith, Gelfand, Hartmann, and Partlow (1979), second- and third-grade children played a marble drop game in which they could earn pennies toward the purchase of a prize. During the first part of the experiment, the children were given a number of opportunities, signaled by a light, to donate pennies to another child playing the game in a nearby room. The experimenter also created situations in which the children could not make a donation (but would think they should have done so) by having the signal light go off before they could respond. Whereas some subjects were merely praised each time they donated a penny, others were given a monetary reward for doing so, and the relationship between their help giving and the reward was spelled out to them. Other subjects were scolded by the experimenter each time they did not avail themselves of an opportunity to help; another group of children were fined each time they failed to donate, and the contingency between their behavior and the punishment was made explicit. Finally, a control group did not receive any type of reward or punishment from the experimenter. Intensive interviews conducted by a second experimenter with each subject showed that children who were rewarded or fined were more likely to attribute their help giving to external pressures than were children in the control group or than were those who received praise or a scolding.

Alternative Explanations for the Foot-in-the-Door Effect

Surprisingly few alternative explanations for the foot-in-the-door effect have been suggested. In part, this is because the self-perception analysis has proven to be a heuristic explanation, leading investigators to examine a wide variety of phenomena such as overjustification effects and the impact of social labels. Other explanations have not generated much research, but they do resurface periodically. Three of these are considered.

Adaptation Level

It has been argued that the small initial request to which people agree establishes a new baseline against which the subsequent larger request is compared (e.g., Schmidt, 1973). In other words, the prior request causes the magnitude of the second to be redefined, making it seem less extreme than it

otherwise would. It is not clear that this explanation would lead one to predict a difference in subsequent compliance between those who refuse and those who agree to the initial request. The perceived relative size of the second request should not depend very much on whether people have complied with the first request. However, several experiments show foot-in-the-door effects to be stronger when those who refuse the first request are excluded from the data analysis (e.g., Snyder & Cunningham, 1975; DeJong & Musilli, Note 3). It should be noted that this explanation cannot easily account for the results of studies looking either at extrinsic justification for the initial act of compliance or at the impact of social labels.

To lay this explanation permanently to rest, two types of studies could be done. First, subjects could be asked to rate the relative size of various requests under differing experimental conditions. Second, the subsequent compliance of those who are merely informed of a request being made of others and those who actually comply with that request could be compared. An adaptation level explanation would predict higher compliance from both of those groups, compared with a control group never before contacted, but self-perception theory would not.

Salience of Social Norms

Harris (1972) suggested that being asked to perform an initial small request makes people more aware of the norm of social responsibility, a norm that prescribes that one should help those who are in need. Like the adaptation level explanation, this argument also leads to the prediction that those who are only informed about the first request will be more compliant with the second demand. It would also lead us to expect a high rate of compliance from those who refuse the initial request, in contrast with available evidence. The major problem with this explanation, however, is that it, too, cannot account for the studies on extrinsic constraints or labels, which are better handled by the self-perception analysis.

Behavioral Consistency

It has also been argued that foot-in-the-door demonstrations show that people try to act consistently with the way they have behaved in the past. Does a person induced to agree to a first request agree to a second in order to sustain a consistent public image? If the effect were due to subjects' attempts at such impression management, then a larger foot-in-the-door effect should result when the second request is made by the same experimenter or involves the same issue or action. The importance of these kinds of variables has not been adequately tested, but the available evidence suggests that similarity between the requests on these dimensions is not terribly important (e.g., Freedman & Fraser, 1966).

A second version of this argument was suggested by Brock (1969), who argued, in essence, that people desire to be psychologically consistent for the sake of a self-image, not a public one. People have behaved in a certain way in the past and continue to behave that way in the future simply because they value consistency. However, it is clear that to manage a self-image in the fashion Brock suggested, people must first code their behavior and understand its attributional meaning. They must decide, essentially, whether the behavior reflects the pressures of extrinsic constraints or reflects their own dispositions or attitudes. Of course, it may be that the self-rewarding consequences of being consistent with one's self-image are what mediate the relationship between changes in self-perception and subsequent behavior. Evidence on this point is lacking.

Conclusions

1. It is concluded that the foot-in-the-door effect can be reliably obtained. Although the number of studies that have failed to demonstrate it unequivocally is surprising, it must be reemphasized that almost all of the studies reviewed showed experimental results in the predicted direction. Furthermore, a number of plausible explanations for failures to replicate can be offered.

2. In addition to the self-perception explanation first offered by Freedman and Fraser (1966), three alternative explanations for the effect have been suggested. It is concluded that these explanations are inadequate accounts of the foot-in-the-door literature.

3. A number of theoretical derivations from the self-perception explanation of the foot-in-the-door effect have been outlined, and the evidence for most is found to be generally supportive. However, investigators have not clearly specified the exact nature of the self-perception changes that are said to mediate the foot-in-the-door effect. Self-perception theory must concern itself with how people classify or code their own behavior and how they are led to form a broad or highly specific inference about their attitudes, traits, or dispositions. Echoing Bem's (1972) admission, it must also be underscored that self-perception theory does not fully explain the link between self-attribution and subsequent behavior. Further work is needed to explore these gaps in the self-perception analysis of the foot-in-the-door effect.

Reference Notes

1. Crano, W. D., & Sivacek, J. *Social reinforcement, self-perception and the foot-in-the-door phenomenon.* Unpublished manuscript, 1978. (Available from W. D. Crano, Department of Psychology, Olds Hall, Michigan State University, East Lansing, Michigan 48824.)
2. DeJong, W. *Consensus information and the foot-in-the-door effect.* Unpublished manuscript, 1978. (Available from W. DeJong, Department of Psychology, Brandeis University, Waltham, Massachusetts 02154.)
3. DeJong, W., & Musilli, L. *Handicapped versus non-handicapped requesters: The effect of pressure to comply with an initial request on the foot-in-the-door phenomenon.* Paper presented at the meeting of the Eastern Psychological Association, Washington, D.C., March 1978.

References

Arbuthnot, J., et al. The induction of sustained recycling behavior through the foot-in-the-door technique. *Journal of Environmental Systems,* 1976–1977, *6,* 353–366.

Baer, R., Goldman, M., & Juhnke, R. Factors affecting prosocial behavior. *Journal of Social Psychology,* 1977, *103,* 209–216.

Baron, R. A. Foot-in-the-door phenomenon: Mediating effects of size of first request and sex of requester. *Bulletin of Psychonomic Science,* 1973, *2,* 113–114.

Batson, C. D., Coke, J. S., Jasnoski, M. L., & Hanson, M. Buying kindness: Effect of an extrinsic incentive for helping on perceived altruism. *Personality and Social Psychology Bulletin,* 1978, *4,* 86–91.

Beaman, A. L., Svanum, S., Manlove, S., & Hampton, C. An attribution theory explanation of the foot-in-the-door effects. *Personality and Social Psychology Bulletin,* 1974, *1,* 122–123.

Bem, D. J. Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6). New York: Academic Press, 1972.

Bem, D. J., & Allen, A. On predicting some of the people some of the time: The search for cross-situational consistencies in behavior. *Psychological Review,* 1974, *81,* 506–520.

Brock, T. C. On interpreting effects of transgression upon compliance. *Psychological Bulletin,* 1969, *72,* 138–145.

Cann, A. A. The effects of performing a small consonant behavior on subsequent attitudes and behavior (Doctoral dissertation, Indiana University, 1975). *Dissertation Abstracts International,* 1976, *36,* 5859B. (University Microfilms No. 76-11,359)

Cann, A., Sherman, S. J., & Elkes, R. Effects of initial request size and timing of a second request on compliance: The foot in the door and the door in the face. *Journal of Personality and Social Psychology,* 1975, *32,* 774–782.

Cialdini, R. B., & Ascani, K. Test of a concession procedure for inducing verbal, behavioral, and further compliance with a request to give blood. *Journal of Applied Psychology,* 1976, *61,* 295–300.

Cialdini, R. B., Cacioppo, J. T., Basset, R., & Miller, J. A. The low-ball procedure for producing compliance: Commitment then cost. *Journal of Personality and Social Psychology,* 1978, *36,* 463–476.

Cialdini, R. B., et al. A reciprocal concessions procedure for inducing compliance: The door-in-the-face technique. *Journal of Personality and Social Psychology,* 1975, *31,* 206–215.

Cook, D. A., Pallak, M. S., & Sogin, S. R. The effect of consensus on attitude change and attribution of causality. *Personality and Social Psychology Bulletin,* 1976, *2,* 248–251.

Cooper, J., Jones, E. E., & Tuller, S. M. Attribution, dissonance, and the illusion of uniqueness. *Journal of Experimental Social Psychology,* 1972, *8,* 45–47.

DeJong, W., & Funder, D. Effect of payment for initial compliance: Unanswered questions about the foot-in-the-door phenomenon. *Personality and Social Psychology Bulletin,* 1977, *3,* 662–665.

Dutton, D. G., & Lennox, V. L. Effect of prior "token" compliance on subsequent interracial behavior. *Journal of Personality and Social Psychology,* 1974, *29,* 65–71.

Even-Chen, M., Yinon, Y., & Bizman, A. The door in the face technique: Effects of the size of the

initial request. *European Journal of Social Psychology*, 1978, *8*, 135–140.

Fish, B., & Kaplan, K. J. Does a "foot-in-the-door" get you in or out? *Psychological Reports*, 1974, *34*, 35–42.

Freedman, J. L., & Fraser, S. C. Compliance without pressure: The foot-in-the-door technique. *Journal of Personality and Social Psychology*, 1966, *4*, 195–202.

Gergen, K. J. *The concept of self*. New York: Holt, Rinehart & Winston, 1971.

Gurwitz, S. B., & Topol, B. Determinants of confirming and disconfirming responses to negative social labels. *Journal of Experimental Social Psychology*, 1978, *14*, 31–42.

Harris, M. B. The effects of performing one altruistic act on the likelihood of performing another. *Journal of Social Psychology*, 1972, *88*, 65–73.

Harris, M. B., Liguori, R., & Joniak, A. Aggression, altruism, and models. *Journal of Social Psychology*, 1973, *91*, 343–344.

Harris, M. B., Liguori, R. A., & Stack, C. Favors, bribes, and altruism. *Journal of Social Psychology*, 1973, *89*, 47–54.

Harris, M. B., & Samerotte, G. The effects of aggressive and altruistic modeling on subsequent behavior. *Journal of Social Psychology*, 1975, *95*, 173–182.

Harris, M. B., & Samerotte, G. The effects of actual and attempted theft, need, and a previous favor on altruism. *Journal of Social Psychology*, 1976, *99*, 193–202.

Kelley, H. H. Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 15). Lincoln: University of Nebraska Press, 1967.

Kraut, R. E. Effects of social labeling on giving to charity. *Journal of Experimental Social Psychology*, 1973, *9*, 551–562.

Lepper, M. R. Dissonance, self-perception, and honesty in children. *Journal of Personality and Social Psychology*, 1973, *24*, 65–74.

Lowman, R. P. Recycling refuse: The effect of the foot-in-the-door technique on attitude and repetitive behavior (Doctoral dissertation, Claremont Graduate School, 1972). *Dissertation Abstracts International*, 1973, *33*, 6114B. (University Microfilms No. 73-14,258)

McArthur, L. A. Perceiving the cause of one's own behavior: An unexpected reversal of Kelley's attribution theory. *Psychological Reports*, 1973, *32*, 983–988.

McArthur, L. A., Kiesler, C. A., & Cook, B. P. Acting on an attitude as a function of self-concept and inequity. *Journal of Personality and Social Psychology*, 1969, *12*, 295–302.

McGuire, W. J., & Padawer-Singer, A. Trait salience in the spontaneous self-concept. *Journal of Personality and Social Psychology*, 1976, *33*, 743–754.

Miller, R. L. Facilitating compliance by manipulating the nature of the comparison: Relative cost vs. reciprocal concession. *Personality and Social Psychology Bulletin*, 1974, *1*, 160–162.

Miller, R. L., Brickman, P., & Bolen, D. Attribution versus persuasion as a means for modifying behavior. *Journal of Personality and Social Psychology*, 1975, *31*, 430–441.

Miller, R. L., Seligman, C., Clark, N. T., & Bush, M. Perceptual contrast versus reciprocal concession as mediators of induced compliance. *Canadian Journal of Behavioural Science*, 1976, *8*, 401–409.

Miller, R. L., & Suls, J. Helping, self-attribution, and size of an initial request. *Journal of Social Psychology*, 1977, *103*, 203–208.

Nemeth, C. Effects of free versus constrained behavior on attraction between people. *Journal of Personality and Social Psychology*, 1970, *15*, 302–311.

Paulhus, D. L., Shaffer, D. R., & Downing, L. L. Effects of making blood donor motives salient upon donor retention: A field experiment. *Personality and Social Psychology Bulletin*, 1977, *3*, 99–102.

Pendleton, M. G., & Batson, C. D. Self-presentation and the door-in-the-face technique for inducing compliance. *Personality and Social Psychology Bulletin*, 1979, *5*, 77–81.

Pliner, P., Hart, H., Kohl, J., & Saari, D. Compliance without pressure: Some further data on the foot-in-the-door technique. *Journal of Experimental Social Psychology*, 1974, *10*, 17–22.

Reingen, P. H. Inducing compliance via door-in-the-face and legitimization of paltry contributions. *Psychological Reports*, 1977, *41*, 924.

Reingen, P. H., & Kernan, J. B. Compliance with an interview request: A foot-in-the-door, self-perception interpretation. *Journal of Marketing Research*, 1977, *14*, 365–369.

Schmidt, L. G. *The foot-in-the-door effect: An inquiry*. Unpublished doctoral dissertation, University of South Carolina, 1973.

Schur, E. M. *Labeling deviant behavior: Its sociological implications*. New York: Harper & Row, 1971.

Seligman, C., Bush, M., & Kirsch, K. Relationship between compliance in the foot-in-the-door paradigm and size of first request. *Journal of Personality and Social Psychology*, 1976, *33*, 517–520.

Seligman, C., Miller, R., Goldberg, G., Gelberd, L., Clark, N., & Bush, M. Compliance in the foot-in-the-door technique as a function of issue similarity and persuasion. *Social Behavior and Personality*, 1976, *4*, 267–271.

Smith, C. L., Gelfand, D. M., Hartmann, D. P., & Partlow, M. E. Y. Children's causal attributions regarding help-giving. *Child Development*, 1979, *50*, 203–210.

Snyder, M., & Cunningham, M. R. To comply or not comply: Testing the self-perception explanation of the "foot-in-the-door" phenomenon. *Journal of Personality and Social Psychology*, 1975, *31*, 64–67.

Steele, C. M. Name-calling and compliance. *Journal*

of Personality and Social Psychology, 1975, *31*, 361–369.

Tipton, R. M., & Browning, S. Altruism: Reward or punishment. *Journal of Psychology*, 1972, *80*, 319–322.

Upton, W. E. Altruism, attribution, and intrinsic motivation in the recruitment of blood donors (Doctoral dissertation, Cornell University, 1973). *Dissertation Abstracts International*, 1973, *34*, 6260B. (University Microfilms No. 74-12,652)

Uranowitz, S. Helping and self-attributions: A field experiment. *Journal of Personality and Social Psychology*, 1975, *31*, 852–854.

Zuckerman, M., Lazzaro, M. M., & Waldgeir, D. Undermining effects of the foot-in-the-door technique with extrinsic reward. *Journal of Applied Social Psychology*, in press.

Received December 1, 1978 ■

U.S. POSTAL SERVICE

STATEMENT OF OWNERSHIP, MANAGEMENT AND CIRCULATION

(Required by 39 U.S.C. 3685)

1. TITLE OF PUBLICATION	A. PUBLICATION NO.	2. DATE OF FILING
JOURNAL OF PERSONALITY AND SOCIAL PSYCHOLOGY	2 8 1 9 4 0	9/30/79
3. FREQUENCY OF ISSUE	A. NO. OF ISSUES PUBLISHED ANNUALLY	B. ANNUAL SUBSCRIPTION PRICE
Monthly	12	$30 Member $60 Nonmember

4. LOCATION OF KNOWN OFFICE OF PUBLICATION (Street, City, County, State and ZIP Code) (Not printers)
1400 N. Uhle Street, Arlington, VA 22201

5. LOCATION OF THE HEADQUARTERS OR GENERAL BUSINESS OFFICES OF THE PUBLISHERS (Not printers)
1200 17th Street N.W., Washington, D.C. 20036

6. NAMES AND COMPLETE ADDRESSES OF PUBLISHER, EDITOR, AND MANAGING EDITOR

PUBLISHER (Name and Address)
American Psychological Association, 1200 17th St., N.W., Wash., D.C. 20036

EDITOR (Name and Address)
Anthony Greenwald, Ohio State Un., Columbus, OH 43210

MANAGING EDITOR (Name and Address)
Anita DeVivo, 1400 N. Uhle St., Arlington, VA 22201

7. OWNER (If owned by a corporation, its name and address must be stated and also immediately thereunder the names and addresses of stockholders owning or holding 1 percent or more of total amount of stock. If not owned by a corporation, the names and addresses of the individual owners must be given. If owned by a partnership or other unincorporated firm, its name and address, as well as that of each individual must be given.)

NAME	ADDRESS
APA	1200 17th St., N.W., Wash., D.C. 20036

8. KNOWN BONDHOLDERS, MORTGAGEES, AND OTHER SECURITY HOLDERS OWNING OR HOLDING 1 PERCENT OR MORE OF TOTAL AMOUNT OF BONDS, MORTGAGES OR OTHER SECURITIES (If there are none, so state)

NAME	ADDRESS
None	

9. FOR COMPLETION BY NONPROFIT ORGANIZATIONS AUTHORIZED TO MAIL AT SPECIAL RATES (Section 132.122, PSM)
The purpose, function, and nonprofit status of this organization and the exempt status for Federal income tax purposes (Check one)

☒ HAVE NOT CHANGED DURING PRECEDING 12 MONTHS ☐ HAVE CHANGED DURING PRECEDING 12 MONTHS (If changed, publisher must submit explanation of change with this statement.)

10. EXTENT AND NATURE OF CIRCULATION	AVERAGE NO. COPIES EACH ISSUE DURING PRECEDING 12 MONTHS	ACTUAL NO. COPIES OF SINGLE ISSUE PUBLISHED NEAREST TO FILING DATE
A. TOTAL NO. COPIES PRINTED (Net Press Run)	7,637	7,655
B. PAID CIRCULATION 1. SALES THROUGH DEALERS AND CARRIERS, STREET VENDORS AND COUNTER SALES	- 0 -	- 0 -
2. MAIL SUBSCRIPTIONS	5,910	6,176
C. TOTAL PAID CIRCULATION (Sum of 10B1 and 10B2)	5,910	6,176
D. FREE DISTRIBUTION BY MAIL, CARRIER OR OTHER MEANS SAMPLES, COMPLIMENTARY, AND OTHER FREE COPIES	282	316
E. TOTAL DISTRIBUTION (Sum of C and D)	6,192	6,492
F. COPIES NOT DISTRIBUTED 1. OFFICE USE, LEFT OVER, UNACCOUNTED, SPOILED AFTER PRINTING	1,445	1,163
2. RETURNS FROM NEWS AGENTS	- 0 -	- 0 -
G. TOTAL (Sum of E, F1 and 2—should equal net press run shown in A)	7,637	7,655

11. I certify that the statements made by me above are correct and complete. SIGNATURE AND TITLE OF EDITOR, PUBLISHER, BUSINESS MANAGER, OR OWNER

12. FOR COMPLETION BY PUBLISHERS MAILING AT THE REGULAR RATES (Section 132.121, Postal Service Manual)

39 U. S. C. 3626 provides in pertinent part: "No person who would have been entitled to mail matter under former section 4359 of this title shall mail such matter at the rates provided under this subsection unless he files annually with the Postal Service a written request for permission to mail matter at such rates."

In accordance with the provisions of this statute, I hereby request permission to mail the publication named in item 1 at the phased postage rates presently authorized by 39 U. S. C. 3626.

SIGNATURE AND TITLE OF EDITOR, PUBLISHER, BUSINESS MANAGER, OR OWNER

PS Form 3526 (Page 1) (See instructions on reverse)

Journal of Personality and Social Psychology
1979, Vol. 37, No. 12, 2240-2252

Salience and the Cognitive Mediation of Attribution

Eliot R. Smith
Department of Sociology
University of California, Riverside

Frederick D. Miller
New York University

Two studies of the impact of salience and informational factors on attribution and memory yielded findings on two general issues. First, manipulations of the amount of thought subjects gave to their attributions and of a delay before responding to attribution questions did not diminish the effect of salience on attribution; in fact, the delay increased the effect. Second, recall of the stimulus material was shown to be influenced by salience and by covariation information (consensus, distinctiveness, and consistency) and to be related to attributions. These findings, together with theory and data from the literature on comprehension and representation of linguistic material in memory, are used to argue that salience is not simply a process by which people make attributions without giving much thought to them. Instead, salience effects reflect the close relationships among the processes of comprehension, remembering, and attribution, and the fact that attributional processing can take place at the time of the encoding and storage of information, as well as at the time of its retrieval from memory.

The literature on the perception of causation currently contains support for two distinct and even opposed theories of the attribution process. Kelley (1967, 1972) has proposed that the perceiver makes attributions after logically weighing information about the covariation of an effect with various possible causes. This covariation information may be observed or assumed (Kelley's, 1972, "causal schema"), but in either case it is processed more or less rationally, in a fashion resembling the analysis of variance (ANOVA). Kelley's model of attribution has come to be known as the ANOVA model for this reason. On the other hand, several recent writers (Pryor & Kriss, 1977; Taylor & Fiske, 1975, 1978) have argued that perceivers use a much simpler information-processing strategy: attributing causality to the most salient plausible cause instead of weighing many possible causes to make a decision. Salience is used synonymously with Tversky and Kahneman's (1974) use of "availability" to suggest some factor that is literally prominent in the perceiver's field of view (Taylor & Fiske, 1975) or that is easily retrievable from memory (Pryor & Kriss, 1977).

A variety of studies have demonstrated that both object salience and the information specified by the ANOVA model *can* influence perceivers in making causal assignments. Taylor and Fiske (1978) present strong evidence that in many everyday situations perceivers give little thought to issues of causation, making attributions to salient stimuli off the "top of the head." These situations are primarily routine ones that involve the perceiver very little. On the other hand, in psychological experiments where information is available and instructions (or implicit demands) urge the subject to consider evidence carefully before making a response, perceivers can respond to information in the fashion Kelley suggests (McArthur, 1972).

The authors are grateful to Susan T. Fiske, Reid Hastie, and Shelley E. Taylor for their comments on an earlier version of this paper. The research reported here was supported by the University of California Academic Senate.

Requests for reprints should be sent to Eliot R. Smith, Department of Sociology, University of California, Riverside, California 92521.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3712-2240$00.75

Is it possible to identify the processes by which both salience and the types of information identified by Kelley influence attributions? That is, going beyond the simple demonstration that different manipulations do influence attributions, research can address the question of how they do. One line of research that may give insights into processes is that aimed at identifying points of demarcation between salience and ANOVA processes (cf. Taylor & Fiske, 1978, Section VII). One prediction is that salience effects might weaken or vanish, and the more thoughtful ANOVA processes take over, when the attributor is highly involved in the situation about which he or she is making causal judgments. Taylor, Crocker, Fiske, Sprinzen, and Winkler (1979) attempted to moderate salience effects on attributions by a variety of manipulations. In several studies they varied distraction, general arousal, and interest in the event to be attributed—all ways of affecting the subjects' degree of involvement in the situation—but were unable to modify the salience effect strongly. This suggests that involvement in events may not lead to more rational, ANOVA-like processing of them; on the contrary, such engrossment may sometimes prevent more active processing. One can readily imagine a harried attributor tossing off causal judgments as needed in response to minimal cues. Study 1 of this article attempted to manipulate the subjects' degree of involvement in attributional *information processing* itself, as distinct from involvement in the *situation*, to determine whether salience effects are weakened when perceivers are forced to give extensive consideration to many possible causes, that is, under conditions that ought to be ideal for the emergence of the more rational information processing described by Kelley's ANOVA model.

The manipulation of salience used in this study was used and validated by Pryor and Kriss (1977). Their Study 1 showed that varying the order of the person and object in a sentence influenced the salience of each word: Specifically, the first-presented sentence element was identified more quickly in a reaction time task. The more salient word was more available in memory. Their Study 2 then showed that this salience manipulation, applied orthogonally to an information manipulation (sentences varying consensus, distinctiveness, and consistency), did affect attributions in the predicted direction. The more salient sentence element (person or object) attracted more attributions from subjects. It can be argued, however, that Pryor and Kriss's (1977) procedure, in its use of simple closed-ended attribution scales as the dependent measures, encouraged subjects to make quick, relatively unthinking attributions and hence encouraged reliance upon the salience cue. The current study focused upon this possibility. This study also examined recall data that may illuminate the processes that underlie the effects of salience and informational factors on attributions.

The present Study 1 conceptually replicates and extends Pryor and Kriss's (1977) Study 2. Pryor and Kriss presented subjects with 16 sentences of the type "Joe likes the film." In a within-subjects experiment they manipulated salience (by varying the order of presentation of subject and object in the stimulus sentence), information (by including additional sentences that contained consensus, distinctiveness, and consistency information pointing to either the subject or object as causal), and verb form (the verb was "like" or "dislike"). They measured attributions of causality to the subject and object on separate 11-point scales but reported analyses only on the difference between these two attributions. The current study largely followed Pryor and Kriss's procedure, with the addition of a between-subjects manipulation intended to influence strongly the amount of time and attention subjects gave to consideration of several possible causes in answering the attribution dependent measures, and a recall measure, administered after the conclusion of the attribution questionnaire, in which subjects attempted to reproduce the stimulus sentences.

Three specific predictions are made for this study.

1. Salience phenomena have been viewed as characteristic of perceivers who give attributional responses on the basis of a quick, sketchy analysis because they are uninvolved

—or are overly engrossed and pressed for time. This view implies that increased thought or information processing concerning attributions should diminish the effects of the salience manipulation.

2. If salience effects are mediated by cognitive availability or similar factors that affect memory, then relationships between attributions and memory should appear. Differential availability of the subject and object of the sentence should influence the way the sentence is recalled as well as the attribution that is made. In general, relationships between memory and attribution should provide information about the cognitive dynamics of salience.

3. Finally, the recall measures may be influenced by the manipulations (salience and information) that are applied to the sentences and are known to affect attributions. Effects of salience on recall would confirm once again (as in Pryor & Kriss's Study 1) that salience operates through an effect on the stimulus sentence's representation in memory (thereby influencing its recall). Effects of covariation information on recall are a more interesting possibility; they would point to a close connection between attribution and the cognitive processes involved in sentence comprehension and recall.

Study 1

Method

Overview. The first study replicates Pryor and Kriss's Study 2, extended in three ways. The first extension is the addition of a between-subjects manipulation intended to influence the amount of thought subjects give to their attributions. Subjects in the little thought condition received the closed-ended dependent measures used in the Pryor and Kriss study, two 11-point scales for attributions to the subject and object of the stimulus sentence (wording was, "To what extent was the given situation caused by [name of person or object]?"). Subjects in the moderate thought condition were instructed to think of and write in their own words one explanation for the event presented in the stimulus sentence and then to code that explanation using the two scales. Subjects in the extensive thought condition were asked to think of and write several (as many as five) different possible explanations for the given event, to consider them all carefully, to indicate the best explanation and, finally, to code that cause using the two closed-ended scales. A larger amount of careful and systematic thought should be given to the final closed-ended responses in the moderate and extensive thought conditions than in Pryor and Kriss's little thought condition.

The second extension is the addition of a recall measure. After subjects completed the experimental booklet that presented the 16 sentences and causal measures, they were given a sheet of paper with instructions to "list below as many [sentences] as you can remember—just the sentences, not the additional information that was given." Each sentence was coded for whether it was recalled with subject first or object first (paralleling the salience manipulation) and whether the subject and object of the sentence were each recalled.

The third extension is of a minor nature. Eight different positive verbs and eight different negative verbs were used in place of Pryor and Kriss's two verbs "like" and "dislike." This was intended to provide confirmation of their unexpected finding of a Verb Form × Information interaction.

Subjects. The subjects were 84 undergraduate and graduate students[1] (28 randomly assigned to each of the three conditions) at the University of California, Riverside, who were paid for their participation. Subjects were run in groups that were heterogeneous as to experimental condition and in which communication among subjects did not take place. The experimenter who administered the questionnaire and debriefed the subjects was blind to subjects' conditions.

Materials. Sixteen verbs were used in sentences, 8 different verbs implying a positive relationship between the subject and object of the sentence (e.g., likes, agrees with, is pleased with, helps, values), and 8 different negative verbs (e.g., dislikes, despises, complains about, is angry with). With this exception, the salience and information manipulations are as described by Pryor and Kriss (1977). Salience refers to the order of presentation of the subject and object, and information to the consensus, distinctiveness, and consistency factors presented in the same manner as in McArthur (1972). The salience manipulation is not equivalent to active versus passive syntactic forms of the sentence, although it does overlap somewhat. Some sentences (such as the first example below) are syntactically passive in the person-salient version (e.g., "Penny is disgusted by the turnips . . .") and active in the object-salient version (as shown). Examples are (object salient with object information):

The turnips at dinner disgust Penny.
The turnips at dinner disgust almost everybody.
Most other vegetables do not disgust Penny.
In the past the turnips at dinner have almost always disgusted Penny.

---

[1] The five graduate-student subjects were all from fields unrelated to social psychology, and exclusion of their data does not materially affect the results.

(person salient with person information):

Phil complains about his co-worker's habits.
Almost nobody else complains about Phil's co-worker's habits.
Phil complains about almost everybody's habits.
In the past Phil has almost always complained about his co-worker's habits.

(object salient with person information):

The vegetables at the Italian restaurant are liked by Tony.
The vegetables at the Italian restaurant are not liked by almost anybody else.
Most other vegetables are liked by Tony.
In the past the vegetables at the Italian restaurant have almost always been liked by Tony.

The salience, information, and verb form (positive/negative) manipulations were all orthogonal, forming a $2 \times 2 \times 2$ design of which each subject completed two entire replications, for a total of 16 different sentences. Four questionnaire forms were used, so that each stimulus sentence was presented to equal numbers of subjects in each of the four Salience $\times$ Information conditions. The order of presentation of the 16 sentences was randomized separately for each subject to avoid systematic order effects. The dependent measures (attribution and recall) were described above.

Results

The results fall into three sections: attributions, recall, and relations between attribution and recall. Attributions and recall were analyzed in a Thought $\times$ Salience $\times$ Information $\times$ Verb Form $\times$ Replication ($3 \times 2 \times 2 \times 2 \times 2$) analysis of variance, with all but thought being within-subjects factors. Replication is always treated as a dummy factor, and any significant effects it may have are ignored; it has no significant interactions that qualify effects discussed below.[2]

Attribution measures. The attribution dependent measures were subject attribution, object attribution, and the difference between them (the measure used by Pryor & Kriss, on which a higher score means more person attribution). In the current data, the within-conditions correlation between subject and object attribution averages $-.470$, $p < .05$, justifying the use of the difference score. In addition, results for all three measures are quite similar.[3] Considering the difference measure, four significant effects appeared, the first three of which were also obtained by Pryor and Kriss.[4] Covariation information strongly affected attributions: The mean attribution score was 5.08 with person-oriented information and -1.74 with object-oriented information. (Fs are shown in Table 1.) Salience also had a significant effect on attributions. The mean was 2.03 when the sentence was presented with the person salient and 1.32 when the sentence was presented with object salient. A significant Verb Form $\times$ Information interaction also appeared, with covariation information having more impact when the verb was negative than when it was posi-

---

[2] The use of an alternative analysis based on quasi F ratios, recognizing that replication as well as subject is a random factor in the design, does not change the essential conclusions. In the current study, subjects contributed much more variance than replications. The analyses reported here are ordinary Fs based on the appropriate interactions with subject as error terms, because this approach is more familiar and because in some analyses to be reported, the use of dichotomous scores renders untrustworthy the high-order interactions required for the computation of quasi Fs.

[3] The results for the object–attribution dependent measure were the same as for the difference measure, except that the effect of verb type was not significant. The results for the person measure were the same, except that the salience effect was only marginally significant ($p < .06$) and a three-way Verb Type $\times$ Salience $\times$ Information interaction reached significance, indicating that with negative verbs, object information and especially object salience gave rise to much less person attribution. The results for the difference measure thus summarize well the results for the two separate dependent variables: Four of the five significant effects on the difference measure (all except verb type) are at least marginally significant on both the person and object measures.

The strength of the person–object attribution correlation, in contrast to the typical findings of a nonsignificant correlation between measures of situational and dispositional attributions (e.g., Miller, Smith, & Uleman, Note 1), may be due to the presentation of the informational manipulations. These sentences strongly imply a person or object attribution for each stimulus sentence and hence may sensitize subjects to the idea that person and object attributions are alternatives.

[4] Verb form also affected attributions, with more person attributions being given for positive than for negative verbs. The Verb Form $\times$ Replication interaction was significant, however, so the verb form effect should be interpreted with caution (and it receives no further attention in this article).

Table 1
Analysis of Variance Results for Difference Attribution Measure and Significant Effects in Study 1

Effect	*MS*	*df*	*F*	*p*
Salience	170.715	1	6.556	.013
(Error) Salience × Unit	26.038	81	—	—
Information	15,614.238	1	205.811	.000
Information × Thought	435.421	2	5.739	.005
(Error) Information × Unit	75.867	81	—	—
Verb Form × Information	435.435	1	13.761	.001
(Error) Verb Form × Information × Unit	31.642	81	—	—

tive (means appear in Table 2). This replication of Pryor and Kriss's finding on verb form with an extended sample of verbs increases confidence in the effect. This interaction may be due to subjects being more careful (i.e., paying more attention to information) in their attributions for negative events, perhaps because such events have bad implications or are unusual (cf. Kanouse & Hanson, 1972).

The other attribution effect involved the between-subjects manipulation and so was not part of Pryor and Kriss's findings. The thought manipulation was successful. Subjects in the extensive thought condition wrote down and considered a mean of 3.1 different explanations before selecting the best. Subjects in the moderate thought condition wrote down one explanation and then coded it, whereas subjects in the little thought condition had only to make two checkmarks on closed-ended attribution scales. It is notable that the data contain no support for Prediction 1, which called for the weakening of salience effects in the extensive thought condition; for the Thought × Salience interaction, $F(2, 81) = .78$, $p = .46$. Salience effects on attributions were as strong when subjects carefully considered several possible explanations as when they simply made two checkmarks on attribution scales. The thought manipulation did interact with information, however, with the impact of covariation information being largest in the little thought condition and smallest in the extensive thought condition. Means appear in Table 2, and *F* test results in Table 1.

Table 2
Means of Difference Attribution Measure for Study 1

Item	Information	
	Person	Object
Thought[a]		
Little	5.683	−2.638
Moderate	5.839	−1.705
Extensive	3.723	−0.862
Verb form[b]		
Positive	4.860	−0.818
Negative	5.304	−2.652

Note. A higher mean indicates more person attribution and less object attribution.
[a] $n = 224$ observations per mean.
[b] $n = 336$ observations per mean.

Recall measures. Four recall measures were analyzed. These were the proportion of sentences recalled with person first, the proportion recalled with object first,[5] the proportion of persons recalled at all, and the proportion of objects recalled at all. Partial recall was fairly frequent, especially of the form "____ complains about his co-worker's habits." This example would have been scored as recalled

[5] These two proportions are independent, as they add to the proportion of sentences recalled at all, and not to 1.0. The order-of-recall variables were scored independently of the initial salience (i.e., order of presentation) of the sentences—that is, any recall of a sentence in a particular order was counted, not just correct recall of order.

with person first, object recalled, person not recalled. The reliability of this coding was near perfect. A second rater checked approximately 30 sentences, with no disagreements on coding.

Both salience and information influenced the order of recall variables. More sentences were recalled with the object first when the object was salient ($M = .125$ vs. .029), $F(1, 81) = 28.31$, $p < .001$. A Salience × Verb Form interaction affected both of the order-of-recall variables. Object salience led to stronger object-first recall, and person salience led to stronger person-first recall, when the verb was negative as compared with positive; for person-first recall, $F(1, 81) = 4.82$, $p < .05$; for object-first recall, $F(1, 81) = 5.66$, $p < .01$.

More important is an interaction of salience with covariation information. Inspection of the Salience × Information interaction means (Table 3) reveals that information has little or no effect on order of recall in the person-salient condition, where nearly all recalled sentences were reproduced accurately with person first—in this condition the mean object-first recall was only .029. When sentences are presented with object salient, however, there is more variation in the recall measures, and covariation information has an effect. For the Salience × Information interaction, for person-first recall, $F(1, 81) = 7.16$, $p < .05$; for object-first recall, $F(1, 81) = 9.16$, $p < .01$. Tests of the simple effects of information within the object-salient condition, using the appropriate pooled error mean square (Winer, 1970, pp. 544–545), reveal that this effect is significant for both the person-first recall and object-first recall measures, $t(81) = 3.07$, $p < .01$, and $t(81) = 2.93$, $p < .01$, respectively. In both cases, covariation information influences recall in the predicted direction, with the factor the information pointed to as the cause more likely to be recalled first. The clear conclusion is that these recall measures suffer respectively from a ceiling effect and a floor effect in the person-salient condition, producing the observed Salience × Information interaction. When the analysis is focused on the object-salient condition in which the absence of ceiling and floor effects allows the recall measures to show more variance, the effects of the covariation information manipulation on recall are significant.[6]

Table 3
Means for Recall of Verbs by Salience and Information for Study 1

Proportion recalled	Information: Person	Information: Object
Person salient (person first)		
Person first	.426	.459
Object first	.047	.012
Object salient (object first)		
Person first	.470	.358
Object first	.098	.152

Note. Each proportion is based on 336 observations.

Correlations between attributions and recall. Finally, we come to the relationships between attributions and recall. These were analyzed as partial correlations to ensure that a true relationship between the variables is present, not a spurious relationship resulting from the variables being affected in similar ways by the manipulated factors (cf. Taylor & Fiske, 1978, section VI.C). The partial correlations of the two attribution measures (person and object) with the four recall measures were computed across responses, controlling for salience, information, and experimental conditions (thought). This test shows

[6] Other effects on recall measures, not of central theoretical significance, can be briefly mentioned: More sentences were recalled person first, and more persons were recalled, in the extensive thought condition than in the other two conditions, $F(2, 81) = 4.54$, $p < .05$. Person recall was influenced strongly by verb form, with more persons presented with positive verbs being recalled (means of .317 vs. .208), $F(1, 80) = 19.05$, $p < .001$. The proportion of objects recalled was influenced by salience, with more object recall when the object was salient, $F(1, 80) = 5.95$, $p < .05$. Both object recall and person-first order of recall were influenced by a three-way interaction of Thought × Verb Form × Salience. However, this effect appears not to be simply interpretable and does not qualify the interpretation of any other significant effects, and it will not be discussed further.

that two of the eight relationships are significant. (The probability of obtaining at least two of eight relationships at the .05 level under the null hypothesis is less than .002 by a binomial test.) The proportion of sentences recalled object first was associated with attributions to the object on the object attribution scale, partial $r(1317) = .08$, $p < .01$, and proportion of persons recalled was associated with higher person attribution on the person scale, partial $r(1317) = .06$, $p < .05$. Thus, relationships between attribution and recall measures do appear when the experimental manipulations are controlled. The small size of these correlations can be attributed to the facts that (a) the attributional measures may be of low reliability (the two items only correlate .47) and (b) these are correlations between a continuous variable and a dichotomy, which can limit the maximum size of correlation coefficients to substantially less than 1. (Correction of these correlations for the attenuation due to the use of a dichotomy yields estimates of .15 and .08 for their "true" sizes, respectively; Harshbarger, 1977, p. 440.)

Discussion

Three basic findings were replicated in this study, and two new relationships were uncovered. Salience affects both recall and attribution, and covariation information affects attribution, as Taylor and Fiske (1975) and Pryor and Kriss (1977) have shown. In addition, information affects recall, in the object-salient condition where the recall measures showed adequate variance. Also, a partial correlation between recall and attributional measures emerged (controlling for the experimental factors).

Contrary to Prediction 1, thought (the extensive thought condition compared to moderate and little thought conditions) did *not* diminish the effects of the salience manipulation: no Salience × Thought interaction appeared. This supports the findings by Taylor et al. (1979) that salience effects remain robust despite a wide variety of manipulations. Rather than minimizing the effects of salience, greater thought appears to decrease the importance of the covariation *information* manipulation. The extensive thought manipulation was specifically intended to influence subjects to consider several explanations and choose the best one, rather than to engage in simple "top of the head" processing of just one causal possibility—the most salient one. The manipulation was evidently successful, as subjects in the extensive thought condition considered a mean of 3.1 different explanations. Yet the impact of salience was not reduced by the manipulation. A couple of points should be made about the thought manipulation. First, the particular manipulation used does not artifactually bias subjects to give more or less weight to any particular kind of information as a determinant of causal attributions (e.g., the covariation information); it asks them only to choose the best cause in the light of all available information. However, the manipulation may have other effects besides the intended one. For example, there may be a motivational effect: In the extensive thought condition subjects may have perceived the task as more important (since they were asked to spend so much time at it) and as a result may have tried harder to be accurate in their attributions. Still, the earlier theory of salience (as based on unthinking top-of-the-head responses) would predict that such a motivational effect would also reduce the impact of salience—contrary to what was found. The thought manipulation thus seems adequate to support the conclusion that we wish to draw from it: Salience effects persist even when subjects give careful consideration to many possible causes of an event.

The fact that salience manipulations have an impact over a broad range of conditions indicates that salience is not just a matter of making attributions without thinking. Instead, salience effects appear to be related to such cognitive processes as the way the stimulus sentence is perceived and encoded in the first place, giving rise to the observed memory–attribution links. These links are both the effects of information on recall (implicating attribution processes in the comprehension and recall of sentences) and the partial correlations between recall and attribution indices.

A simple model can account for the findings. In carrying out the experimental task, subjects must first comprehend the stimulus materials, then report their attribution, then (in the recall task) report their memory of the sentence itself. Results of Study 1 show that salience and information affect *both* the attribution and recall results and also that there is a partial correlation between attribution and recall indices. These effects can most parsimoniously be accounted for by the hypothesis that attributional processing and recall both operate from a single internal, encoded representation of the stimulus sentence to which the subject refers in answering the questions he or she is posed. The experimental factors that affect both attribution and recall can thus parsimoniously be assumed to affect the nature of the internal representation and thereby, indirectly, the dependent variables. It is easy to see that salience may affect the way a sentence is encoded, but the hypothesis that covariation information does so as well is novel. It has been generally assumed that such informational manipulations affect attribution but do not affect memory per se (Pryor & Kriss, 1977; Taylor & Fiske, 1975). Here, however, information is shown to affect recall, so parsimony suggests that the information factor too may operate through an effect on the representation of the sentence in memory. This proposition will be further tested in Study 2.

The subject's reporting of the attribution could then take place in either or both of two ways. The subject could retrieve from memory an attribution that was made during encoding and report it, *or* the subject could retrieve a representation of the sentence as well as of the additional consensus, distinctiveness, and consistency information and make a causal judgment at the time of retrieval. This argument that attributions are sometimes made at the time information is stored in memory, rather than subsequently on the basis of data retrieved from memory, has two noteworthy implications. First, as mentioned above, it is contrary to Pryor and Kriss's (1977) interpretation of salience. They raise two arguments against the idea that attributions are encoded and retrieved from memory: Since their subjects were not told that the experiments involved an attribution task they had no reason to make attributions; and covariation information did not influence the speed of recall in their study, as would be expected if attributions were involved in information encoding and storage. In response it can be argued that (a) social psychologists have often claimed that attributions are frequently made by people in their everyday lives, not just when experimenters ask for them. If we did not believe this, there would be little point in studying attributions. (b) Subjects in the Pryor and Kriss study, as well as in the current study, received the covariation informational manipulations with each sentence. Since these strongly point to either the subject or the object as the cause, it would be surprising if reading these stimuli did not generate some causal thinking. (c) In the current study, information was shown to influence two recall measures (not measures used by Pryor & Kriss).

Second, the notion that attributions can be made as information is stored in memory is consistent with many current theories of sentence comprehension and memory in cognitive psychology that give important places to inferred (attributed) causality. Among others, Kintsch (1974), Norman, Rumelhart, and the LNR Research Group (1975), and Schank (1975) have all proposed theories that involve the representation of "cause" links between propositions. In comprehending a text like "A burning cigarette was carelessly discarded. The fire destroyed many acres of forest," people make the attribution that the cigarette caused the fire and seem to store that inferred causal link as if it were explicitly stated in the text, when they are tested after a 15-minute delay (Kintsch, 1975). The cognitive theories and studies support the idea that attributional (cause-inferring) processing is intrinsically involved in the initial comprehension of sentences and therefore that it goes on all the time, not just when a subject is asked an attributional question. Such theories also imply that the cause generated by the initial attributional processing is represented in memory in some form. Hence relations between recall and causal responses,

such as those obtained in Study 1, are to be expected.

Our tentative model thus holds that subjects comprehend the stimulus material: They process the sentence as influenced by the salience and information factors and transform it into a representation in memory. Some attributional inferences occur in this process, and causal material is included in the memory representation (Kintsch, 1974, 1975). If subjects are immediately asked to give an attribution, they may report the cause they have stored in memory—but on the other hand, the additional covariation information may still be available to them, so subjects may be able to carry out further attributional processing in response to the question. If a delay intervenes between the study of the material and the attributional question, Kintsch's results suggest that subjects will be more likely to answer the question directly from the memory representation. The additional covariation information will in general have been lost from memory during such a delay. Only the stimulus sentence itself will be retrievable because it was more intensively studied by the subjects and because it is more central; such propositions are better recalled (Kintsch, 1974). The prediction of this model is thus for a stronger relationship of attribution and recall indices after a delay than immediately and for a greater influence of salience on attributions (via its influence on the memory representation) after the delay. These hypotheses parallel Kintsch's (1975) results showing that after a delay causal relations seem to be retrieved as part of a person's encoded memory for the event itself, not as separately encoded information.

Study 2 was conducted as a test of these ideas. In particular, it was predicted that (a) the effect of information on recall found in Study 1 should be replicated within the object-salient conditions where some variation exists in the order-of-recall measures; (b) the partial correlation of attribution and recall variables should be replicated; (c) this attribution–recall relationship should increase with time after the presentation of the sentences, since other information (e.g., the covariation information) should have decayed from memory and the subject will be forced to base his attribution and recall responses on the same encoded memory representation of the sentence (cf. Kintsch, 1975); (d) the effect of salience on attributions should also be stronger after the delay, since that effect is presumed to be mediated by the memory representation.

Study 2

Method

Overview. Subjects read sentences similar to those in Study 1, together with the salience and information manipulations, presented as materials for a study of memory. After studying the sentences for 8 minutes, some subjects (randomly assigned to the immediate condition) were asked to recall as many sentences as they could in 10 minutes, writing each one and then answering the two closed-ended subject and object attribution questions about it (essentially Pryor & Kriss's questions). During this time, the other subjects (those in the delayed condition) performed an irrelevant anagrams task. A 20-minute delay then followed, during which the subjects listened to and took notes on a lecture on an unrelated topic. Finally, the memory and attribution dependent measure was administered to all subjects, again for 10 minutes. Thus, immediate recall subjects filled out the measure beginning only seconds after they had finished studying the sentences, and delayed subjects filled it out 30 minutes later.

Subjects. The 107 subjects were volunteers drawn from introductory social psychology courses at the University of California, Riverside (26 subjects) and at New York University (81 subjects). The study was run during a class period, prior to exposing the students to material concerning attribution in the course.

Materials. The 16 sentences used in this study were written to have a person as subject and a nonhuman object, to eliminate ambiguity in the wording of the attribution questions (see below). Some sentences from the first study were retained and some new ones were written. The salience and information manipulations were as described for Study 1. In this study, to simplify the experimental materials and the task of balancing cell sizes in two locations 3,000 miles apart, sentences were nested within (Salience × Information), so that just one form of stimulus presentation needed to be prepared. That is, each sentence was presented in just one Salience × Information condition.

The dependent measure consisted of 16 occurrences of three questions: "Write down one of the sentences here, as exactly as you can remember it—just the sentence, not the additional information about it" (i.e., the information manipulation), followed by person and object attribution questions. These were the

same as those used in the little thought condition in Study 1, except that instead of including the person's name and the object from the sentence (which would have constituted a powerful recall cue for each sentence), the questions asked, "To what extent was the event described in the sentence caused by the *person* named in the sentence" and "caused by the *object* the person was reacting to."

Coding of variables. From the dependent measure, each stimulus sentence was coded as recalled with person first, with object first, or not recalled; the person was coded as recalled or not; and the object was coded as recalled or not. Four recall measures identical to those in Study 1 result. Since attribution measures are available in this design only for sentences that were recalled, analyses involving the attribution variables must exclude sentences that were not recalled. For such analyses the two order-of-recall variables are redundant (since a recalled sentence must be scored as either recalled person first or recalled object first) and only one can be used. The person and object attribution scales were used in their original form, with a higher score meaning in each case less person or object attribution. A difference score was not used in this study, since the correlation of the two attribution scales was only −.26, not strong enough to justify using the difference as a single summary measure.

Analysis. Attributions cannot be analyzed in a repeated-measures design in this study because they are not given for sentences the subject did not recall. Recall, however, can be so analyzed. The ANOVA design is Time × Sentence Nested Within (salience by information), the latter three all being within-subjects variables. Attributions and the relationships among recall and attribution variables can be analyzed with the sentence as the unit of analysis. Since a subject typically recalled several sentences, the observations in this mode of analysis are not independent of each other, giving rise to some degree of bias. The bias is small, however, since the variance due to sentences is generally many times greater than that due to subjects in these analyses, and this analysis is the most effective available for examining the questions of interest.[7]

Results

It was predicted first that this study would replicate an effect found in Study 1, an impact of information on the order of recall of a sentence, within the object-salient condition. Unfortunately, no such effect was found, and the reason became clear in examining the ANOVA table: The design used, with sentences nested within information, had very low power to detect such effects, since the sentence factor contributed much variance (sentence was significant at the .001 level for both the order-of-recall dependent measures). This indicates that the design should have crossed Sentence × Information Condition, as Study 1 did, yielding higher power for tests of information effects (cf. Kenny & Smith, Note 2).

The second prediction was for a within-conditions relationship between attribution and recall. This was examined by computing correlations between the three recall variables (order of recall, person recalled, object recalled) and the two attribution variables (person and object attributions) separately for the four combinations of Time × Salience. (Information need not be considered, since it did not significantly affect recall; thus, it cannot generate spurious correlations between recall and attribution measures.) One of the 12 correlations from the immediate subjects was significant: In the object-salient condition, object recall was associated with less person attribution, $r = .14$, $p < .05$. For the delayed subjects, 4 of 12 correlations reached the .05 level, all in the object-salient condition. All three recall variables (object-first order of recall, person recall, and object recall) were associated with more person attribution ($r = -.16$, $-.22$, and $-.17$, respectively). Object recall was also associated with more object attribution, $r = -.20$. Overall, obtaining 5 of 24 correlations at the .05 level by chance

[7] The use of analyses that may contain technical violations of assumptions has precedents in the attribution literature. McArthur (1972) found herself in a situation similar to ours, argued that the bias due to ignoring subject effects would be "negligible," and proceeded with an analysis that is conceptually similar to the ones we perform. Our data permit a reassurance like the one she was able to give: The correlation of recall responses given by the same subject is negligible relative to item effects. In addition, we have conducted reanalyses to test the effects of the violations of assumptions. Removal of subject effects from the attribution variables (by subtracting each subject's mean score from each observation) and reanalyses of these corrected scores show that the major points of the paper are unaffected by the nonindependence in the original data. This is the case with the partial correlations between memory and attribution in Studies 1 and 2, the ANOVA on attribution in Study 2, and the path models in Study 2. (The other analyses reported in the paper explicitly incorporate the subject variable and so are not subject to the criticism of nonindependence of observations.)

alone has a binomial probability of less than .002: a clear demonstration of relationships between recall and attributions.

The third prediction was that the relationship between recall and attributions should be stronger when recall was delayed. This issue can be illuminated by path diagrams showing the influences of the experimentally manipulated factors on recall and attribution variables, separately for the immediate and delayed recall conditions. Using the conventions of path analysis to draw the diagrams (Duncan, 1975), the following are the results for the most sensitive recall and attribution variables, object-first recall and person attribution.[8]

In Figure 1, the straight arrows represent the causal effects (numerically, the standardized regression coefficients) of the manipulated variables on the two response variables. The residual causes u and v represent all other causes of the responses (i.e., error variance). The curved arrow between u and v represents a correlation between those residual causes, generated theoretically by an unmeasured factor that has a causal impact on *both* recall and attribution (in this case, the factor is presumed to be aspects of the memory representation of the sentence). Note that the model shows neither the recall nor attribution dependent variables as causing the other. This is a consequence of the theoretical viewpoint sketched above: The recall and attribution processes must work from the same encoded representation of the sentence (so one can expect to find a partial correlation between them, represented in the diagram by the correlation between the residual causes u and v). However, neither can be presumed to cause the other, even though the recall dependent measure comes first in time—attributions (those made during the initial processing of the sentence) may influence recall as well as the other way around.

In the immediate condition, there seems to be a separation of the recall and attribution processes: Salience affects recall, and the informational manipulation affects attribution. Furthermore, the residual correlation between recall and attribution is nonsignificant. However, the delayed condition shows a different

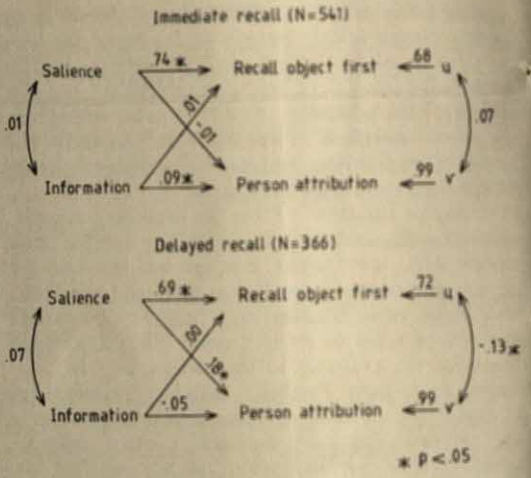


Figure 1. Path diagrams showing relations among manipulated factors, recall, and attribution in immediate and delayed conditions, Study 2. ($p < .05$.)

pattern of effects. Salience still strongly affects order of recall, but it now also affects attribution. The information effect has vanished. Finally, the partial (residual) correlation is now significant. (These three correlations or path coefficients differ significantly between the immediate and delayed conditions.) An

[8] Detailed results for the other combinations of variables are not presented here for reasons of space. Briefly, the object attribution measure never shows significant relations in these analyses, so only the person scale need be considered. There are three nonredundant recall measures, object-first recall, person recall, and object recall. (Since only sentences that are recalled have attributions, results using the person-first order-of-recall variable are identical to those with the object-first variable except for having the opposite sign.) The major conclusions are the following: (a) For two of the three analyses (object-first recall and person recall), the delayed recall-attributional correlation is significant and significantly different from the immediate one; (b) in all three cases, there is a significant delayed effect of salience on attribution, significantly greater than the immediate effect; (c) for object recall only, there is a significant delayed effect of information on recall—such that information implying that the object is causal leads (after the delay) to a greater likelihood of recalling the object. This effect supports our model but is not discussed further, since it appears with only one of the three measures. Researchers on salience, at least since Taylor and Fiske (1975), have often obtained theoretically expected effects with only a subset of dependent measures; the exact form of the most appropriate measure is a matter for further study.

analysis of variance on the person attribution variable verifies that the impact of salience is stronger after the delay: Salience × Time interaction, $F(1, 12) = 8.39$, $p < .05$.

Discussion

This pattern of findings lends support to the idea that after a delay (though not immediately after study of the sentences) recall and attribution are based on a single encoded representation of the sentence. The effect of salience on that encoded representation seems to be stronger than the effect of the information manipulation: the effect of salience on attribution was significantly stronger after the delay than before. On the other hand, the impact of information was significantly weaker (and nonsignificant) after the delay. This pattern may be accounted for by the salience manipulation's status as part of the sentence itself, whereas the covariation information was presented as separate sentences, subject to memory decay during the 30-minute delay.

Several puzzles remain in the results. First, the residual correlation between recall order and attribution in the delayed condition is in the wrong direction. That is, order of recall has a relation to attributions opposite to that of the order of presentation of the sentence. The reason for this is unknown; the exact form of the representations of sentences in memory is a matter of some controversy (cognitive theories differ widely on the issue). Similarly, we do not yet know what specific characteristics of the representation are responsible for observed differences in order of recall or attribution. Salience does appear to affect the memory representation, but the development of sensitive and relatively pure measures of the affected characteristics remains a topic for future research. Second, the recall–attribution partial correlations in Study 1, although significantly different from zero, were small. This may also be due to the use of measures that are only indirectly related to the true variables of interest, or the effects may actually be relatively small in magnitude (and thus difficult to study). In either case, the elucidation of the relationships among salience, attributions, and the cognitive processes and representations involved in sentence comprehension is a goal that deserves the further research efforts that will be required.

Conclusions

In summary, this article has attempted to accomplish three things. It has presented evidence that one effect of salience—the dependence of attributions upon the order of presentation of person and object in a stimulus sentence—is surprisingly robust. The effect holds up even when the subject is instructed to invent several possible causes and pick the best one and becomes *stronger,* rather than weaker, after a delay that forces subjects to rely more on an encoded memory. This finding led to the idea that salience processing involves more than just top-of-the-head responses but may reflect the basic cognitive processes underlying sentence comprehension and the storage of events in memory. (This analysis could be extended to visual manipulations of salience, as in Taylor & Fiske, 1975, by assuming a verbal representation of encoded visual material; cf. Chase, 1978.) One implication of the data is that strict lines of demarcation between salience processes and more thoughtful attribution processes do not exist. Rather, one must seek to understand how salience and informational factors exert influence on the perceiver across a series of stages of encoding, processing, and recall.

Second, two new findings concerning memory are presented (in addition to the replication and extension of Pryor & Kriss's finding that salience affects memory as well as attributions): an effect of information on recall (in Study 1) and a relationship between attributions and characteristics of the recall of a sentence (in both studies). Further research is clearly needed to investigate the effects of different types of stimuli and different delay periods and particularly to identify the most appropriate and sensitive recall measures with which to elucidate the cognitive processes underlying attribution. The present findings point to a close relationship between the processes that subjects use in reporting attri-

butions and in reporting recall; in particular, we advance the hypothesis that both of these processes work from an encoded representation of the stimulus that the subject initially forms. This interpretation is supported by the fact that both the recall–attribution relationships and the effects of salience are increased by a delay (Study 2). Although the support of the hypothesis by these findings is inferential, theory and data from the literature on text comprehension and representation, cited earlier, also argue strongly for such an encoding process, which can involve attributional processing at the time of the encoding and storage of a stimulus as well as at the time of its retrieval. Thus attribution can mediate memory, as well as memory mediating attribution.

Third, an implicit message of these arguments has been that the formulation of explicit theories of information processing in an experimental task can be a productive approach for future research. Instead of asking about whether one variable mediates another in the sense that a statistical relationship between them can be established, we should attempt to understand how the subject goes about performing the task and therefore how the results of a process such as encoding or memory can influence other subsequent processes. Such explicit theories of stages in information processing lend themselves readily to test by means that have been used in cognitive psychology with great success for the past several years (Schneider & Shiffrin, 1977; Shiffrin & Schneider, 1977; Taylor, 1976).

Reference Notes

1. Miller, F. D., Smith, E. R., & Uleman, J. *Can we measure causal attributions?* Manuscript submitted for publication, 1978.
2. Kenny, D. A., & Smith, E. R. *A note on the analysis of designs in which subjects receive stimuli only once.* Manuscript submitted for publication, 1978.

References

Chase, W. F. Elementary information processes. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes: Vol. 5. Human information processing.* Hillsdale, N.J.: Erlbaum, 1978.

Duncan, O. D. *Introduction to structural equation models.* New York: Academic Press, 1975.

Harshbarger, T. *Introductory statistics: A decision map approach.* New York: Macmillan, 1977.

Kanouse, D. E., & Hanson, L. R. Negativity in evaluations. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior.* Morristown, N.J.: General Learning, 1972.

Kelley, H. H. Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 15). Lincoln: University of Nebraska Press, 1967.

Kelley, H. H. Causal schemata and the attribution process. In E. E. Jones et al. (Eds.), *Attribution: Perceiving the causes of behavior.* Morristown, N.J.: General Learning Press, 1972.

Kintsch, W. *The representation of meaning in memory.* Hillsdale, N.J.: Erlbaum, 1974.

Kintsch, W. Memory representations of text. In R. L. Solso (Ed.), *Information processing and cognition.* Hillsdale, N.J.: Erlbaum, 1975.

McArthur, L. The how and what of why: Some determinants and consequences of causal attribution. *Journal of Personality and Social Psychology,* 1972, *22,* 171–193.

Norman, D. A., Rumelhart, D. E., & the LNR Research Group. *Explorations in cognition.* San Francisco: Freeman, 1975.

Pryor, J. B., & Kriss, M. The cognitive dynamics of salience in the attribution process. *Journal of Personality and Social Psychology,* 1977, *35,* 49–55.

Schank, R. *Conceptual information processing.* Amsterdam: Elsevier, 1975.

Schneider, W., & Shiffrin, R. M. Controlled and automatic human information processing: I. Detection, search, and attention. *Psychological Review,* 1977, *84,* 1–66.

Shiffrin, R. M., & Schneider, W. Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review,* 1977, *84,* 127–190.

Taylor, D. A. Stage analysis of reaction time. *Psychological Bulletin,* 1976, *83,* 161–191.

Taylor, S. E., Crocker, J., Fiske, S., Sprinzen, M., & Winkler, J. The generalizability of salience effects. *Journal of Personality and Social Psychology,* 1979, *37,* 356–367.

Taylor, S. E., & Fiske, S. T. Point of view and perceptions of causality. *Journal of Personality and Social Psychology,* 1975, *32,* 439–445.

Taylor, S. E., & Fiske, S. T. Salience, attention, and attribution: Top of the head phenomena. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 11). New York: Academic Press, 1978.

Tversky, A., & Kahneman, D. Judgment under uncertainty: Heuristics and biases. *Science,* 1974, *185,* 1124–1131.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1970.

Received December 11, 1978 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 12, 2253-2261

Effects of "Voice" and Peer Opinions on Responses to Inequity

Robert Folger, David Rosenfield, Janet Grove, and Louise Corkran
Southern Methodist University

Two experiments investigated the effects of "voice" (participating in allocation decision making by expressing one's own opinion about the preferred allocation) on responses to an inequitable allocation. In addition to subjects' (female college students) either having or not having voice, Experiment 1 manipulated (a) whether the allocation made by a "decision maker" (supposedly another subject but actually the experimenter) was or was not biased (due to self-interest) and (b) whether the subject did or did not learn that a "co-worker" believed the allocation to be inequitable. Experiment 2 (with female high school students) manipulated the presence or absence of voice and involved only a self-interested decision maker; also, a note from a co-worker either supported the decision maker's allocation or confirmed the subject's opinion that the allocation was inequitable. In both experiments, the impact of voice was mediated by knowledge about the co-worker's opinion. When subjects had no knowledge of the co-worker's opinion (Experiment 1) or knew that the co-worker's opinion coincided with the decision maker's allocation (Experiment 2), there was evidence for a "fair process effect": Voice subjects expressed greater satisfaction than those with no voice.

How do people know that they have been treated fairly? According to equity theory (Adams, 1965; Walster, Berscheid, & Walster, 1973), a distribution of outcomes is considered fair (equitable) if the ratio of outcomes to inputs is constant across people. Apart from considerations of equity, however, fairness judgments may also be affected by whether a distribution is the result of an acceptable decision-making procedure (see the distinction between distributive and procedural justice in Folger, 1977; Leventhal, 1976; and Thibaut & Walker, 1975). Deutsch (1975), in discussing how "injustice of decision-making procedures" affects the perception of justice, makes the following argument: "There is much social psychological research which would suggest that . . . [procedural] injustice is the most fundamental. The research to which I am referring indicates that people are more apt to accept decisions and their consequences if they have participated in making them" (p. 139).

Indeed, many classic studies support Deutsch's position. Lewin, Lippitt, and White (1939), for example, found that the unity of a group was greater under democratic leaders than under autocratic or laissez-faire leaders. Similarly, Shaw (1955) found group members to be more satisfied under nonauthoritarian than under authoritarian leaders. Another classic example is Leavitt's (1951) research on communication networks, which found that the satisfaction of group members increased as the centrality of their position (opportunity to influence decisions) increased. Reviewing related research, Lawler (1975) concluded that workers' participation in decision making can lead to both greater worker satisfaction and higher productivity.

There is, however, an ambiguity in this research. Positive feelings may reflect the *consequences* of participation rather than par-

Preparation of this manuscript was facilitated by a National Science Foundation Postdoctoral Fellowship to the first author and by National Institute of Mental Health Grant 30968-01 to the first two authors. Philip Brickman graciously provided comments on an earlier draft.

Requests for reprints should be sent to Robert Folger, Department of Psychology, Southern Methodist University, Dallas, Texas 75275.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3712-2253$00.75

ticipation per se (i.e., increased participation in decision making might ordinarily lead to people actually obtaining more favorable outcomes). Although those given an opportunity to make their preferences known can hope to influence the outcome, it is clear that they may at times be unsuccessful. It is important, therefore, to distinguish between (a) the response to an opportunity to exercise a "voice" in decision making (cf. Folger, 1977; Hirschman, 1970)[1] and (b) the response to the outcome of having exercised voice. A key issue is whether voice is in itself sufficient to induce greater acceptance of decisions that result in inferior outcomes.

Thibaut and Walker (1975) provide an encouraging answer to this question in suggesting that the opportunity to present evidence supporting one's own case (voice) has a remarkable effect on a defendant's satisfaction with the verdict. Their research has demonstrated (1975; see also LaTour, 1978; Walker, LaTour, Lind, & Thibaut, 1974) that the more such voice is available, the more an otherwise intolerable outcome (e.g., a guilty verdict imposed on those who think themselves innocent) becomes relatively acceptable.

We will refer to Thibaut and Walker's pattern of results as the *fair process effect*. This effect refers to cases in which greater satisfaction results from giving people a voice in decisions. Thibaut and Walker's results suggest that satisfaction with the procedure may also generalize to other aspects of the situation (e.g., the distribution of outcomes). Our investigation concerns the generality of the fair process effect, and we begin by examining an experiment that produced the opposite results —a pattern that we will term the *frustration effect*.

In a study by Thibaut, Friedland, and Walker (1974), each of four subjects was assigned the role of a "corporation" in an industrial simulation. A fifth participant, actually a confederate, was assigned the role of a "government" empowered to collect taxes according to one of two fiscal policies, one being fair and the other being inequitable (the following account pertains only to inequitable conditions). In the "participation" condition, the fiscal policy was allegedly determined by a vote of all participants (a voice procedure). Bogus feedback about this vote showed that two corporations had voted for the equitable policy and two for the inequitable policy and that the government had cast the tie-breaking vote that established the inequitable policy. In the "no-participation" condition, the government decided on the inequitable policy without consulting the corporations.

The results showed that the proportion of taxes paid to taxes owed was approximately twice as high among no-participation subjects as among participation subjects (i.e., those who had voice were *more* likely to express discontent with the inequitable policy by cheating on their taxes). Thibaut et al. explained these results, which are clearly opposite from a fair procedure effect, by suggesting that discontent is intensified whenever a participative process raises hopes for a desired outcome and those hopes are dashed.

Thus the proposed mechanism whereby voice can cause heightened displeasure is increased frustration due to raised expectations (hence the term *frustration effect*). But if voice enhances frustration when outcomes are inferior, why was the opposite, fair process, effect evident in the LaTour (1978) and Walker et al. (1974) studies, which involved unfavorable outcomes? One possible answer (Thibaut, Note 1) is that although the decision maker in these latter studies was a neutral or "disinterested" third party, the government in the Thibaut et al. (1974) study was clearly voting for its own self-interest. However, Folger (1977) found a fair process effect under some conditions in which the

[1] The original use of the term *voice* by Hirschman (1970) was in the context of alternative responses to declining outcomes. At the same time, however, a more general usage was suggested by passages such as the following (pp. 16, 30): "Voice is political action par excellence"; "voice is nothing but a basic portion and function of any political system, known sometimes as 'interest articulation.'" It is in this sense of interest articulation that we use the word *voice* in place of longer locutions such as "opportunity to express opinions and preferences or to present facts relevant to one's position in the context of decision making."

allocator did benefit directly from this inequity, as was the case in the Thibaut et al. study. Since this result demonstrates a fair process effect even when the allocator's self-interest could have biased the decision, it appears that some other aspect of the Thibaut et al. study was responsible for the frustration effect.

We suggest that the frustration effect in the study of Thibaut et al. was due to the feedback subjects did or did not receive about the other subjects' opinions. Since no vote was allowed in the no-participation condition, subjects in this condition did not know the fiscal policy the other corporations favored. In contrast, each subject in the participation condition learned that in addition to himself, one other corporation had opposed the inequitable policy; thus these subjects had social support for their opinions.

The importance of such social support is highlighted by Asch's (1952) conformity studies. When the majority's unanimity was broken by a single confederate, the ordinarily substantial rate of conformity was markedly reduced. Perhaps a similar phenomenon occurred when participation subjects learned that one other corporation agreed with their own opinion: This support from a solitary peer may have reduced these subjects' tendencies to conform to the government's opinion, thereby leaving them more dissatisfied with the inequitable result. Unfortunately, since voice and feedback about others' opinions were confounded in the Thibaut et al. experiment, it is impossible to determine which factor was responsible for the obtained effect. The present line of investigation involves two experiments designed to shed light on this issue by manipulating procedure and feedback independently of one another.

It is predicted that voice will interact with the feedback the subject receives concerning the other subject's opinion. The exercise of voice should lead to more positive responses when subjects do not receive any feedback from others supporting their own opinion that the decision was unfair, conditions similar to those that yielded a fair process effect in the LaTour (1978) and Walker et al. (1974) studies. However, when subjects receive information from others that confirms their beliefs and indicates that the final decision was inequitable (as in the Thibaut et al. study), then voice may not lead to a fair process effect. Additionally, Experiment 1 will include a manipulation of whether the decision maker is biased or unbiased. As noted above, Thibaut has suggested that voice may not lead to a fair process effect if the decision maker has some external inducement that might have biased his or her decision.

Experiment 1

Method

Subjects. The subjects were 82 female undergraduates enrolled in an introductory psychology course. They received extra credit for participating in the experiment. Although the gender was restricted primarily for convenience, it should be noted that previous research has suggested that females are more likely than males to prefer an equal division of outcomes (e.g., Leventhal & Lane, 1970).

Procedure. The participants, scheduled in groups of three, were seated in separate rooms and listened to taped instructions. They were told that the study was simulating business decision making and that a random drawing would assign subjects to the roles of "decision maker" and "workers" (each person was actually assigned the worker role). Workers were to make small words from the letters of a larger word and to help evaluate the decision maker's effectiveness. There would be two sessions of four word-making trials, and on each trial the decision maker would select from a list of words a subset that would be given to the two workers (making the selection so that the number of possible smaller words would be maximized). The subjects were also told that they would receive lottery tickets for a $50 prize as additional compensation. Nine tickets were to be distributed during each of the two work sessions, with three of those tickets going to the decision maker. The decision maker would decide how to divide the remaining six tickets between the two workers.

Subjects in the *unbiased* decision-maker condition were told that whoever drew the role of decision maker would keep that role for both work sessions, which meant that she would receive the same number of tickets regardless of how she chose to allocate the workers' tickets. In the *biased* condition, the person initially chosen as decision maker would have that role for one session only. During the second session, one of the workers would become the new decision maker and would be in a position to determine how many tickets the original, first-session decision maker would receive during the second session. Thus, the decision maker's ticket allocation during the first

session might be biased because she would have an extrinsic reason to favor one worker (the one who would be the next decision maker) over the other.

In all conditions, the drawing was rigged so that the subject believed that she would be a worker (Worker B) during both sessions. In the biased condition, she also believed that the other worker for the first session (Worker A) would be the decision maker during the second session. Note that subjects thought the initial decision maker knew which first session worker would be the decision maker during the second session (Worker A).

After the first session, every subject was asked to complete ("for our records") an "opinion card" on which she wrote her ideas of the fair way to divide the lottery tickets for that session. In the *voice* conditions, the experimenter then added the following statement:

> Even though the decision maker will be making the final decision about how the lottery tickets will be divided, one of you workers will have an opportunity to let her know what you think is fair before she makes her decision. We will have a drawing and the worker who wins the drawing will get to have her opinion card shown to the decision maker.

Subjects in the *mute* (no voice) condition were not given these additional instructions.

Before the subject filled out her opinion card, the experimenter counted the total number of smaller words that she had made (supposedly to be reported to the decision maker) and told the subject how many words the other worker had allegedly formed (one more than the subject's total). The experimenter then left, taking the opinion card with her in the voice conditions. The experimenter administered the feedback manipulation upon her return. Subjects in the *inequity-confirmed* conditions were told that the co-worker had suggested an even split of the six tickets, whereas the decision maker had chosen to give that co-worker four of the tickets and to give the subject only two. The information about the co-worker's opinion thereby reinforced the subject's feelings that a 3–3 split was fair and confirmed that the 4–2 decision was inequitable. Subjects in the *no-feedback* conditions were only told about the decision maker's allocation (a 4–2 split favoring the other worker). It should be noted that in the biased condition, the 4–2 split favored the worker who was to become the decision maker during the next session. Thus that decision could apparently have been due to the decision maker's desire to be treated favorably during the second session.

Subjects were then given a questionnaire to evaluate the effectiveness of the decision maker. Included were items concerning the subject's feelings about the fairness of the allocation and about the fairness of the decision-making process, each answered on a scale anchored by "very unfair" (1) and "very fair" (11). Another item assessed subjects' general satisfaction with the work situation on a scale ranging from "very dissatisfied" (1) to "very satisfied" (11). Finally, two questions asked the subjects to rate their decision maker on scales of "incompetent" (1) to "competent" (11) and "inefficient" (1) to "efficient" (11). Answers to these two items were summed to obtain an average rating of the decision maker.

After completing the questionnaire, the subjects were told that there was not enough time left to finish the experiment. The experimenter then attempted to assess any suspicions that the subjects had and debriefed them. The subjects were also told that they would each receive three tickets for the lottery. A lottery drawing actually took place at the end of the semester, and the winner was given her $50.

Results and Discussion

The results were analyzed by a $2 \times 2 \times 2$ multivariate analysis of variance. This MANOVA revealed that the Feedback (No Feedback vs. Inequity-Confirmed) × Procedure (Voice vs. Mute) interaction was significant, $F(4, 71) = 2.77$, $p < .05$. Also, the procedure main effect was marginally significant, $F(4, 71) = 2.07$, $p < .09$. Given these results, univariate analyses of variance (ANOVAs) were examined to determine the nature of these effects.

The univariate ANOVAs showed significant procedure main effects on the fairness of the process, the ratings of the decision maker, and the fairness of the allocation, $F(1, 74) = 7.45$, $p < .01$; $F(1, 74) = 4.30$, $p < .05$; and $F(1, 74) = 5.72$, $p < .05$, respectively. As Table 1 shows, subjects in voice conditions generally expressed more positive feelings than those in the mute conditions. Note, however, that the mute–voice differences are always larger within the no-feedback condition; these differences are generally quite minimal in the inequity-confirmed condition and are even reversed slightly on the satisfaction measure.

In fact, the procedure main effect is qualified by the significant Feedback × Procedure interaction, indicating that the main effect was due almost entirely to the large differences within the no-feedback condition. This interaction was significant on the univariate analyses of the fairness of the decision-making process and the subject's satisfaction with the work situation, $F(1, 74) = 5.34$, $p < .05$, and $F(1, 74) = 4.89$, $p < .05$,

Table 1
Responses to Inequity as a Function of Procedure and Feedback, Experiment 1

Measure	Inequity confirmed		No feedback	
	Mute	Voice	Mute	Voice
Fairness of the process	6.7	7.0	5.6	8.8
Fairness of the decision	4.9	6.4	6.0	7.8
Satisfaction	8.1	7.7	7.6	9.5
Ratings of decision maker*	14.6	15.3	15.0	18.4

* Higher numbers indicate more positive ratings of the decision maker.

respectively. As Table 1 demonstrates, voice did not lead to substantially more positive reactions than did mute in the inequity-confirmed conditions ($p > .50$ for each of the four dependent measures), but it did lead to substantially more positive feelings for those subjects who were given no feedback and hence were not as convinced that they had been treated inequitably ($p < .05$ for fairness of the process and satisfaction with the work situation).

This experiment thus substantiates that voice sometimes tends to mitigate the distress associated with inequitable treatment; however, the significant Feedback × Procedure interaction indicates that voice alleviates distress only when the subjects do not receive a second opinion from their co-workers suggesting that the co-workers also feel that the decision was inequitable. When such an opinion is received (which reinforces the belief that the decision was unfair), voice does not make the subjects feel that the inequitable decision was any more satisfying.

There was little evidence of any strong tendency for knowledge of a supportive opinion to make voice any *less* satisfying than the mute procedure, as it did in the Thibaut et al. study. However, it is possible that increased displeasure due to voice may be unique to situations in which decision makers stand to benefit directly from the allocation, as they did in the Thibaut et al. study. In our study, the decision maker could benefit indirectly in the biased conditions (if giving more lottery tickets to the worker who would be the next allocator would increase the chances of being overrewarded by her in return), but the indirect nature of this advantage may have reduced its salience, as is reflected in the absence of any significant differences between the biased and unbiased conditions. Since Experiment 2 placed the decision maker in a position to benefit directly from reducing the payments to other people, further discussion of frustration effects will be postponed until after the presentation of that experiment.

Experiment 2

Method

Subjects. In response to telephone solicitations, 61 female high school students participated. Each understood that she would receive at least the minimum wage for participating for 1 hour in a study about "working conditions."

Procedure. Experiment 2 was conducted similarly to Experiment 1, with several exceptions. First, there was no biased–unbiased manipulation (Experiment 2 used biased conditions only). During each work session, the decision maker supposedly had $3.00 to divide between herself and the two workers. Thus, by giving the workers less (her decision was allegedly to give the workers $.75 each), she was able to keep more for herself ($1.50). In keeping with this procedure, it was further announced that the decision maker would remain in that role for the entire experiment (three sessions).

Another difference was that the feedback manipulation was stronger. In the *inequity-confirmed* conditions, subjects were told that their co-worker had indicated on her opinion card that an even split of the $3.00 was fair. Since the decision maker supposedly gave the workers only $.75 while keeping $1.50 for herself, this feedback confirmed the subjects' feelings that the decision was inequitable. In the *inequity-disconfirmed* conditions, the subjects again received information about their co-worker's opinion, but this time the co-worker supposedly had indicated that $.75 for the workers and $1.50 for the decision maker was fair. Another difference between the first and second studies was that the decision maker in the latter was not told how many words the subjects had formed. Thus her decision was made (allegedly) without taking productivity into account.

The final difference involved dependent measures. In addition to the measures used in Experiment 1, Experiment 2 also asked subjects to indicate how certain they were that the division of money that they had written down on their opinion card (before learning about the decision maker's allocation or

Table 2
Responses to the Inequity as a Function of Procedure and Feedback, Experiment 2

Measure	Inequity confirmed		Inequity disconfirmed	
	Mute	Voice	Mute	Voice
Fairness of the process	5.5	4.9	6.2	8.7
Fairness of the decision	3.7	2.7	5.5	6.0
Satisfaction	7.3	6.7	7.1	7.7
Ratings of decision maker[a]	13.1	14.5	16.2	15.7
Payment to decision maker	79.7	70.3	85.7	107.0
Certainty	9.9	9.4	9.1	6.3

[a] Higher numbers indicate more positive ratings of the decision maker.

their co-worker's opinion) was the fair way to split the money, on an 11-point scale from "not at all certain" to "very certain" (57 out of 61 had originally indicated that an equal split was fair). Another additional measure was assessed just after the second session. Subjects were told that there was not enough time to complete the third session, so the experiment would have to be terminated. This problem had supposedly come up before, however, and in those cases it had been decided that since the decision maker was not selecting any words for the last session, there was no reason why the decision maker should distribute the money for that session. Thus subjects were told that one of the workers would decide how to divide the money for that session, and a drawing took place to decide which worker would make the decision, arranged so that the subject always won. The subject was then asked to make the allocation, and the amount given to the decision maker was used as a behavioral measure of her satisfaction with the decision maker.

Results and Discussion

The results were initially analyzed by a 2 × 2 MANOVA. This MANOVA indicated that both main effects and the interaction were significant, $F(6, 51) = 2.46$, $p < .05$, for the procedure main effect; $F(6, 51) = 4.35$, $p < .005$, for the feedback main effect; and $F(6, 51) = 2.61$, $p < .05$, for the Feedback × Procedure interaction. Univariate ANOVAs were then examined to determine the nature of these effects.

The Feedback × Procedure interaction was either significant or marginally significant on the fairness of the decision-making process ($p < .05$), the amount of money given to the decision maker ($p < .06$), and the subjects' certainty about their original opinion of the fairness of an equal division ($p < .05$). The cell means (see Table 2) and appropriate contrasts show that voice led to more positive feelings (in the form of more bonus money given to the allocator, greater perceptions of fairness, and less certainty that an equal split was fair) when the feedback from the other worker indicated that the decision was not inequitable (inequity-disconfirmed conditions, $ps < .05$), but that mute–voice differences tended to be slightly (although nonsignificantly) in the opposite direction in the inequity-confirmed conditions.

The feedback main effect was also significant on each of these measures, as well as on the subjects' ratings of the decision maker and the fairness of the decision ($ps < .05$). The cell means show that subjects were generally more content with their low outcome when co-worker's feedback indicated that there was no inequity than when it confirmed that the inequity existed. It is important to note that this difference was substantial only in the voice conditions, however, and was usually very small (or even slightly in the opposite direction) in the mute conditions.

Finally, the procedure main effect was significant only on the certainty measure ($p < .005$). Voice subjects were significantly less certain that an even split was fair, but consistent with the interaction, this was significant only in the inequity-disconfirmed condition ($p < .05$).

The obtained main effect for feedback is not surprising, given the evidence that a person's discontent with a leader will be affected by whether a fellow member of the group endorses or expresses displeasure with the leader

(Michener & Tausig, 1971; Michener & Lyons, 1972). What is striking, however, is the nature of the Feedback × Procedure interaction: Contrary to the trend within the inequity-confirmation conditions, inequity-disconfirmation subjects displayed more discontent under mute than under voice procedures. The latter result is noteworthy since it means voice has a mitigating effect on discontent over and above that due to a disconfirmation of one's own opinion. Voice–disconfirmation subjects were not only *less* distressed than mute–disconfirmation subjects about their inequitable treatment but also seemed even somewhat pleased in an absolute sense. Voice–disconfirmation subjects' ratings of the fairness of the process averaged nearly 9 on an 11-point scale, their average allocation to the decision maker was more than $1 of the available $3, and the certainty measure showed that voice–disconfirmation subjects even came to entertain some doubts about their initial judgment of what was fair.

General Discussion

The results from both of these experiments display a similar pattern, in which the voice procedure is associated with more positive affect than the mute procedure is under some circumstances (no-feedback conditions, Experiment 1; inequity-disconfirmation conditions, Experiment 2), whereas this tendency is virtually neutralized (Experiment 1) or even slightly reversed (Experiment 2) under other circumstances (inequity-confirmation conditions). This pattern is contrary to the commonsense prediction that a voice procedure should always produce less discontent than a mute procedure would because it is fairer, but the results are consistent with our contention that the positive impact of voice tends to be neutralized when the viewpoint expressed by a fellow recipient confirms one's own opinion that the allocation was unfair. An understanding of why such a confirmation mitigates the desirability of voice can be gained by first considering why voice should be desirable.

Voice may be desirable for two reasons. First, voice may be preferable to mute procedures because the latter are based on incomplete information. A mute procedure may not take into account the claims of disputants and hence may entail an inferior, or at least suspect, decision. Second, being given a voice in the decision is considered to be a fairer procedure than being given no voice. Someone who is given voice at least has a chance to defend his/her position and present his/her side of the issue. The general preference for voice over mute procedures (Houlden, LaTour, Walker, & Thibaut, 1978; LaTour, 1978; Thibaut, Walker, LaTour, & Houlden, 1974; Walker et al., 1974) is so strong that disputants with the preponderance of evidence in their favor prefer a voice procedure over a mute procedure even though it may allow their opponents to present claims more strongly and could hence promote their opponents' position. This preference is shared by the decision makers themselves when they are given absolute authority and thus might "efficiently" resolve the dispute without having to consult the disputants.

Given the desirability of voice, it might be expected to affect positively the acceptability of a decision even when that decision provides undesirable outcomes. Our results show this to be the case under two conditions: when people do not know whether anyone else would agree that they have been cheated (Experiment 1) and when they discover that a co-worker's opinion disconfirms their belief that they have been cheated (Experiment 2). Thus, when subjects were not certain that they had been treated inequitably, the voice procedure was rated fairer than the mute procedure and there was a more favorable reaction to the decision maker and to the entire situation.

On the other hand, when people learn that someone else agrees that they have been denied their just deserts (inequity-confirmation conditions), the positive impact of voice seems to be negated; indeed, our results show that when supportive social "evidence" is available, the fairness of the allocation procedure becomes essentially irrelevant. Presumably the raison d'être of having a fair procedure is to prevent inequities by improving the quality of information on which the allo-

cation decision is based, but this possible advantage is obviously not present when it is *clear* that an injustice has been done. Thus, although the voice procedure could have potentially been fairer than the mute procedure, when the voiced opinions seem to have been completely ignored such a procedure is no longer perceived to be any better than a mute procedure. Support for this view comes from our "fair process" measure, which showed that subjects in the inequity-confirmation conditions did *not* rate the voice procedure as any fairer than the mute procedure.

Although the overall pattern of results does show that some circumstances are more likely than others to promote the fair process effect, the evidence regarding when voice procedures will lead to greater *discontent* (the frustration effect) is less conclusive. Frustration effect tendencies appeared only in Experiment 2, where they were nonsignificant. The main reason that the present frustration effect tendencies were less pronounced than those in the Thibaut et al. (1974) experiment may be that voice and inequity-confirmation were confounded in the Thibaut et al. study, whereas they were manipulated independently in the present experiment. The discontent expressed by participation (voice) subjects in the Thibaut et al. study may have been due to the fact that the voice subjects were given another opinion that confirmed their feelings that the decision was inequitable. Subjects in their no-participation (mute) conditions did not receive any such supporting opinion. Since opinions confirming the inequity can substantially increase discontent (as shown in Experiment 2 and in Michener & Lyons, 1972, and Michener & Tausig, 1971), this confirming opinion may have been responsible for the obtained frustration effect in their study.

We should also point out, however, that other studies have at times found frustration effects. For example, Austin, Williams, Worchel, Wentzel, and Siegel (Note 2) found that the provision of voice in a mock trial led to greater discontent when the verdict was highly unfavorable but led to greater satisfaction when the verdict was favorable. Furthermore, the provision of voice under circumstances in which outcomes improve after voice (although cumulative outcomes remain inequitable) has been observed to create a frustration effect in at least two instances (certain conditions of an experiment by Folger, 1977, and an earlier study by Thibaut, 1950). These results indicate that the frustration effect found by Thibaut et al. may not have been due solely to the confound between voice and inequity-confirmation. This possibility, coupled with our finding that the positive effects of voice can be neutralized under certain circumstances, indicates the need for further study of the conditions under which voice can affect one's satisfaction either positively or negatively. Field studies, where involvement level and applicability are high, would be especially helpful.

Reference Notes

1. Thibaut, J. Personal communication, October 25, 1976.
2. Austin, W., Williams, T. A. III, Worchel, S., Wentzel, A. A., & Siegel, D. *Effect of mode of adjudication, presence of defense counsel, and favorability of verdict on observers' evaluation of a trial proceeding: An empirical study of procedural justice.* Unpublished manuscript, 1978. (Available from William Austin, Department of Psychology, Gilmer Hall, University of Virginia, Charlottesville, Virginia 22901.)

References

Adams, J. S. Inequity in social exchange. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2). New York: Academic Press, 1965.

Asch, S. *Social psychology*. New York: Prentice-Hall, 1952.

Deutsch, M. Equity, equality, and need: What determines which value will be used as the basis of distributive justice? *Journal of Social Issues*, 1975, *31*, 137–150.

Folger, R. Distributive and procedural justice: Combined impact of "voice" and improvement on experienced inequity. *Journal of Personality and Social Psychology*, 1977, *35*, 108–119.

Hirschman, A. O. *Exit, voice and loyalty: Responses to declines in firms, organizations, and states.* Cambridge, Mass.: Harvard University Press, 1970.

Houlden, P., LaTour, S., Walker, L., & Thibaut, J. Preference for modes of dispute resolution as a function of process and decision control. *Journal of Experimental Social Psychology*, 1978, *14*, 13–30.

LaTour, S. Determinants of participant and observer satisfaction with adversary and inquisitorial modes

of adjudication. *Journal of Personality and Social Psychology*, 1978, *36*, 1531–1545.

Lawler, E. E. III. Pay, participation and organizational change. In E. L. Cass & F. G. Zimmer (Eds.), *Man and work in society*. New York: Van Nostrand Reinhold, 1975.

Leavitt, H. J. Some effects of certain communication patterns on group performance. *Journal of Abnormal and Social Psychology*, 1951, *46*, 38–50.

Leventhal, G. S. *Fairness in social relationships*. Morristown, N.J.: General Learning Press, 1976.

Leventhal, G. S., & Lane, D. W. Sex, age, and equity behavior. *Journal of Personality and Social Psychology*, 1970, *15*, 312–316.

Lewin, K., Lippitt, R., & White, R. Patterns of aggressive behavior in experimentally created "social climates." *Journal of Psychology*, 1939, *10*, 271–299.

Michener, H. A., & Lyons, M. Perceived support and upward mobility as determinants of revolutionary coalitional behavior. *Journal of Experimental Social Psychology*, 1972, *8*, 180–195.

Michener, H. A., & Tausig, M. Usurpation and perceived support as determinants of the endorsement accorded formal leaders. *Journal of Personality and Social Psychology*, 1971, *18*, 364–372.

Shaw, M. A comparison of two types of leadership in various communicants. *Journal of Abnormal and Social Psychology*, 1955, *50*, 127–134.

Thibaut, J. An experimental study of the cohesiveness of underprivileged groups. *Human Relations*, 1950, *3*, 251–278.

Thibaut, J., Friedland, N., & Walker, L. Compliance with rules: Some social determinants. *Journal of Personality and Social Psychology*, 1974, *30*, 782–801.

Thibaut, J., & Walker, L. *Procedural justice: A psychological analysis*. Hillsdale, N.J.: Erlbaum, 1975.

Thibaut, J., Walker, L., LaTour, S., & Houlden, P. Procedural justice as fairness. *Stanford Law Review*, 1974, *26*, 1271–1290.

Walker, L., LaTour, S., Lind, E. A., & Thibaut, J. Reactions of participants and observers to modes of adjudication. *Journal of Applied Social Psychology*, 1974, *4*, 295–310.

Walster, E., Berscheid, E., & Walster, G. W. New directions in equity research. *Journal of Personality and Social Psychology*, 1973, *25*, 151–176.

Received December 30, 1978 ■

New JPSP Format Delayed Until April

Recent issues of this journal carried an announcement that the *Journal of Personality and Social Psychology* would appear in January 1980 under new policies and in a new sectioned format. However, in order to publish the backlog of manuscripts accepted under present policies, the first three issues of 1980 will continue to carry articles accepted under Clyde Hendrick's Acting Editorship.

This journal will appear in April as a sectioned journal edited by Melvin Manis, Ivan D. Steiner, and Robert Hogan, as previously announced.

Robert S. Daniel, Chair, APA
Publications and Communications Board

Journal of Personality and Social Psychology
1979, Vol. 37, No. 12, 2262-2277

Interpersonal Attraction and the Polarity of Similar Attitudes: A Test of Three Balance Models

Abbas Tashakkori
Shiraz University, Shiraz, Iran

Chester A. Insko
University of North Carolina

In a variation of Byrne's anonymous-stranger technique, a PDP-11 computer was used to provide feedback on five attitude issues for each of the cells of a design created by three levels of the *p-x* variable (−2, 0, +2) and three levels of the *o-x* variable (−2, 0, +2). Measurements of *p-o* sentiment were used to test the adequacy of three models for the quantification of balance: the equal-weights tetrahedron model, the unequal-weights tetrahedron model, and the Feather model. Correlations of obtained and predicted measures for the three models were .48, .51, and .52, respectively, with various problems noted for each model. The results further revealed a predicted shift in the polarity of similar attitudes, but a predicted difference between ambivalent and indifferent neutral attitudes was not supported.

Byrne (1969, 1971) has reported a sizable amount of evidence indicating that pre-acquaintance attraction is a linear function of the proportion of similar attitudes. Similarity is defined by Byrne as "any response on the same side of the neutral point as the subject's response, and dissimilarity as any response on the opposite side of the neutral point" (1971, p. 75). As Byrne is aware, such a definition of similarity ignores the polarity of the attitude. Thus on a +3 to −3 scale, +3,+3 agreement should produce as much attraction as +1,+1 agreement. In his own words, "In effect, each item has been treated as if it were a 2-point scale with mild, moderate, and strong feelings on a given side of an issue defined as equivalents" (1971, p. 75).

As a result of some research done by Nelson (1965), Byrne did expand his formula for predicting attraction so as to include the discrepancy between the subject's and the other's attitude (for a given side of the scale), as well as the proportion of attitudes on the same side of the scale. Nelson's research found that small discrepancy produced more attraction than did large discrepancy. According to Byrne, "If a subject is strongly committed to racial integration, for example, he not only prefers integrationists to segregationists, he prefers those strongly in favor (like himself) to those mildly in favor of integration" (1971, p. 77). This expansion of the formula, however, still did not take into account the polarity of the attitude. Byrne's formula predicts, for example, that two people mildly in favor of integration should be just as attracted as two people strongly in favor of intergration.

Consideration of the possible effect of polarity (the difference between +3,+3 agreement, and +1,+1 agreement, for example) leads fairly directly to concern with 0,0 agreement, or attraction between two people who agree in having a neutral attitude. Since Byrne uses 6-point scales (possibly to simplify his definition of similarity), the whole issue of agreement between two people with neutral attitudes does not arise. Nonetheless, agreement between two people with neutral attitudes does pose a theoretical issue of some interest.

There are two quantitative formulations of balance theory (Heider, 1946, 1958) that have been developed or extended by Wellens and Thistlethwaite (1971a) so as to imply that attraction increases with increased po-

Requests for reprints should be sent to Chester A. Insko, Department of Psychology, University of North Carolina, Chapel Hill, North Carolina 27514.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3712-2262$00.75

larity of similar (or agreeing) attitudes. These are Wiest's (1965) tetrahedron model, and Feather's (1967) model. As is explained below, the Feather model, as Wellens and Thistlethwaite have extended it, only predicts a change with a shift from 0 polarity, whereas the Wiest model predicts a change with every increment or decrement in polarity.

William Wiest had the ingenious idea of conceptualizing Heider's *p-o-x* triad as a cube. One dimension of the cube consists of the minus to plus values of the *p* to *o* sentiment relation, one the minus to plus values of the *p* to *x* sentiment relation, and one the minus to plus values of the *o* to *x* sentiment relation. Four of the corners of the cube represent balanced triads (the product of the three signs is positive), and four of the corners represent imbalanced triads. Connecting the balanced corners with straight lines creates a tetrahedron (a three-sided pyramid) inside the cube. Wellens and Thistlethwaite's (1971a, 1971b) representation of Wiest's three-dimensional space and tetrahedron is presented in Figure 1. In Figure 1 the *p-o* relation is symbolized as *Z*, the *p-x* relation as *X*, and the *o-x* relation as *Y*.

Wiest assumed that all balanced triads are on the surface of or within the tetrahedron. However, the assumption regarding balanced triads within the tetrahedron was modified by two further considerations. First, since points near the center of the tetrahedron are characterized by the absence of any valence for the three relations, they could better be described as "vacuously balanced" (1965, p. 3). Agreeing with Harary's (1959) postulated "tendency toward completeness," Wiest assumes that there is a force toward making "cognitive elements relevant to one another" (1965, p. 6) that moves points away from vacuous balance to balance (or from the center to the surface of the tetrahedron). The second consideration having a bearing upon the distribution of points (or triads) within the tetrahedron is also a postulated force—Osgood and Tannenbaum's (1955) assumed tendency toward maximum polarization of sentiments or attitudes. Osgood and Tannenbaum argued that attitudes have a tendency to drift toward simpler, extreme judgments

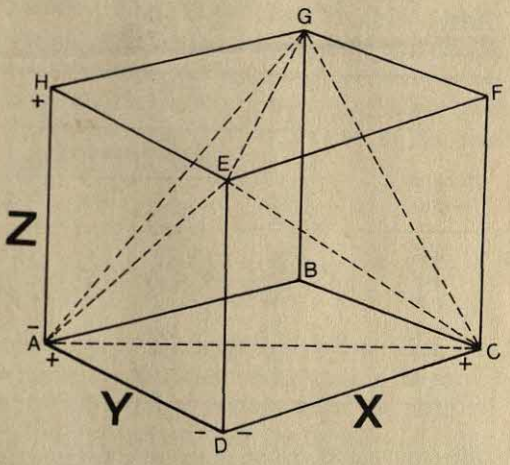


Figure 1. Wiest's three-dimensional space as defined by the three relations (or dimensions) *p-x* (*X*), *o-x* (*Y*), and *p-o* (*Z*). (From "An analysis of two quantitative theories of cognitive balance" by A. R. Wellens and D. L. Thistlethwaite, *Psychological Review*, 1971, *78*, 141–150. Copyright 1971 by the American Psychological Association. Reprinted by permission.)

and away from more difficult, discriminating judgments. Wiest agrees with this postulated tendency and assumes that it is a second force moving points toward the surface of the tetrahedron. Thus Wiest postulates that the balanced points are, by and large, on the surface of the tetrahedron.

We believe that Osgood and Tannenbaum's assumption of movement toward maximal polarity is more reasonable if the neutral attitudes, or sentiments, are characterized by ambivalence rather than indifference (cf. Kaplan, 1972). An ambivalent–neutral attitude is an imbalanced one and thus according to balance theory should be unstable. No special or additional assumption is required. In the case of indifferent attitudes, however, we are skeptical that there is any movement toward polarity. It should be further noted that if there is movement toward polarity on all three dimensions, the points should move to the corners of the tetrahedron and not just generally to the surfaces.

Wiest interpreted his tetrahedron model as having implications regarding the correlation between any two of the relations, holding the third constant at some value. The predictions were derived by "slicing" the tetrahedron at

Table 1
Predictions of the Upper- and Lower-Boundary Formulas When k = 2

p-x value	*o-x* value: Upper boundary −2	−1	0	+1	+2	*M*	Lower boundary −2	−1	0	+1	+2	*M*
+2	−2	−1	0	1	2	.0	−2	−1	0	1	2	.0
+1	−1	0	1	2	1	.6	−1	−2	−1	0	1	−.6
0	0	1	2	1	0	.8	0	−1	−2	−1	0	−.8
−1	1	2	1	0	−1	.6	1	0	−1	−2	−1	−.6
−2	2	1	0	−1	−2	.0	2	1	0	−1	−2	.0
M	.0	.6	.8	.6	.0		.0	−.6	−.8	−.6	.0	

different points along a given relation (or dimension) and then examining the expected scatterplot for the relation between the two remaining relations. Wiest tested the tetrahedron model by calculating certain correlations among the sociometric ratings of fifth, sixth, and seventh grade children. The children initially indicated their liking for each of the other children in the class, and then indicated the extent to which the highest, intermediate, and lowest of these rated children liked, or were liked by, each of the other children in the class. Wiest confirmed the prediction that the mean correlation between *p*'s liking of *o* and *o*'s liking of *q* (or *q*'s liking of *o*) should be largest for the most liked *q* and smallest for the least liked *q*.

Wellens and Thistlethwaite (1971a) extended Wiest's model so as to enable prediction from specified values for two of the relations to some value for the third relation. If we specify two values for X and Y on the "bottom" of the cube in Figure 1 and then project this point into the cube, the projection will intersect the tetrahedron at its lower boundary and will then emerge from its upper boundary. Wellens and Thistlethwaite present the following two equations for the predicted values on the upper and lower boundaries:

$$Z_u = k - |X - Y|$$

$$Z_L = |X + Y| - k,$$

"where Z_u represents the upper boundary limit for Z; Z_L represents the lower-boundary limit for Z; $|X - Y|$ and $|X + Y|$ correspond to the absolute values of the real difference and sum, respectively, of the two given relations; and the integer k is a constant representing the highest positive scale value possible on each of the three interelement relations" (p. 142).

Illustrative predictions for the upper- and lower-boundary equations are presented in Table 1. We have specified the dependent variable as the *p* to *o* relation (Z), and have assumed a 5-point plus-to-minus scale so that $k = 2$. Since our main present concern is with the polarity of similar attitudes, the most interesting values are in the diagonal running from lower left to upper right. Note that the upper-boundary formula predicts values that are the same (+2), that is, do not differ as a function of same-sign polarity. Except for the specification of the 0,0 case, this agrees with Byrne's assumption. As Table 1 also indicates, however, the lower-boundary formula predicts a very marked effect for same-sign polarity.

Wellens and Thistlethwaite present a "generalized tetrahedron formulation" that predicts a single value for Z,

$$Z = aZ_u + bZ_L,$$

where a and b are weights for the range limits. Wellens and Thistlethwaite point out that the values for a and b will vary according to the assumptions made regarding the distribution of points between the upper and lower boundaries. Wiest assumed that the points tend to go to the surface of the tetrahedron. If it is also assumed that the points drift equally toward the two boundaries, we have a further

basis for what Wellens and Thistlethwaite refer to as an "equal-weights" tetrahedron model in which a and b are both assigned values of .5. Other distributions of points are, of course, possible, and it is worth noting that any symmetrical distribution of points could provide justification for the equal-weights model. To the extent that the distribution of points is skewed, one boundary will be weighted more heavily than the other. It does appear implicit in Wellens and Thistlethwaite's formulation that $a + b = 1$. Otherwise Z would not be an average of Z_U and Z_L.

Illustrative predictions for the equal-weights tetrahedron model are given in Table 2. Note in particular that the values along the lower left to upper right diagonal show an effect for polarity. As long as some weight is given the lower-boundary formula, the generalized tetrahedron model will predict greater attraction with greater same-sign polarity. Wellens and Thistlethwaite (1971a) found that their data were best fit by a model in which the upper boundary was weighted .75 and the lower .25. As the predicted values for this unequal-weights model show (see Table 2), there is a clear effect for same-sign polarity.

The second approach to the quantification of balance theory flows from Feather's (1967, 1971) discrepancy principle. This principle states that when two relations are of the same sign, small discrepancies will be associated with a strong positive, third relation, and when two relations are of opposite sign, large discrepancies will be associated with a strong negative, third relation. Feather (1971) wrote simple formulas for predicting the missing relation when the signs of the known relations are the same or opposite. Wellens and Thistlethwaite's (1971a) version of the two formulas is as follows:

$$Z_s = k - |X - Y|$$

$$Z_o = \frac{|X - Y|}{-2},$$

where Z_s is the missing relation when the signs of the two known relations are the same, Z_o is the missing relation when the signs of the two known relations are opposite, k is the highest possible scale value, and X and Y are the two known relations. When either X or Y is 0, Wellens and Thistlethwaite recommend averaging the values for Z_s and Z_o. Such a procedure seems more reasonable for ambivalent–neutral attitudes than for indifferent–neutral attitudes.

Illustrative predictions for the Feather model are given in Table 3. Note that the same-design diagonal values show an effect for polarity only at the 0,0 point. Since the same-sign formula is identical to the upper-boundary formula (which, however, also applies to opposite signs), it is evident that the polarity shift at the 0,0 point is due to the formula for opposite signs.

Wellens and Thistlethwaite (1971a) used a role-playing procedure to test the implications of the various models for p to o attraction. The result indicated that although all models provided a significant fit to the data,

Table 2
Predictions of the Equal-Weights and Unequal-Weights Tetrahedron Models When k = 2

	o-x value											
	Equal weights						Unequal weights					
p-x value	−2	−1	0	+1	+2	*M*	−2	−1	0	+1	+2	*M*
+2	−2	−1	0	1	2	.0	−2	−1	0	1	2	.0
+1	−1	−1	0	1	1	.0	−1	−.5	.5	1.5	1	.3
0	0	0	0	0	0	.0	0	.5	1	.5	0	.4
−1	1	1	0	−1	−1	.0	1	1.5	.5	−.5	−1	.3
−2	2	1	0	−1	−2	.0	2	1	0	−1	−2	.0
M	.0	.0	.0	.0	.0	—	.0	.3	.4	.3	.0	—

Note. For equal weights $a = .5$, $b = .5$. For unequal weights, $a = .75$, $b = .25$.

Table 3
Predictions of the Feather Model When Neutral Attitudes are a Composite (Average) of the Same-Sign and Opposite-Sign Predictions and k = 2

p-x value	*o-x* value −2	−1	0	+1	+2	*M*
+2	−2	−1.5	−.5	1	2	−.20
+1	−1.5	−1	.25	2	1	.15
0	−.5	.25	1.0	.25	−.5	.10
−1	1	2	.25	−1	−1.5	.15
−2	2	1	−.5	−1.5	−2	−.20
M	−.2	.15	.10	.15	−.2	

an unequal-weights tetrahedron model (see Table 2) was descriptively superior to the equal-weights tetrahedron and Feather models. Wellens and Thistlethwaite (1971b) replicated this basic result but also found that for incomplete triads with *p-x* and *o-x* dependent variables "the results were best described by an equal-weights version of Wiest's tetrahedron model" (p. 82). Wellens and Thistlethwaite (1971b) speculated that perhaps the lesser importance, or lower weight, of the lower-boundary formula for predicting the *p-o* relation may be due to subjects' reluctance to imagine negative relations with others. (Note the abundance of negative relations for the lower-boundary formula in Table 1.)

The present investigation is a further study of the adequacy of the various balance models where the *p-o* relation, interpersonal attraction, is the dependent variable. The present investigation differs from the two groundbreaking experiments by Wellens and Thistlethwaite in three salient respects. First, the general experimental procedure did not involve hypothetical role playing but a variation of Byrne's anonymous-stranger technique. Second, instructions regarding the extreme values at the endpoints were used in an attempt to improve the measurement. Third, Kaplan's (1972) technique was used to measure and distinguish ambivalent and indifferent, neutral attitudes. Each of these matters will be considered in turn.

The first procedural change was to use a variation of Byrne's anonymous-stranger technique. Each subject received information regarding an anonymous stranger's attitude on five different issues. For four of the five issues the subject's attitude was −2, 0, or +2, and for four of the five issues the other's attitude was −2, 0, or +2. The exact patterns used are shown in Table 4. For example, in the +2,0 cell there are three instances in which the subject's attitude is +2 and the other's attitude is 0, one instance in which the subject's attitude is +1 and the other's attitude is 0, and one instance in which the subject's attitude is +2 and the other's attitude is (randomly) either +1 or −1. Such patterns were used in preference to completely uniform pairings in an attempt to create verisimilitude and to prevent a problem noticed during pilot testing. During pilot testing some of the subjects in cells with identical subject and other attitudes told us that they felt as if they were judging themselves. The slight variation eliminated the problem (and also reassured us that small scale differences are meaningful). The use of such a five-item procedure required that we initially pilot test so as to obtain a sufficient number of items (or item responses) to enable any subject to be randomly placed in any cell. The procedure also required that a computer be used to select the appropriate items regarding which feedback is to be given.

Table 4
Patterns of the Feedback Items

p-x value	*o-x* value −2	0	+2
+2	−2	0	+2
+2	−2	0	+2
+2	−1	−1/+1	+1
+2	−2	0	+2
+1	−2	0	+2
0	−2	0	+2
0	−2	0	+2
0	−1	−1/+1	+1
0	−2	0	+2
−1/+1	−2	0	+2
−2	−2	0	+2
−2	−2	0	+2
−2	−1	−1/+1	+1
−2	−2	0	+2
−1	−2	0	+2

Table 5
Mean Predicted Attraction in Each Cell of the Design Based on the Equal-Weights Tetrahedron, the Unequal-Weights Tetrahedron, and the Feather Models

p-x value	Equal weights *o-x*				Unequal weights *o-x*				Feather (1967) *o-x*			
	−2	0	+2	*M*	−2	0	+2	*M*	−2	0	+2	*M*
+2	−1.61	−.01	1.60	−.01	−1.61	.09	1.60	.04	−1.81	−.36	1.60	−.23
0	−.01	.00	.01	.00	.09	.80	.11	.45	−.35	.70	−.34	.18
−2	1.61	.02	−1.64	.00	1.61	.13	−1.64	.06	1.61	−.32	−1.82	−.21
M	.00	.00	.00	—	.05	.45	.05	—	−.23	.18	−.22	—

Note. The slight variations from symmetry in these values are due to minor variations in the number of items per subject and to random variation in the feedback patterns (see Table 4). Note also that the marginals reflect the greater number of subjects in the 0 conditions.

From a theoretical perspective the problem is that the Wiest and Feather models apply most directly to one and not five issues. We decided to calculate the Wiest and Feather predictions for each item and then average across the five items (weight each of the items by $1/n$). Mean predicted results for the equal-weights tetrahedron model, the unequal-weights tetrahedron model, and Feather's model are presented in Table 5. All three models predict an interaction, but only the equal-weights model does not predict *p-x* and *o-x* main effects.

The second procedural change involved an attempt to improve the measurement. Anderson (e.g., 1974) has repeatedly emphasized the advantage of end-anchors for rating scales. Anderson maintains that rating scales are biased near their endpoints and that the precaution of using end-anchors overcomes possible ceiling effects that may result from such bias. According to Anderson, end-anchors are "stimuli that are higher or lower in value than the regular experimental stimuli" (1974, p. 245). We did not use end-anchors in this sense of the term, but we did attempt to incorporate the spirit of the recommendation. Thus subjects were initially instructed that the −3,+3 extremes of the scale were only for extraordinarily positive or negative feelings and that such feelings were not common. For the case of the initial instructional example, "North Carolina," they were told that the endpoints should not be used unless "you feel that there definitely isn't any place in the whole world better than it or worse than it." A reminder that the endpoints were only for extreme feelings was periodically repeated on each subject's viewing screen. With this procedure the percentage of endpoint endorsement for all items was 5.86.

Although the initial instructions and periodic reminders may provide for a more uniform interpretation of the scales, the general "anchoring" procedure does create some uncertainty regarding the k term in the various models. With a 7-point scale, $k = 3$, whereas with a 5-point scale, $k = 2$. There are a number of considerations, however, that lead us to believe that the most reasonable interpretation is for k to equal 2. Twenty-five percent of the subjects did not endorse a single endpoint on any of the scales. For these subjects, then, none of the items evoked feelings as extreme as the abstract anchor, and the effective scale was a 5-point scale. On the other hand, 75% of the subjects did endorse an endpoint for at least one item. Although there was some scatter in the items producing endpoint endorsement, the tendency was more apparent for some items than for others. Six of the 50 items had endpoint endorsements above 10%. For example, "homosexuality" produced 19% endorsement at −3, and "one true religion" produced 9% endorsement at +3 and 10% endorsement at −3. Since the computer only selected items that a given subject had endorsed at −2, 0, or +2, it is arguable that the 75% of the subjects who endorsed one or more endpoints really discovered their own idiosyncratic anchors, and thus for purposes of the items utilized in the

manipulations there was a functional 5-point scale. To the extent that the 75% of the subjects who did endorse one or more endpoints and the 25% of the subjects who did not endorse any endpoints at all had functional 5-point scales, no subjects should have endorsed an endpoint of any of the three dependent variables. In fact only 1 of the 224 subjects did so.

Beyond this it is important to note that whether k is 2 or 3 has less effect on the predicted values than might be initially thought. There is no effect at all for the equal-weights model, since the k terms exactly cancel out. For the unequal-weights model, the predicted values in each cell are all altered by the same constant, so that the magnitudes of the predicted main effects and interactions are exactly the same. For the Feather model the variation in k affects only the cells involving same signs or 0, with the result that the predicted main effects do not vary in magnitude, but the predicted interaction is somewhat smaller when $k = 2$. As far as the predicted effects for the analysis of variance are concerned, then, the exact value for k makes little or no difference. With regard to the correlation between obtained and predicted effects, the value for k makes no difference at all (since k determines intercept but not slope). Finally, with regard to the mean squared deviations of obtained and predicted effects, the value for k makes a difference only for the unequal-weights and Feather models. Accordingly, these values have been computed for both $k = 2$ and $k = 3$.

The third and final salient differences from the Wellens and Thistlethwaite studies relates to the attempt to distinguish between indifferent and ambivalent neutral attitudes. Since so much of the difference between the present study and the typical study done by Byrne relates to the inclusion of neutral attitudes, we felt it incumbent upon us to be as analytic as possible regarding such attitudes —even though neither the Wiest nor the Feather version of balance theory has been elaborated so as to take note of the distinction. Kaplan (1972) has developed a very interesting technique for measuring ambivalence. After rating an issue or concept on 7-point, bipolar, semantic-differential scal subjects use a 4-point, unipolar scale (0 to 3 to rate a concept's good qualities ignoring i negative ones, and then a 4-point, unipol scale (0 to −3) to rate a concept's bad qua ities ignoring its positive ones. Reassuringl Kaplan found that the bipolar ratings (A were highly correlated with the simple sum the unipolar positive (A_p) and the unipol negative (A_n) ratings. More relevant for pre ent purposes, however, Kaplan used the ra ings to develop a mathematical index of am bivalence (AMB). According to Kaplan,

$$AMB = TA - POL$$

(ambivalence equals total affect, TA, minu polarity, POL), where

$$TA = A_p + |A_n|$$

and

$$POL = |A_p + A_n| \simeq |A|.^1$$

It is assumed that a neutral attitude with lo ambivalence has high indifference.

For the present study half of the neutra p-x attitudes were high ambivalence and hal low ambivalence (or high indifference). Sinc an initial pilot test seemed to indicate tha the distinction had little effect, we decided t incorporate an analogous distinction into th o-x relation. Perhaps, for example, ambiv alent subjects are more attracted to ambiv alent others than to indifferent others. Sinc we were somewhat skeptical that all subject could adequately cope with feedback for thre different scales for each of the five issues, w adopted a somewhat simpler procedure. Sub jects were told that the other subject whos attitudes they were seeing had not been aske to mark the unipolar scales, as they had done but simply to check whether they were "con flicted" or "not conflicted." The meaning o the term "conflicted" was carefully describe so as to involve the inclusion of both positiv and negative qualities. Accordingly, half o the subjects receiving neutral feedback ha

[1] Kaplan (p. 369) uses $|A|$ and $|A_p + A_n|$ inter changeably in discussion of polarity. In the present study the second term was used in the calculation of ambivalence ($POL = |A_p + A_n|$).

the "conflicted" alternative checked for all five issues and half had the "not conflicted" alternative checked for all five issues. The subjects receiving either −2 or +2 feedback had the "not conflicted" alternative checked. It is mathematically impossible for subjects with highly polarized attitudes to be conflicted or ambivalent.

Since we were initially uncertain regarding how to include the above ambivalence–indifference distinction in the model testing, we decided to do a two-step analysis in which the first analysis related solely to the p-x neutral cells. If the ambivalence–indifference distinction made no difference, the cells would be collapsed. If the distinction did make a difference, the two types of neutral attitudes would be separately included with the remaining cells of the design.

Method

Subjects

The subjects were 224 students (80 male and 144 female) from the introductory psychology class at the University of North Carolina who participated in partial fulfillment of a course requirement.

Independent Variables

The design can be conceived of as involving two, four-level factors (p-x by o-x). The four levels of the p-x factor are −2, 0 indifference, 0 ambivalence, and +2. The four levels of the o-x variable are −2 nonconflicted, 0 nonconflicted, 0 conflicted, and +2 nonconflicted.

The p-x factor was manipulated by having the PDP-11 search for five items that conformed to the subject's randomly assigned condition. If the subject were assigned to a 0 attitude condition, the computer calculated the ambivalence scores for all these items, ranked the scores, and then selected either the four highest or the four lowest. A similar procedure was used for the fifth, +1 or −1, item. Mean ambivalence for all five items was .19 for 0 indifference and 1.63 for 0 ambivalence.[2] Whether the fifth item was +1 or −1 was randomly determined.

Although the initial set of 50 items had been selected so as to maximize the probability of always finding five appropriate items for any randomly assigned condition, in 17 of the 224 cases we were forced to use fewer than five items. In 11 cases four items were used, and in 6 cases three items were used. Since omitting these subjects from the various multivariate and univariate tests made no difference in the results, all subsequently reported analyses include these 17 subjects.

The o-x factor was manipulated by having the computer provide feedback regarding the other's attitude according to the pattern in Table 4. Information was also provided regarding whether the subject was conflicted. Except for the 0 conflicted case, the nonconflicted alternative was always indicated.

Procedure

The experiment was conducted on a PDP-11 computer. Subjects were tested in separate, soundproof cubicles. The instructions and other information were presented to each subject on a cathode ray tube screen, and subjects responded by using different keys on a teletypewriter.

Upon arrival subjects were informed, consistent with information provided on the "sign-up" sheet, that they were going to participate in two different experiments, one concerned with attitude measurement and one concerned with impression formation on the basis of limited information. Subjects were then directed to separate cubicles and were instructed regarding the use of the teletype. Each subject was randomly assigned to a cell of the design, and the cell number was fed into the computer.

In the first experiment subjects were presented with 50 attitude items relating to a miscellaneous set of issues (supersonic planes, gun-control laws, the honor system, communes, nuclear energy, divorce, etc.). Subjects initially responded to the items on a 7-point, bipolar scale, then on a 4-point, positive, unipolar scale, and finally on a 4-point, negative, unipolar scale. The bipolar scale was labeled: extremely in favor, 1 = (+3); strongly in favor, 2 = (2); mildly in favor, 3 = (+1); undecided or neutral, 4 = (0); mildly against, 5 = (−1); strongly against, 6 = (−2); extremely against, 7 = (−3). Instructions for the bipolar scale used the example of "North Carolina":

> You should type the number of any of these options which best represents your feelings about North Carolina. For example, if you feel very positively about North Carolina you should type 2 (strongly in favor). The two endpoints 1 (+3) and 7 (−3) are indicators of very extreme feelings which are not common in the majority of cases. You may use them only if you have an extraordinarily positive or negative feeling about an issue (e.g., in the case of North Carolina, if you feel that there definitely isn't any place in the whole world better than it or worse than it!).

The above instructions, along with additional information indicating that the 0 point could indicate either conflicting feelings or indifference, were followed by a practice example. The first 4 items and

---

[2] If $POL = |A|$, mean ambivalence is .67 for indifference, and 1.88 for ambivalence.

then every 10th item were preceded by the following reminder:

> Please indicate your feelings about the following statement by typing the number of your selected option. Remember that options 1 (+3) and 7 (−3) are for very, very extreme feelings that are not common in many cases.

Consistent with Kaplan's (1972) suggestion, the positive, unipolar rating was introduced with instructions to rate the positive qualities, ignoring the negative. The example was "ice cream" (taste vs. calories).

Following the negative, unipolar rating, the second experiment was introduced. Subjects were asked to press a button if they wished to participate in the second experiment. The instructions were in the vein of the typical "judgment about others on the basis of limited information," except for the inclusion of detailed information regarding the other's indication of whether he or she was "conflicted" or "not conflicted." The other was described as some other University of North Carolina undergraduate of the same sex as the subject. At the time this person rated the items, the unipolar scales had not been used; rather a simple decision regarding the conflicted–nonconflicted dichotomy was made. Again an example was given. The feedback regarding the other's bipolar responses followed the pattern in Table 4. The information regarding the other's response to each item was shown on the screen for 20 sec, and at the same time a copy was printed on the teletype for later reference.

After the fifth item was presented, the subjects were given the several items of Byrne's Interpersonal Judgment Scale (IJS). Subjects were again reminded that the endpoints were for extreme feelings.

Finally, subjects were taken to a different room and were asked to fill out a brief questionnaire that asked for further ratings of the other, as well as information regarding their general impressions of the other and also opinions they had regarding the purpose of the experiment.

Dependent Variables

The main dependent variable is the 7-point rating of liking from the IJS. The IJS also includes an item regarding liking to work with the other person in an experiment. The final questionnaire contained one final item of related interest. This was a 7-point rating of the extent to which the other "would like you if you were to meet." As Heider stated balance theory, there was an expectation that liking and perceived reciprocal liking would be correlated. We regard the liking measure as the most direct measure of attraction. Preference as a work partner (working) and reciprocal liking are less direct measures of attraction.

Results

MANOVA

A two-factor (4 × 4) multivariate analysis of variance (MANOVA) of the three main dependent variables (liking, working, and reciprocal liking) revealed a significant interaction, $F(26, 602) = 4.41$, $p < .01$, and no main effects for p-x, $F(9, 501) = 1.62$, $p < .11$, or o-x, $F(9, 501) = 0.01$, $p < .51$. These results are in agreement with the predictions of the equal-weights tetrahedron model and in partial disagreement with the unequal-weights and Feather models (since they also predicted main effects). Since, as is explained below, the different types of neutral attitude produced no significant variation, we will initially examine these dependent variables with the different types of neutral attitude collapsed. Such a procedure facilitates testing of the various models.

Liking

A 3 × 3 analysis of variance (ANOVA) of the liking scores revealed a significant interaction, $F(4, 215) = 20.29$, $p < .01$, and nonsignificant main effects for p-x, $F(2, 215) = 1.36$, and o-x, $F(2, 215) = .59$. A trend analysis of the interaction revealed both a significant linear by linear component, $F(1, 215) = 69.08$, $p < .01$, and a significant quadratic by quadratic component, $F(1, 215) = 11.40$, $p < .01$. The liking means are in Table 6. As the interaction indicates, there is a tendency for the means to increase across the top row and to decrease across the bottom row. A trend analysis of just the top, +2, row reveals a significant linear trend, $F(1, 215) = 28.37$, $p < .01$, and a nonsignificant quadratic trend, $F(1, 215) = .83$. Across the middle, 0, row the linear trend is nonsignificant, $F(1, 215) = .42$, but the quadratic trend is significant, $F(1, 215) = 7.09$, $p < .01$. Across the bottom, −2, row the linear trend is significant, $F(1, 215) = 41.32$, $p < .01$, as is also the quadratic trend, $F(1, 215) = 4.32$, $p < .01$. Finally, in view of our particular interest in polarity it is relevant to note that similar extreme attitudes (+2,+2 or −2,−2) pro-

Table 6
Mean Liking, Preference As a Co-worker, and Reciprocal Liking

p-x value	Liking *o-x*				Working *o-x*				Reciprocation *o-x*			
	−2	0	+2	*M*	−2	0	+2	*M*	−2	0	+2	*M*
+2	−.64	.14	1.43	.27	−.50	−.14	1.14	.09	−.43	.50	1.36	.48
0	.21	.64	.04	.38	.21	.34	.18	.27	.36	.89	.39	.63
−2	1.64	−.18	−.86	.11	1.07	−.14	−.64	.04	1.43	.29	−.21	.45
M	.36	.31	.16	—	.25	.10	.21		.43	.64	.48	

Note. The marginals reflect the greater number of subjects in the 0 conditions.

duced more attraction than similar neutral attitudes, $F(1, 215) = 13.94$, $p < .01$.

All of the theoretical models predict interactions, but only the equal-weights model does not predict main effects. Since no main effects were obtained, this result provides support for the equal-weights model. On the other hand, the nature of the obtained interaction is only partially in accord with the equal-weights model. As the theoretical values in Table 5 indicate, the equal-weights model predicts a linear by linear interaction with an increasing linear trend across the +2 row, no trend across the 0 row, and a decreasing linear trend across the −2 row. These predicted linear effects are significant. There are also significant quadratic effects, however. Particularly noteworthy is the quadratic trend across the 0 row produced by the rise in the 0,0 cell. This quadratic or curvilinear trend is interesting because such an effect is predicted by both the unequal-weights and Feather models (see Table 5).

Finally, there is also a significant quadratic trend across the −2 row. This curvilinear trend appears to result from the fact that the −2,−2 mean (1.64) is not as discrepant from the 0,0 mean (−.18) as is the −2,+2 mean (−.86). Stated somewhat differently, the same-sign mean is more polarized in the positive direction than the opposite-sign mean is in the negative direction. Although the curvilinear trend is not significant across the top, +2, row, there is a similar tendency for the same-sign (+2,+2) mean (1.43) to be more polarized in the positive direction than the opposite-sign (−2,+2) mean (−.64) is in the negative direction. An overall test of the difference between the absolute values of these same-sign and opposite-sign means (for the corner cells) is significant, $F(1, 215) = 8.60$, $p < .01$.

Table 7 presents the correlations and mean squared deviations between obtained and predicted values for the various Wiest and Feather formulations. For sake of comparability with Wellens and Thistlethwaite analyses, the calculations have been done separately for upper- and lower-boundary formulas and also for three different formulations of the Feather model. The composite model uses the average of the same- and opposite-sign formulas when *p-o* or *o-x* is neutral. The same-sign predictions are based on the same-sign formula for neutral attitudes, and the opposite-sign predictions are based on the opposite-sign formula for neutral attitudes. In all other respects the three Feather formulations are identical. The two Wiest models differ according to whether the weights are equal ($a = .5$, $b = .5$) or unequal ($a = .75$, $b = .25$). In terms of mean squared deviations from pre-

Table 7
Correlations Between and Mean Squared Deviations of Obtained and Predicted Liking Values

Basis of prediction	*MSD* ($k=2$)	*MSD* ($k=3$)	*r*
Upper-boundary formula	1.36	2.79	.48*
Lower-boundary formula	2.42	5.01	.28*
Equal-weights model	1.22	1.22	.48*
Unequal-weights model	1.13	1.34	.51*
Feather same-sign	1.41	2.93	.46*
Feather opposite-sign	1.79	1.93	.50*
Feather composite model	1.24	1.39	.52*

* $p < .01$.

dictions, the equal-weights, unequal-weights, and Feather models are descriptively smallest, and the lower-boundary formula is descriptively largest. This is true for both $k = 2$ and $k = 3$. Since no appropriate tests of significance are possible, some care should be exercised in interpreting these results.

In terms of the correlation of predicted and obtained results, the coefficients for the three models are approximately the same. In this instance it is possible to do tests of significance—assuming one is not concerned with the problem of repeated testing. As it turned out, however, that concern is not relevant, since comparisons among all possible pairs of the three models (equal weights, unequal weights, and Feather composite) produced p values greater than .05 in every case.

The data in Table 7 descriptively indicate that the lower-boundary formula has both the largest mean squared deviation and the lowest correlation. A stepwise regression analysis, however, reveals that the lower-boundary formula is necessary to the Wiest formulation. With only the best predictor (upper-boundary formula) in the model, the correlation is .48. When the lower-boundary formula is added, the (multiple) correlation increases to .51 and the difference is statistically significant, $F(1, 221) = 29.00$, $p < .01$. The multiple regression equation for $k = 2$ is as follows:

$$\text{Liking} = .14 + .49\ (\text{upper}) + .20\ (\text{lower}).$$

With $k = 3$, the intercept changes from .14 to −.15.

It is interesting to note that the above equation gives greater weight to the upper boundary than to the lower. The difference between the two weights is, in fact, significant, $F(1, 221) = 7.36$, $p < .01$. Thus the assumption of "equal weights" is not justified for the present data.

In view of the fact that the generalized tetrahedron formula assumes a 0 intercept, it is of interest to calculate the weights or slopes for the upper and lower boundary when the scores have been standardized so as to guarantee a 0 intercept. These weights, .43 (upper) and .18 (lower), are not appreciably changed from the above values of .49 and .20. When the weights are rescaled so as to sum to 1, they are approximately .70 and .30—not greatly different from Wellens and Thistlethwaite's "unequal weights" of .75 and .25.

Examination of the lower-left to upper-right diagonal values in Table 1 makes it evident that the lower-boundary formula predicts a polarity shift for same-sign values. The reason for this is that the $|X + Y|$ term of the lower-boundary formula is exactly equal to a direct measure of polarity, $|X| + |Y|$, for same-sign values. This suggests the use of a stepwise regression to demonstrate further the importance of polarity. The two terms are $|X - Y|$ from the upper-boundary formula and the direct measure of polarity, $|X| + |Y|$. The best predictor, $|X - Y|$, correlates .48 with liking, and the correlation increases to .52, significantly higher, $F(1, 222) = 38.20$, $p < .01$, when polarity is added. For same-sign values the first term, $|X - Y|$, is a measure of discrepancy. This suggests that the limited success of the tetrahedron model is at least partially due to the fact that it includes simultaneous measures of both discrepancy and polarity.

Working

Table 6 reports the mean working (preference for working within an experiment) results. An analysis of variance revealed a significant interaction, $F(4, 215) = 10.53$, $p < .01$, and no significant main effect for p-x, $F(2, 215) = 1.17$, $p < .31$, or o-x, $F(2, 215) = .50$, $p < .61$. Both the linear component of the interaction, $F(1, 215) = 37.93$, $p < .01$, and the quadratic component of the interaction, $F(1, 215) = 4.12$, $p < .05$, are significant. Trend analyses of the separate rows revealed a linear effect in the +2 rows, $F(1, 215) = 18.7$, $p < .01$, and a linear effect in the −2 row, $F(1, 215) = 19.78$, $p < .01$. No other trends, linear or quadratic, are significant. The quadratic interaction appears to result partially from the somewhat greater deviation from the 0 level of the same-sign than from the opposite-sign corner means. A comparison of the absolute value of the same-sign corner cells with the absolute values of the opposite-sign corner means is marginal, $F(1, 215) = 3.85$, $p < .06$. These results are

a somewhat weak parallel of the liking results. As was the case for liking, similar extreme attitudes (+2,+2 or −2,−2) produced more attraction than similar neutral attitudes, $F(1, 215) = 12.14$, $p < .01$.

Reciprocal Liking

Table 6 reports mean reciprocal liking (perceived reciprocal liking) results. An analysis of variance revealed a significant interaction, $F(4, 215) = 17.54$, $p < .01$, and no significant main effect for *p-x*, $F(2, 215) = 1.20$, $p < .30$, or *o-x*, $F(2, 215) = 1.49$, $p < .23$. Both the linear component of the interaction, $F(1, 215) = 59.93$, $p < .01$, and the quadratic component of the interaction $F(1, 215) = 8.90$, $p < .01$, are significant. Trend analyses of the separate rows revealed significant linear effects in the +2 row, $F(1, 215) = 32.51$, $p < .01$, and −2 row, $F(1, 215) = 27.52$, $p < .01$. There is also a significant quadratic trend in the 0 row, $F(1, 215) = 10.94$, $p < .01$. No other trends are significant. A comparison of the absolute value of the same-sign corner cell with the absolute value of the unlike corner cells is significant $F(1, 215) = 23.67$, $p < .01$. Except for the lack of a significant quadratic trend in the −2 row, these results parallel those for liking. Again, similar extreme attitudes (+2,+2 or −2,−2) produced a greater effect than similar neutral attitudes, $F(1, 215) = 40.49$, $p < .01$.

Reactions to Neutral Others

In view of the fact that one of the major purposes of the present study is to examine the effect of neutral attitudes, it is of interest to note some of the written comments regarding neutral others. The postexperimental questionnaire contained a question that asked subjects to indicate whether or not they had any specific individual in mind when responding to questions about the other person. It is interesting to note that regardless of the type of response (yes or no), some subjects elaborated by expressing negative attitudes toward the stranger when he or she had neutral attitudes. Here is a sample of some of the comments: "someone who is wishy-washy or uninformed," "a blah sort of person in that he did not feel any way about any issue," "someone indifferent to the society," "(someone with) little feelings about current issues," "(one who) didn't care about anything," "some dumb female who didn't care about anything but her own private world," "(I don't) know anybody who has no opinion," "(a person who) didn't have any opinion about anything," "a person who wouldn't make up his mind on anything,"and "someone strongly individualistic enough to not really care."

Table 8
Mean Liking of Ambivalent and Indifferent Neutral Subjects

p-x value	*o-x* value				
	-2_n	0_n	0_c	$+2_n$	*M*
0_{ind}	.86	.93	.93	−.14	.64
0_{amb}	−.43	.57	.14	.21	.12
M	.21	.75	.54	.03	—

Note. Subscript *n* symbolizes nonconflicted and subscript *c* conflicted.

Even some of the subjects who were themselves neutral expressed negative feelings about the neutral other. Here are some of their comments: "a freshman (who) didn't seem to be extremely well-informed, close to myself," "a friend of mine who is always neutral about everything," "(a person who) is not well acquainted with current events, just like me," "(a person who) was not very intelligent, someone who was indecisive on a lot of topics," "a very indecisive person," "a very unopinionated (person)," and "(a person) who does not keep up with current events." In view of the fact that these comments came in response to a very indirect question (regarding whether or not they had a specific individual in mind), their negative tone is somewhat surprising. Some subjects at least appear to have generalized beyond the small amount of conveyed information so as to regard the other as "empty headed."

Type of Neutral Attitude

Table 8 presents the mean liking results for the ambivalent and indifferent levels of

the *p-x* factor and the cross-cutting levels of the *o-x* factor (some of which are conflicted and some of which are nonconflicted). As previously indicated, a 4 × 4 multivariate analysis of variance for liking, working, and reciprocal liking revealed that neither main effect (*p-x* or *o-x*) is significant. Our planned comparisons, however, relate to the interaction within the four center cells (see Table 8). Specifically, it was predicted that attraction would be greater for the indifferent–nonconflicted and ambivalent–conflicted cells than for the ambivalent–nonconflicted and indifferent–conflicted cells. A multivariate test of this interaction is nonsignificant, as are also each of the three univariate tests. In the case of liking and working, the interaction is in fact in the "wrong" direction.

In order to be somewhat more certain about the above findings, an internal analysis was done by selecting subjects who differed more extremely on ambivalence–indifference. The results were not appreciably changed.

Within-Cell Distributions

One final matter relates to the within-cell distribution of liking scores. Wellens and Thistlethwaite (1971a, 1971b) report that their within-cell distributions were not noticeably bimodal and took this as evidence against Wiest's assumption that the points inside the tetrahedron go toward the two surfaces or boundaries. Such a test, however, rests on the assumption that individuals will markedly differ in the extent to which upper- or lower-boundary forces predominate—an assumption that may or may not be true. In any event, we also did not notice any marked tendency toward bimodality in the nine cells of the design. Most of the cells appeared unimodal with moderate skew. The third moments (measures of skewness) for each of the cells are as follows: .065, −.431, −.430, −.264, .487, .268, 1.375, −2.175, .70. Assuming that individuals markedly differ in the extent to which upper- or lower-boundary forces predominate, the equal weights model would assume symmetrical distributions.

Discussion

The obtained results leave little doubt that a same-sign shift in polarity from +2,+2 or −2,−2 to 0,0 has a marked effect on interpersonal attraction. This result was obtained for all measures, direct and indirect. Beyond this, the results give some qualified support to the equal-weights tetrahedron, unequal-weights tetrahedron, and Feather model, with no unqualified support for any one of the models. The equal-weights tetrahedron model is the only one that predicts the obtained pattern of an interaction and no main effects. According to the equal-weights tetrahedron model, however, the interaction should be entirely linear (see Table 5). In fact, both the linear and quadratic interactions are significant for all three dependent variables (liking, working, and reciprocal liking). One aspect of this quadratic interaction is a tendency for a curvilinear trend to appear in the 0 row. Such a trend (which is not predicted by the equal-weights model but is predicted by the unequal-weights and Feather models) is significant for both liking and reciprocal liking and nonsignificant but in the same direction for working (see Table 6). These ANOVA results, then, are not entirely in accord with any of the three models.

The mean squared deviations of the obtained from predicted values and the correlation between the obtained and predicted values are approximately the same for all three models (see Table 7). The three versions of the Feather formulation are fairly similar, with a slight descriptive edge going to the composite model. The most apparent thing from the Table 7 measures is the relatively lower correlation and higher mean squared deviation for the lower boundary formula. This impression is further buttressed by the finding that the multiple regression weight for the upper boundary is significantly greater than for the lower boundary. This rather clearly indicates that the data are not consistent with the equal-weights model. Overall, then, we are left with a somewhat mixed picture.

One of the most apparent problems is the failure of the subjects to report as much dis-

like as any of the models predict. Particularly noteworthy is the failure of the opposite-sign (or dissimilar) 2,2 cells to be as extreme in the negative direction as the same-sign (or similar) 2,2 cells are in the positive direction (see Table 6).[3] Judging from the numerous studies reported by Byrne, our subjects' reluctance to express extreme dislike when there is high dissimilarity is a typical occurrence. What is responsible for this phenomenon? A brief digression suggests a possible answer.

Aronson and Worchel (1966) have advanced the interesting idea that the similarity–attraction effect is due, "at least in part, to an implicit assumption that people who hold attitudes similar to our own will like us" (p. 157). By extrapolation one could also infer that there is an implicit assumption that people who hold attributes dissimilar to our own will dislike us. The reciprocal liking results for the present study (see Table 6) do indeed indicate that similarity information does have implications for perceived reciprocal liking (and disliking). Insko, Thompson, Stroebe, Shaud, Pinner, and Layton (1973) obtained similar results, and, in addition, found further support for this "implied evaluation" hypothesis with an experimental manipulation of evaluative feedback. As predicted by this hypothesis, the slope of the similarity–attraction effect was less when evaluation feedback was either positive or negative than when it was absent. Similar results were obtained by Byrne and Rhamey (1965) and also by Clore and Baldridge (1970).

There is, therefore, evidence in support of this implied-evaluation hypothesis. Since balance theory, as stated by Heider (1958), clearly asserts that reciprocal, *p-o* liking–disliking is balanced, the implied-evaluation hypothesis can be regarded as a type of balance explanation of the similarity–attraction effect. It is an explanation that focuses attention on balance in the *p-o*, *o-p* cycle, as opposed to the emphasis upon balance in the *p-o-x* cycle. There is nothing in balance theory that states before the fact which cycle or cycles will be important. Rather, balance theory simply points to the importance of any implicit aspects of *p*'s phenomenology and assumes that the basic tendency to maintain or achieve balance always operates. It can be further noted that balance theory allows for the possibility of any of a number of causal sequences among the three variables of similarity, *p-o* sentiment, and *o-p* sentiment. Aronson and Worchel's hypothesis suggests that similarity affects *p-o* sentiment directly and also indirectly via the mediation of *o-p* sentiment. Other possibilities exist, however; for example, that similarity causes *p-o* sentiment, which causes *o-p* sentiment, which feeds back to *p-o* sentiment (cf. Insko et al., 1973).

For present purposes, however, it is interesting to note that the (direct or indirect) inference from similarity to reciprocal liking is stronger than is the inference from dissimilarity to reciprocal disliking. As the data in Table 6 indicate (and consistent with the less marked but still evident results for liking), the opposite-sign 2,2 cells are not as extreme in the negative direction as the same-sign 2,2 cells are in the positive direction. Why should this be? There is nothing in the *p-o*, *o-p* cycle per se that would predict such a result. It is, however, evident that for the typical person of high self-esteem the simultaneous occurrence of *p-o* liking and *o-p* liking produces agreement regarding the worth of the self, but the simultaneous occurrence of *p-o* disliking and *o-p* disliking produces disagreement regarding the worth of the self. Common sense certainly suggests that there is some tendency to avoid disagreement regarding the worth of the self. What is the reason for such avoidance? Social comparison theory (Festinger, 1954), of course, predicts a general avoidance of all disagreement—assuming that the *o*s are similar. On the other hand, balance theory more simply predicts the same effect. Assuming a positively evaluated self, and that disagreement implies negative evaluation, it is imbalanced for the self (+) to be

---

[3] It is interesting to note that with $k = 3$, the unequal-weights and Feather models actually predict that the opposite-sign, corner means should not be as negative as the same-sign, corner means are positive. The obtained data, however, show this tendency much more markedly than is predicted under this circumstance.

(+) negatively evaluated (−). In this instance the verb "to be" implies equality, or similar grouping, and is thus a positive unit relation. Whatever the theoretical basis for this avoidance of disagreement phenomenon, its assumed reality provides a possible explanation for the failure of the opposite-sign, high-dissimilarity cells to produce the predicted amount of disliking. The inference from dissimilarity to *p-o* disliking would be weakened because of the suggestion of *o-p* disliking, which (for the typical high self-esteem person implies disagreement regarding the worth of the self. Such a complex feedback system could account for the lesser polarity of the opposite-sign, corner cells. Incorporation of such speculation into a model that includes some version of the Wiest or Feather formulations represents a major theoretical challenge.

Aside from such complexities, close examination of the Feather model makes apparent that it predicts some fairly implausible results in the neighborhood of 0,0. Thus with X and Y both equal to .001, the predicted $Z = 2$; however, with opposite signs so that $X = -.001$ and $Y = .001$, the predicted $Z = .001$. The problem arises from the fact that the same- and opposite-sign formulas predict such differing results, and as both X and Y approach 0, such large differences do not appear reasonable.[4]

Finally, it is clear that the experiment did not support our initial expectations regarding the difference between ambivalent– and indifferent–neutral attitudes. We think that Kaplan's technique is an interesting one, and that the distraction between the two types of attitudes may indeed be important in some research contexts—for example, the problem of stability or instability of an attitude structure. Our data, however, do not indicate that the distinction is important in the context of interpersonal attraction. In the course of our research, one problem with Kaplan's scheme did become evident. Kaplan does not distinguish between indifferent–don't-care attitudes and indifferent–don't-know attitudes. Possibly this was less of a problem in the recent past than in these times of increasingly "unprepared" students.[5] In general, however, this issue needs to be addressed. It may be that the more that is known about an object, the less likely it is that the attitude toward the object will be indifferent.

[4] We are indebted to the action editor for pointing this out to us.

[5] We were amazed to learn that some subjects felt that "socialized medicine" had to do with the kind of training that MDs receive.

References

Anderson, N. H. Information integration theory: A brief survey. In D. H. Krantz, R. C. Atkinson, D. R. Luce, & P. Suppes (Ed.), *Contemporary developments in mathematical psychology* (Vol. 2). San Francisco: Freeman, 1974.

Aronson, E., & Worchel, P. Similarity versus liking as determinants of interpersonal attractiveness. *Psychonomic Science*, 1966, *5*, 157–158.

Byrne, D. Attitudes and attraction. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 4). New York: Academic Press, 1969.

Byrne, D. *The attraction paradigm*. New York: Academic Press, 1971.

Byrne, D., & Rhamey, R. Magnitude of positive and negative reinforcement as a determinant of attraction. *Journal of Personality and Social Psychology*, 1965, *2*, 884–889.

Clore, G. L., & Baldridge, B. The behavior of item weights in attitude–attraction research. *Journal of Experimental Social Psychology*, 1970, *6*, 177–188.

Feather, N. T. A structural balance approach to the analysis of communication effects. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 3). New York: Academic Press, 1967.

Feather, N. T. Organization and discrepancy in cognitive structures. *Psychological Review*, 1971, *78*, 355–379.

Festinger, L. A theory of social comparison processes. *Human Relations*, 1954, *7*, 117–140.

Harary, F. On the measurement of structural balance. *Behavioral Science*, 1959, *4*, 316–323.

Heider, F. Attitudes and cognitive organization. *Journal of Psychology*, 1946, *21*, 107–112.

Heider, F. *The psychology of interpersonal relations*. New York: Wiley, 1958.

Insko, C. A., Thompson, V. D., Stroebe, W., Shaud, K. F., Pinner, B. E., & Layton, B. D. Implied evaluation and the similarity–attraction effect. *Journal of Personality and Social Psychology*, 1973, *25*, 297–308.

Kaplan, K. J. On the ambivalence–indifference problem in attitude theory and measurement. *Psychological Bulletin*, 1972, *77*, 361–372.

Nelson, D. *The effect of differential magnitude of reinforcement on interpersonal attraction*. Unpublished doctoral dissertation, University of Texas, 1965.

Osgood, C. E., & Tannenbaum, P. H. The principle of congruity in the prediction of attitude change. *Psychological Review*, 1955, *62*, 42–55.

Wellens, A. R., & Thistlethwaite, D. L. An analysis of two quantitative theories of cognitive balance. *Psychological Review*, 1971, *78*, 141–150. (a)

Wellens, A. R., & Thistlethwaite, D. L. Comparison of three theories of cognitive balance. *Journal of Personality and Social Psychology*, 1971, *20*, 82–92. (b)

Wiest, W. M. A quantitative extension of Heider's theory of cognitive balance applied to interpersonal perception and self-esteem. *Psychological Monographs*, 1965, *79*(14, Whole No. 607).

Received January 9, 1979 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 12, 2278–2285

Personality Structure and the Circumplex

Douglas N. Jackson
University of Western Ontario
London, Canada

Edward Helmes
London Psychiatric Hospital
London, Canada

A computer simulation employing the responses of hypothetical respondents answering solely in terms of Jackson's threshold model for stylistic responding reproduced closely the circumplex structure reported by Wiggins as representing adjective trait endorsements in the interpersonal domain of personality. Wiggins' data are interpreted as defining two general dimensions, the first representing the differential salience of trait desirability, and the second, differential thresholds for responding desirably in the middle range, resulting in the endorsement of many or few traits as self-descriptive. These processes are considered of fundamental importance in self-description and presentation but are stylistic in nature and are to be distinguished from trait content. The circumplex structuring of trait content is considered too simple a model for describing the complexities of personality and is judged to be inconsistent with much of modern measurement theory, which highlights convergent and discriminant validity, simple structure, and the mutual independence of traits.

There is much that is commendable in Wiggins' (1979) approach to the classification of personality. It is *systematic* in that it seeks to distinguish definitionally a number of domains of enduring personality dispositions or traits; it is *scholarly* in that it seeks to incorporate a number of previous conceptualizations; it is *integrative* in that it seeks to relate, relatively exhaustively, diverse traits; and it is *empirically based* because certain data are presented that purport to support a particular interpretation of trait structure. We do not take issue with the value to personality theory of systematic taxonomic efforts. In this paper we take issue primarily with Wiggins' interpretation of his empirical data and its supposed relevance to the structuring of personality trait content in the interpersonal domain.

This work was supported by Research Grant 410-77-0320-R1 from the Social Sciences and Humanities Research Council of Canada and was completed while Douglas N. Jackson was at the Pennsylvania State University on a Social Sciences and Humanities Research Council of Canada Leave Fellowship. The authors thank Jerry S. Wiggins for uncovering an error in a previous draft.

Requests for reprints should be sent to Douglas N. Jackson, Department of Psychology, University of Western Ontario, London, Ontario, Canada N6A 5C2.

Wiggins distinguished interpersonal traits from those derived from other domains, such as those based on "temperament," "character," and "qualities of mind as manifested in thought, perception, and speech." Within the domain of interpersonal traits, he has identified eight theoretical variables, labeled (with accompanying abbreviations): Gregarious-Extraverted (NO), Ambitious-Dominant (PA), Arrogant-Calculating (BC), Cold-Quarrelsome (DE), Aloof-Introverted (FG), Lazy-Submissive (HI), Unassuming-Ingenuous (JK), and Warm-Agreeable (LM). He hypothesizes that measures of adjacent variable should be correlated, and if arrayed in a circle, correlations should decrease between variables as a function of their mutual distance along the circumference of the circle. Wiggins proceeded to identify 16 specific trait dimensions, and on the basis of rational criteria as well as empirical item response statistics and desirability ratings, he selected eight adjectives to define each trait. Similar pairs of traits were combined into broader categories yielding eight measures corresponding to the eight theoretical traits. These were intercorrelated and factored by principal components. A plotting of the two largest dimensions yielded the circular array of traits, ap-

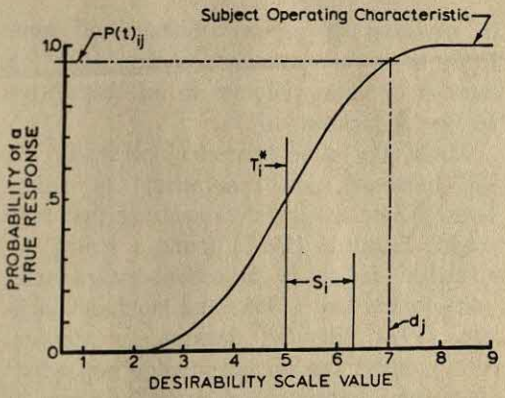
Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3712-2278$00.75

parently confirming the hypothesized structure. Wiggins seems to attach psychological importance to this structure, corresponding to what Guttman (1954) termed a circumplex.

We shall present evidence that Wiggins' findings regarding personality trait structure are subject to a very plausible alternative interpretation. We shall also suggest that his approach to the measurement of personality is not consistent with modern measurement theory and shall argue that as a model for describing behavioral consistency, a circumplex structuring of personality traits is an incomplete oversimplification. The latter is not to say that important processes may not underlie the reported circumplex structure. We shall argue that although they are not uniquely relevant to particular traits, these processes are of fundamental importance in personality assessment and interpersonal behavior.

An Interpretation of Processes Underlying the Circumplex

There is now considerable evidence that responses to personality questionnaires and judgments of the personality of others are related to the desirability level of the items being judged or endorsed. The probability of a true response to a personality statement or adjective by a group of respondents, for example, has been found to be a function of the judged desirability of the item. This finding has been observed over a wide variety of personality item content and respondent populations. Of course, this functional relationship, as Norman (1967) has noted, applies to item statistics and scale values, not to individual differences. But it is possible to compute individual probabilities of responding to items at similar levels of desirability by grouping items with similar scale values. When this is done, individuals show diverse patterns of responding (Jackson & Messick, 1969). Some respond in a manner highly related to desirability scale values; some respond as if they were unaware of desirability. For the former one might say that desirability is *salient*; for the latter it apparently is not. Furthermore, when



Derivation of the Probability of a True Response on Item j, for Subject i with the following Parameters:

T_i^* = Desirability Responding Threshold = 5
S_i = Salience Parameter = 1.25
d_j = Desirability of Item = 7
$P(t)_{ij}$ = Probability of a True Response = .94

Figure 1. Description of the threshold model.

one plots the functional relation between an individual's frequency of a true response to items having different levels of desirability scale values, one finds that some persons tend to endorse many characteristics as self-descriptive in the middle range of desirability, whereas others do so only at relatively higher levels of desirability. This implies different *thresholds* for responding desirably.

These two parameters, the individual salience of the desirability dimension (or the respondent's *sensitivity* to desirability) and the threshold for responding desirably in interaction with a stimulus parameter, the judged desirability scale value of a personality item, were the central features of the threshold formulation for stylistic responding, proposed a decade ago (cf. Voyce & Jackson, 1977; Jackson, Note 1). This model is represented in Figure 1. Using this formulation, in which salience corresponds to the *slope* of the individual function relating probability of responding true and the threshold corresponds to some median probability of a true response, processes underlying previous findings (Edwards, 1963; Jackson & Messick, 1961, 1962a) have been clarified. Measures of salience and threshold, for example, have been found to show highest loadings in the two-dimensional circumplex structure found

in measures of psychopathology and were found to be very strongly associated with the number of items endorsed as self-descriptive (Voyce & Jackson, 1977).

Might the same processes operate in the self-characterization endorsement of adjectives? There is suggestive evidence that they might. Edwards (1957) found a strong desirability factor in adjective endorsement. Jackson and Lay (1968) and Morf and Jackson (1972; Bentler, Jackson, & Messick, 1971) report two factors in endorsing adjectives, one associated with the tendency to endorse desirable adjectives (related to responding desirably to personality items) and the other associated with endorsing many adjectives (linked in a second-order factor to the tendency to endorse many personality items as self-descriptive). A similar circumplex patterning of correlations and factor loadings was reported for three factor analyses of the Minnesota Multiphasic Personality Inventory (MMPI; Jackson & Messick, 1961, 1962a, b). The question arises, what would happen if a group of respondents answered the Wiggins adjectives in terms of the threshold model, holding constant all other phenomena? Would they yield a circumplex similar to that of Wiggins? With human respondents one has no way of knowing with certainty that they would follow instructions to answer in terms of the threshold model, even if it could be communicated to them, and to ignore other influences, such as item content. But it is a relatively simple matter to program a computer to simulate the responses of a number of hypothetical subjects instructed to answer solely in terms of the threshold model (Rogers, 1971).

Computer Simulation of Adjective Endorsements

A computer program was written to simulate the responses of 500 subjects. The primary source of scale values for social desirability for the adjectives was Norman (Note 2). Three adjectives (dominant, appreciative, and enthusiastic) were not in the Norman list, and scale values were obtained from the Goldberg data (Note 3). A scale value for one adjective (organized) could not be located and was given a neutral scale value of 5.0. All scale values are based on the mean of the mean ratings of the two sexes reported by Norman (Note 2) and Goldberg (Note 3).

The simulation program predicted responses to the 128 adjectives for each of the 500 subjects on a 9-point Likert scale. The program first generated two independent, random numbers from a standard normal distribution. One number was rescaled, using a predetermined value for mean and standard deviation, to the salience parameter. Corresponding values were used to rescale the second number as a threshold. The value of threshold was then converted to the equivalent value of the intercept of the straight line illustrated in Figure 1, using a value of 5 for the value of the ordinate in the equation of the line.

To predict item responses, the values of the threshold (intercept) and the salience (slope) were used in the linear prediction equation with the social desirability scale values for the adjectives. The predicted response was rounded to the nearest integer value (1 to 9), which was then taken as the value of the response to that adjective. Responses were then summed to form scores for the eight scales.

The threshold model describes the process by which a person ascribes traits as self descriptive at different levels of desirability. This adds a slight complication in the case of negations. Consider the trait *uncrafty*. A "false" (or "uncharacteristic") response to *uncrafty* implies the presence of the trait *crafty*. Thus in the case of items containing a negation, it is necessary to consider the desirability scale value of the unnegated adjective. The Wiggins scale contains 46 such negations in four octant scales. The social desirability scale values for unnegated forms of these adjectives (obtained from Norman, Note 2) were used in the simulation. Such values for the unnegated forms of four adjectives could not be located. For these adjectives, the complement of the scale value for the negation was used. Referring to Figure 1, a false response to a negation has the same interpretation as a true response to the original. This has an important bearing on the

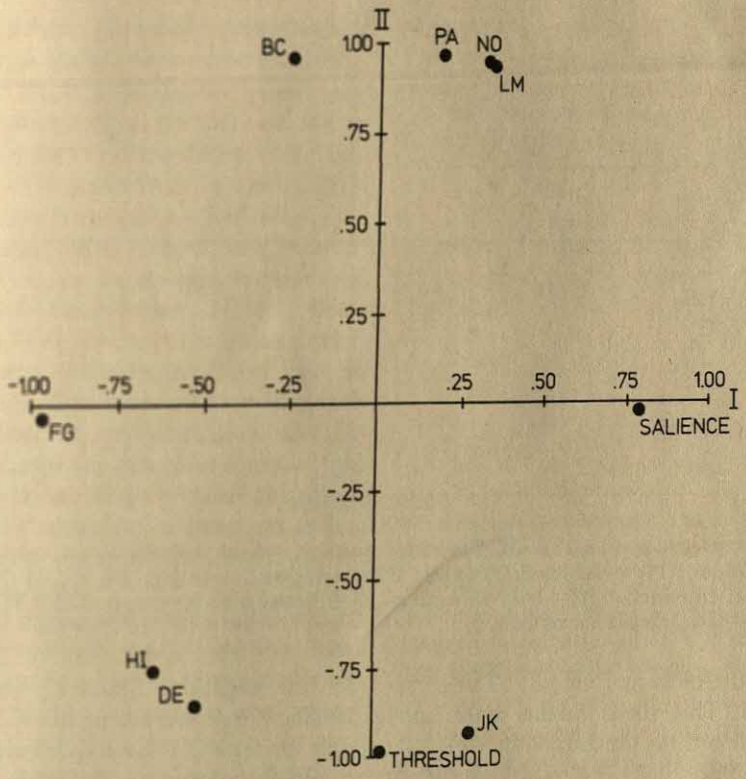


Figure 2. Plot of first two principal components from computer simulation. (PA = Ambitious-Dominant. BC = Arrogant-Calculating. DE = Cold-Quarrelsome. FG = Aloof-Introverted. HI = Lazy-Submissive. JK = Unassuming-Ingenuous. LM = Warm-Agreeable. NO = Gregarious-Extraverted.)

interpretation of the Wiggins results, as it requires the complementation of the predicted responses. It will be demonstrated that this can be interpreted as accounting for the second principal component.

Two principal components were extracted from the correlation matrix for the "responses" to the items of eight scales of the 500 hypothetical respondents. Two components accounted for 94.4% of the variance. These are plotted in Figure 2. A comparison with the Wiggins figures indicates a very similar structure. The differences consist of an overlap of NO and LM and a different ordering of the DE-FG-HI triad. It should be noted that the salience parameter loads highly on Factor 1 and the threshold parameter has a substantial loading on Factor 2. The strikingly similar pattern of results suggests that processes underlying the threshold model provide a plausible interpretation of the Wiggins data.

The precise interpretation of the two dimensions underlying the Wiggins data is facilitated by an examination of the properties of his scales. Wiggins does call attention to what he terms the "confound" between desirability and his scales, but the extent of this relationship is of some interest. First, note that in Figure 2 the four scales having the least desirable content (BC, DE, FG and HI) have substantial negative loadings, whereas the other four scales have neutral or positive loadings. In addition, a slight (approximately 30°) orthogonal rotation of the axes for the data shown in Wiggins' Figure 4 will give a complete separation of the scales high and low in social desirability. This, combined with the high loading of the salience parameter on this factor, would argue for an interpretation in terms of social desirability. Figure 3 presents a plot of the average desirability scale values and the scale means for ratings of accuracy of self-description as obtained from Norman

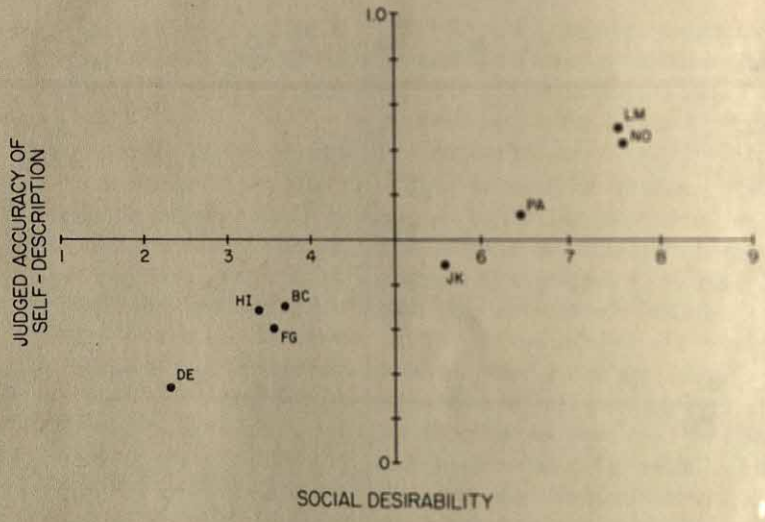


Figure 3. Correlation of social desirability scale values and judged accuracy of self-endorsement for Wiggins' scales. (PA = Ambitious-Dominant. BC = Arrogant-Calculating. DE = Cold-Quarrelsome. FG = Aloof-Introverted. HI = Lazy-Submissive. JK = Unassuming-Ingenuous. LM = Warm-Agreeable. NO = Gregarious-Extraverted.)

(Note 2) and Goldberg (Note 3). The correlation is .98. This sheds further light upon the interpretation of the first principal component. It implies that the average respondent is sensitive to the mean desirability of the adjectives and that these desirability scale values are colinear with the probability of the average respondent's endorsing items on the different scales. But some respondents are more sensitive than others to desirability (Jackson & Messick, 1969). High scorers on the first principal component are thus those who show higher relationships between their own probabilities of responding to an item and the item's judged desirability. Low scorers on the second principal component respond less in terms of the desirability scale values of the items.

The second principal component appears to reflect the differential tendency to endorse traits as self-descriptive (cf. Bentler, Jackson, & Messick, 1971; Morf & Jackson, 1972; Voyce & Jackson, 1977) and is dependent upon the presence of negations in the Wiggins traits. Table 1 presents the number of negated items on each scale together with the loading of the scale on the first principal component in the simulation study and average loadings from Wiggins (1979). When the simulation was undertaken without the special provision for the negations described above, negative loadings disappeared, along with the circumplex structure. It is hypothesized that they would also disappear in Wiggins' data, had he not chosen negations to comprise the majority of items in the scales forming the lower part of the circumplex. An examination of Figure 1 illustrates how, in terms of the threshold model, endorsing an item at a given level of desirability yields a negative correlation with the endorsement of a negated item at the same level of desirability. Those who endorse many traits as self-descriptive at a given level of desirability will tend to reject many negations at the same level of desirability. Since Wiggins engaged in item selection for the purpose of confirming a hypothesized structure, it is possible that his inclusion of many negations (even highly contrived ones like "uncrafty," "unwily," and "warmthless") for the lower two quadrants was dictated by the fact that the inclusion of such negations was necessary for obtaining the circumplex structure.

Plausibility of a Content Interpretation to the Circumplex

Although the findings in the simulation are very similar to the data reported by Wiggins, the success of the simulation does not *prove*

that the hypothesized threshold processes are operating, particularly since the simulation results do not duplicate Wiggins' results exactly. These differences may in part be due to differences between the desirability of an adjective when it is judged on its desirability and its desirability when an individual attributes the adjective to himself (see Jackson & Helmes, 1979), or to changes in adjective desirability since the Norman (Note 2) data were obtained. The forcing of a linear model onto a nonlinear situation may also have a bearing upon the differences between Wiggins' results and those of the simulation. But in weighing the competing interpretations of the data, it would be useful to consider the plausibility of the evidence as it bears on a denotative content interpretation of the two principal components.

Ever since Thurstone's pioneering work in the 1930s in multiple factor analysis, the weight of opinion has been in the direction of seeking evidence for the existence of characteristics of intellect or personality in the form of *simple structure*, that is, of identifying a resolution of test vectors such that a few tests defining a factor had high loadings, with the remainder loading negligibly. There are now literally hundreds of studies in the personality area confirming the applicability of the Thurstonian approach, that is, of expecting separate factors for each trait, defined by some analytic or judgmental criterion of simple structure. It is true that circumplexes such as those reported by Wiggins have been reported by others (Benjamin, 1974; Leary, 1957; Schaefer, 1959; Stern, 1970), but in all of these studies the potential role of response or judgmental styles was uncontrolled. Given the independence of our present simulation study, as well as that of Rogers (1971), from any particular content, it would be expected that the results of the above authors could just as easily have been simulated using parameters derived from the threshold model. On the other hand, when response styles were explicitly investigated, circumplexes consistently appeared (Jackson & Messick, 1961, 1962a, 1962b; Jackson & Pacine, 1961; Voyce & Jackson, 1977). Jackson (Note 1) found that 28 MMPI "scales" devised by ranking all 560 items in terms of judged desirability and keying alternate 20 item sets "true" and "false" yielded a regular, replicated circumplex array. True keyed scales formed a regular "fan" array from least to most desirable, all with positive loadings on the second factor, whereas false keyed scales also formed a similar regular array, but with negative loadings on the second factor. Voyce and Jackson (1977) showed that these findings did not depend upon use of MMPI items and could be obtained with frequency of endorsement scale values as well as with those for desirability. When the role of response biases is curtailed in personality assessment, however, as with the Personality Research Form, simple structure, rather than a circumplex, has been reported (Nesselroade & Baltes, 1975; Stricker, 1974).

Table 1
Number of Negated Items in Wiggins Scales, Loadings on Second Principal Component of Simulation, and Average Loadings on the Corresponding Wiggins Component

Scale	No. of negated items	Loading on first component	*M* loading from Wiggins (1979)
PA	0	.98	.81
BC	0	.96	.44
DE	11	−.85	−.14
FG	7	−.04	−.61
HI	11	−.73	−.79
JK	16	−.94	−.53
LM	0	.94	.07
NO	1	.94	.55

Note. All scales have 16 items. The data on the actual loadings were kindly made available to us by Wiggins. PA = Ambitious-Dominant. BC = Arrogant-Calculating. DE = Cold-Quarrelsome. FG = Aloof-Introverted. HI = Lazy-Submissive. JK = Unassuming-Ingenuous. LM = Warm-Agreeable. NO = Gregarious-Extraverted.

There is an important reason why traits should define independent, distinct, uncorrelated factors—such traits are more likely to yield evidence of convergent and discriminant validity. With all traits arrayed in only two dimensions and in general showing substantial correlations, it is unlikely that they would meet any of the Campbell and Fiske (1959) criteria for multitrait–multimethod validity.

In fact, before proffering or accepting the circumplex structure as a framework for understanding the content of personality, it would be appropriate to investigate and confirm the heteromethod convergent and discriminant validity of the vectors in such a space. Our own view, consistent with the findings of Jackson and Lay (1968), Kusyszyn and Jackson (1968), Morf and Jackson (1972), and others, is that the discriminant validity of traits is to be found, if at all, in the *residual* factor scores, after removing the influence of the two large principal components. In this sense, the reliabilities of individual scales reported by Wiggins can be regarded as spurious, representing not what is uniquely relevant to a given trait but what it shares with other traits. In the case of the Wiggins' data, each scale score is largely predictable from the others. A concept of *differential reliability*, as opposed to classical reliability, would distinguish the contribution to reliability of general factors, due to method and global content variance, from those due to unique trait variance.

But what is one person's method variance may be another person's trait variance (Campbell & Fiske, 1959; Jackson & Messick, 1958). The identification of two broad dimensions relevant to presentation of self may be considered an important and general finding of wide relevance to human functioning, including such diverse areas as psychopathology, socialization, and interpersonal influence. Wiggins has contributed importantly by calling attention to these dimensions, but we wish to emphasize the relative independence of these dimensions from particular trait-relevant stimuli. By an appropriate selection of stimuli and negations, Wiggins' structure could be changed dramatically. Putting one's best foot forward can be done within limits in the context of a variety of traits, and the desirability of behaviors relevant to particular traits can be manipulated. So let us not confuse content and style. (And let it not be said that we maintain that the Wiggins scales are devoid of content variance. Wiggins has confounded content and style). Both are crucial in understanding personality, but these concepts are not interchangeable. This issue might be resolved by identifying a new set of adjectives, each substantively linked to the individual scales but diverse in desirability. Half of the items should be worded positively, and the remaining, negatively. With the addition of marker scales for desirability responding, content and style factors might be distinguished, particularly if a sufficient diversity of desirability among items within each scale could be identified.

In a more general sense, in addition to methodological and empirical reservations, we are not impressed with the psychological or personological import of the circumplex. Despite what ancient philosophers have observed about perfect geometric figures, we doubt its explanatory value. Personality is almost certainly more complex than that which can be represented realistically in a two-dimensional plane. Even when considering only the interpersonal domain, such representations, although perhaps seductive in their simplicity, can hardly be expected to do justice to the subtlety and multiplexity of the variety of behavior. Similarly, although there is indeed diversity in the words used to describe traits, we would prefer to go beyond single words and use, among other things, larger units of the natural language as it has evolved over thousands of years, to define and to assess personality characteristics.

Reference Notes

1. Jackson, D. N. A threshold model for stylistic responding. In C. Hanley (Chair), *New interpretations of response style and content in personality assessment*. Symposium presented at the meeting of the American Psychological Association, San Francisco, 1968.
2. Norman, W. T. *2,800 personality trait descriptors: Normative operating characteristics for a university population*. Ann Arbor: Department of Psychology, University of Michigan, 1967.
3. Goldberg, L. R. *Test-retest and other item statistics for an adjective check list* (Research Monograph No. 5, Vol. 3). Eugene, Oregon: Oregon Research Institute, 1963.

References

Benjamin, L. S. Structural analysis of behavior. *Psychological Review*, 1974, *81*, 392–425.
Bentler, P. M., Jackson, D. N., & Messick, S. Identification of content and style: A two-dimensional

interpretation of acquiescence. *Psychological Bulletin*, 1971, *76*, 186–204.

Campbell, D. T., & Fiske, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 1959, *56*, 81–105.

Edwards, A. L. Social desirability and probability of endorsement of items in the Interpersonal Check List. *Journal of Abnormal and Social Psychology*, 1957, *55*, 394–396.

Edwards, A. L. A factor analysis of experimental social desirability and response set scales. *Journal of Applied Psychology*, 1963, *47*, 308–316.

Guttman, L. A new approach to factor analysis: The radex. In P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences*. New York: Free Press, 1954.

Jackson, D. N., & Helmes, E. Basic structure content scaling. *Applied Psychological Measurement*, 1979, *3*, 313–325.

Jackson, D. N., & Lay, C. H. Homogeneous dimensions of personality scale content. *Multivariate Behavioral Research*, 1968, *3*, 321–337.

Jackson, D. N., & Messick, S. Content and style in personality assessment. *Psychological Bulletin*, 1958, *55*, 243–252.

Jackson, D. N., & Messick, S. Acquiescence and desirability as response determinants on the MMPI. *Educational and Psychological Measurement*, 1961, *21*, 771–790.

Jackson, D. N., & Messick, S. Response style on the MMPI: Comparison of clinical and normal samples. *Journal of Abnormal and Social Psychology*, 1962, *65*, 285–299. (a)

Jackson, D. N., & Messick, S. Response styles and the assessment of psychopathology. In S. Messick & J. Ross (Eds.), *Measurement in personality and cognition*. New York: Wiley, 1962. (b)

Jackson, D. N., & Messick, S. A distinction between judgments of frequency and of desirability as determinants of response. *Educational and Psychological Measurement*, 1969, *29*, 273–293.

Jackson, D. N., & Pacine, L. Response styles and academic achievement. *Educational and Psychological Measurement*, 1961, *21*, 1015–1028.

Kusyszyn, I., & Jackson, D. N. A multimethod factor analytic appraisal of endorsement and judgment methods in personality assessment. *Educational and Psychological Measurement*, 1968, *28*, 1047–1061.

Leary, T. *Interpersonal diagnosis of personality: A functional theory and methodology for personality evaluation*. New York: Ronald Press, 1957.

Morf, M. E., & Jackson, D. N. An analysis of two response styles: True responding and item endorsement. *Educational and Psychological Measurement*, 1972, *32*, 329–353.

Nesselroade, J. R., & Baltes, P. B. Higher order factor convergence and divergence of two distinct personality systems: Cattell's HSPQ and Jackson's PRF. *Multivariate Behavioral Research*, 1975, *10*, 387–407.

Norman, W. T. On estimating psychological relationships: Social desirability and self-report. *Psychological Bulletin*, 1967, *67*, 273–293.

Rogers, T. B. The process of responding to personality items: Some issues, a theory and some research. *Multivariate Behavioral Research Monographs*, 1971, *6*(2).

Schaefer, E. S. A circumplex model for maternal behavior. *Journal of Abnormal and Social Psychology*, 1959, *59*, 226–235.

Stern, G. G. *People in context: Measuring person–environment congruence in education and industry*. New York: Wiley, 1970.

Stricker, L. J. Personality Research Form: Factor structure and response style involvement. *Journal of Consulting and Clinical Psychology*, 1974, *42*, 529–537.

Voyce, C., & Jackson, D. N. An evaluation of a threshold theory for personality assessment. *Educational and Psychological Measurement*, 1977, *37*, 383–408.

Wiggins, J. S. A psychological taxonomy of trait-descriptive terms: The interpersonal domain. *Journal of Personality and Social Psychology*, 1979, *37*, 395–412.

Received January 9, 1979 ■

Journal of Personality and Social Psychology
1979, Vol. 37, No. 12, 2286–2296

Externality as a Function of Obesity in Children: Pervasive Style or Eating-Specific Attribute?

Philip R. Costanzo and Erik Z. Woody
Duke University

The developmental sources of the link between stylistic externality and food-related externality found in the obese by Schachter and others were explored by testing whether the externality phenomena that have been found to differentiate obese and normal adults are also discriminators of obese and normal children. The results suggest that obese children as young as 7–12 years of age show an external responsiveness to salient food cues but not yet a generally external perceptual style. The implications of these findings for the development of obese externality are examined.

The most influential and intuitively appealing theory concerning the distinctive psychogenic characteristics of obese eating has been advanced by Schachter and his associates (cf. Schachter & Rodin, 1974). Stated most simply, Schachter's theory consists of two relatively independent hypotheses: (a) The eating behavior of the obese is less influenced by internal physiological cues signaling satiety or hunger (e.g., blood sugar level, gastric motility, etc.) than that of normal weight individuals, and (b) the eating behavior of the obese is more influenced by salient external cues (e.g., time of day, the taste and ready accessibility of foods, etc.) than that of normals. Although both of these hypotheses have met with fairly convincing confirmation in the research of Schachter and his colleagues (see Schachter & Rodin, 1974), it is the second hypothesis that has stood the test of independent replication (cf. Johnson, 1970; McArthur & Burstein, 1975; Nisbett & Gurwitz, 1970).

In attempting to provide a clear theoretical explanation for the externally moderated eating style of overweight individuals, Schachter (1971) has advanced the broad hypothesis that *any* stimulus (food related or not) above a certain intensity level is more likely to evoke a response from an obese than from a normal weight subject. In short, Schachter has argued that the externality evident in the eating behavior of the obese is only one manifestation of a pervasive external style. A number of studies support this broader hypothesis. Obese subjects show faster disjunctive reaction times, better short-term recall and lower tachistoscopic recognition thresholds than normal weight subjects do (Rodin, Herman, & Schachter, 1974), indicating a greater tendency for a salient stimulus to trigger a response in the obese. In addition, both the time estimations (Pliner, 1973; Rodin, 1975) and reaction times (Rodin, 1973) of the obese are more responsive to salient external–contextual cues than are similar responses of normals. Finally, there is evidence that obese individuals are significantly more field dependent on Witkin's rod–frame test (RFT) than normals are (Karp & Pardes, 1965; McArthur & Burstein, 1975), indicating a dispositional dependence upon salient contextual cues for the obese.

The correspondence between general external hyperreactivity and food related externality found in the obese led Rodin and Slochower (1976) to propose that externality

We would like to express thanks to the personnel of North Elementary School, Roxboro, North Carolina, for their help with this project and to Professor Irving Alexander for stimulating our thoughts on the questions raised in this article.

Requests for reprints should be sent to Philip R. Costanzo, Department of Psychology, Duke University, Durham, North Carolina 27706.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3712-2286$00.75

in general style leads to overeating, which in turn leads to obesity. In testing the plausibility of this causal sequence Rodin and Slochower (1976) performed a study that assessed environmentally mediated weight gain among preadolescent and adolescent normal weight girls who varied in general external responsiveness. More specifically, Rodin and Slochower used multiple assessments to classify 9–15-year-old camp-bound girls on external responsiveness. They predicted that the more external girls would evidence greater weight gain contingent upon the shift occasioned by an 8-week summer camp experience in a food-rich novel environment.

The findings were equivocal with respect to the primary hypothesis. Instead of finding a linear relationship between stylistic externality and weight gain, Rodin and Slochower found a nonmonotonic relationship between these variables. Thus, both the weight gainers and weight losers were found to be more external on the premeasured index than those girls whose weights remained stable across the 8-week experience. The obtained results were quite similar for the small contingent of overweight campers included in the sample. What one might conclude from the Rodin and Slochower study is that among a sample of normal weight girls one can predict weight change (but not simply weight gain) with a modicum of accuracy. Yet these data fall short of providing evidence for the causal role of generalized externality in producing environmentally mediated overeating.

In the current investigation we have addressed questions that are similar to those inquired about in the Rodin and Slochower study. That is, the current study has endeavored to examine the relative plausibility of the potential sources of the stylistic externality – food-related externality link suggested by Schachter's viewpoint and by the past research literature. It was felt that an important first step in examining the role of generalized externality in obesity would involve discerning whether obese children when compared to normal weight children would exhibit enhanced responsiveness to both external food-related cues and external nonfood cues when exposed to the kind of experimental arrangements which have characterized the research using adult samples. This strategy of inquiry introduces a broader developmental perspective on the issue of causal determination of the eating style of the obese than is evident in work such as Rodin and Slochower's study.

If general externality is a predisposing style resulting in overeating and obesity, then one should expect that children who are already obese will evidence greater external hyperreactivity than their nonobese counterparts will. This proposal, of course, presumes that childhood obesity is continuous with obesity in adulthood. The evidence would appear to justify such a presumption. A number of studies have indicated that those who are obese in childhood tend to remain obese through adulthood (Heald & Hollander, 1965; Knittle, 1972; Mullins, 1958; Rimm & Rim, 1976; Rose & Mayer, 1968). In adult follow-up studies of earlier childhood public health examinations (Abraham, Collins, & Nordsieck, 1971; Abraham & Nordsieck, 1970) it was found that 86% of overweight boys and 80% of overweight girls became overweight adults, against 42% of average weight boys and 18% of average weight girls.

An extremely poor prognosis for juvenile-onset obesity was given by Stunkard and Burt (1967), who empirically estimated that the more than 4–1 odds against an obese child becoming a normal weight adult rise to 28–1 if weight reduction does not occur by adolescence. Finally, it is important to note that the primary source of obese subjects used in adult inquiries into general and food specific externalities have been college-aged young adults (19–22 years of age). According to Grinker's (1973) criterion, which because of juvenile lipid cell development sets age 19 as the dividing line between obesity of juvenile onset and that of adult onset, most previous laboratory studies of obese externality are most likely to have used samples of subjects characterized by juvenile onset of overweight. Thus, although a definitive but expensive longitudinal study exploring the developmental vicissitudes of external responsiveness in obese individuals would be highly desirable, a study probing both general and

food-specific externalities in obese children would seem to warrant cross-sectional comparisons with the findings from methodologically similar adult inquiries, since the overlap in childhood and young adult obese populations appears to be quite substantial.

Accordingly, the current study posed three primary questions: (a) Is there a differential dependence on external food related cues for obese versus normal weight children? (b) Is there a differential dependence on and responsivity to external perceptual and cognitive processing cues for obese versus normal children? (c) Is there a relationship between a child's dependence upon external cues in the eating process and his dependency upon external cues in the perceptual–cognitive area?

The differential responsivity to food-related external cues by obese and normal children was examined in the current study by manipulating the accessibility of food. The specific manipulation used was derived from the work of Schachter and Friedman (1974) and McArthur and Burstein (1975). In both of these studies, obese and normal subjects were presented with either shelled *or* unshelled nuts to nibble on. In accordance with externality theory, both studies convincingly demonstrated that obese subjects eat significantly fewer nuts with shells on than with shells off, whereas the ad lib nut eating of normals is largely unaffected by the presence of a shell.

In order to examine differences in non-food-related externality between obese and normal children, two assessments were made. The first consisted of Witkin's rod–frame test of field dependence; the second was a measure of obese and normal children's differential estimations of the elapsed time of a boring versus an interesting film. The RFT measure had been used in the McArthur and Burstein study. As noted above, McArthur and Burstein found that obese subjects ate significantly fewer unshelled nuts than their normal weight counterparts did. These same subjects were assessed for field dependence on the RFT task, and it was found that obese subjects were significantly more field dependent than normal weight subjects were. More important, it was found that external food related behavior was significantly related t degree of field dependence. That is, McArthu and Burstein found a quite significant corre lation ($r = -.58$) between the number o unshelled nuts eaten and RFT-assessed fiel dependence. Furthermore, McArthur an Burstein found a sizable correlation ($r = .62$ between percentage overweight and degree o field dependence. Since these findings consti tute the strongest and clearest support for th link between perceptual–cognitive externalit and food-related externality, McArthur an Burstein's experimental arrangements wer replicated with 7–12-year-old children in th current study. Because past longitudinal stud ies of field dependence and psychological differentiation (cf. Faterson & Witkin, 1970) have indicated that the RFT is both reliable for children and is evidenced by cross-aged stability, it appeared reasonable for use with a childhood sample.

Obese and normal children's estimations of the elapsed time of a boring versus an interesting film (shown in the presence of a bowl of either shelled or unshelled nuts) were used as a second measure of generalized externality because it has been cited by Leon and Roth (1977) as the single strongest evidence for generalized externality differences between obese and normal weight individuals. Both Pliner (1973) and Rodin (1975) have demonstrated that obese estimates of time elapsed are significantly more affected by interest cues in the surrounding environment than are the time estimates of normal weight individuals. The current study partially derives from the Rodin (1975) investigation, which very clearly demonstrated that obese subjects tend to overestimate the length of time of a boring tape and underestimate the length of time of an interesting tape, whereas the difference in time estimation of the boring and interesting tapes for normals is not significant.

In summary, the current investigation has endeavored to discern whether the externality phenomena that have been found to differentiate obese and normal adults are also discriminators of obese and normal children. The externality view of obesity suggests a strong

causal relationship between general and food-specific externality. It is believed that the developmental continuity or discontinuity of the adult phenomena in a sample of children with a high probability of maintaining an obese adjustment into young adulthood will help to provide data on the plausibility of the application of this causal assumption to the life span.

Method

Design Overview

The experiment proper was run in two parts. In the first part obese and normal children were asked to watch and rate either a boring or an interesting film of 10 minutes' duration while in the presence of a large bowl of either shelled or unshelled peanuts. This arrangement yielded a 2 (Weight Group) × 2 (Film) × 2 (Shelled–Unshelled Nuts) between-subjects factorial design. The dependent measures assessed were (a) the amount of ad lib eating of peanuts, (b) the subjects' estimate of the time taken by the observed film, and (c) the subjects' rating of the interestingness of the observed film.

For the second part of the study, all children were individually escorted to a darkened room and were administered eight trials of Witkin's rod–frame measure of field dependence–independence. These data were evaluated in a 2 (Weight) × 2 (Sex) factorial design. In addition to the analyses of variance on the separate dependent variables, correlations were computed (within and across sex and weight) among percentage overweight, quantity of nuts consumed, and field dependence.

Subjects

Subjects were drawn from the second through fifth grades of North Elementary School, a Roxboro, North Carolina, school of about 350 children. A sample of 56 males and females (28 predesignated obese and 28 predesignated normal weight) was run through the several conditions of the study. The initial sample was selected through teacher nomination. Teachers were asked to indicate which children in their class appeared distinctly overweight, which appeared normal weight, and which appeared distinctly underweight. From their nominations, groups of 28 obese and 28 normal weight subjects were assessed on three measures under the conditions indicated above: (a) their eating of shelled or unshelled nuts, (b) their time estimations of the length of a boring or interesting film, and (c) their RFT-measured field dependence. Inclusion in the final sample of subjects was determined by the subjects' actual measured deviation from average weight for age and height and not the teacher rating. Because of the instability of weight in growing children (Garn, Clark, & Guire, 1975), a 25% overweight *minimum* criterion was used for classification of an overweight subject. The criterion for inclusion as a normal weight subject was that the child fall between −10% and 10% deviation from average weight for age and height. Children were weighed and measured in the school's annual health screening, which was fortuitously scheduled 1 week after the completion of our study. Of the 28 teacher-referred overweight children, only 1 failed to meet our minimum criterion for inclusion in the sample. The 27 overweight children included in the final sample ranged from 25% to 124% overweight for their age and height (the National Center for Health Statistics, 1970, studies were used for purposes of comparing our sample children to national averages). On the average the 27 overweight children in the experimental sample were 48.8% overweight, with a range of 40.8% to 52.7% overweight in the four Shelled Nut–Unshelled Nut × Boring Film–Interesting Film cells of study. The average age of the overweight children was 10.0, with a range of 9.7 to 10.5 in the four experimental cells. The average height of the overweight groups was 57.8 inches (1.47 m; cell means ranged from 57.2 inches to 58.1 inches, or 1.45 to 1.48 m). Finally, the average weight of the obese sample was 118.2 lbs. (cell means ranged from 115.0 to 124.6 lbs).

Teachers were not quite so accurate in their normal weight referrals. Of the 28 subjects designated as normal weight, 20 approximated our a priori criterion of between −10% and 10% deviation from normal weight. The 8 subjects discarded from the normal weight experimental cells included 2 children who were appreciably underweight (−17% and −18% below average weight) and 6 subjects who were of borderline overweight status (14% to 24% overweight). The final sample of normal weight subjects ranged from −12% to 8% deviation from average age/height weight. The mean deviation from average weight for the normal sample was −2.3% (with a range of −4.6% to 1.1% across the four experimental cells). The average age of the normal weight sample was 10.0 years, with a range of means of 9.5 to 10.7 in the experimental cells. The average height and weight of the normal weight sample were 56.1 in. and 73.1 lbs., respectively (with a cell mean range of 55.4 in. and 69.3 lbs. to 57.0 in. and 77.8 lbs.). Although the 7 discarded mild to moderately overweight subjects (the 6 who had been teacher designated as normal weight and 1 who had been teacher designated as obese—but whose objective weight status was not found to correspond to teacher rating) were excluded from the univariate analyses comparing obese with normal children on the primary dependent measures, their data were included in appropriate correlational analyses using percentage overweight as a criterion measure. Thus, the correlational analyses contained 54 subjects whose weights varied continuously from −12% to 124% deviation from average weight.

Procedure and Materials

One experimenter escorted each subject from his classroom to the trailer and introduced him to a second experimenter in the trailer. The second experimenter asked the subject to take a seat. This seat faced both a projection screen and a table upon which was a bowl of peanuts, shelled or unshelled, according to the assigned condition. This table was lit by three 40-watt incandescent lights. These provided enough illumination to make the bowl of nuts salient but not so much as to obscure the projected image.

In the nuts-with-shells condition, the fragments of three emptied shells were in the bowl. The second experimenter said, "We're trying to see what kids think of different kinds of movies. We'd like to show you a movie and then get your opinion of it, OK? Help yourself to some nuts. I'll be back when the movie is over." Meanwhile the experimenter had shelled a couple of nuts in the nuts-with-shells condition. In either condition he left, munching a couple of nuts. The movie was projected through the open door of the two adjoining rooms. The projection distance was about 10 feet (3.05 m), making a rather small but bright and easy-to-see image. Subjects were then shown an interesting or boring film. For the "interesting" condition, *Toot, Whistle, Plunk and Boom*, a 10-minute Disney cartoon about the origin and development of musical instruments, was shown. For the "boring" condition, 10 minutes of *The Right of Age*, a child-educational film dealing with protective services available for the aging, was shown.

At the end of the film the experimenter returned and asked three questions. For the first question the subject was presented with a clock face with movable hands and was asked, "How long did the movie seem to take? Would you like to show me on this clock face? Here's the time we started—now what time would you think it is now? I want your impression of how much time went by." Subjects with wristwatches were asked not to look at them until they had guessed. If the subject got the two hands mixed up or had some other readily apparent difficulty, the experimenter assisted him a bit. The second question was, "How much do you think other kids would like this movie?" The subject's response to to this question was recorded on a 6-point scale ranging from "dislike a whole lot" to "like a whole lot." The third question asked, "Was this movie interesting to you?" Again the response was recorded on a 6-point scale ranging from "very boring" to "very interesting."

Following these queries, the first experimenter escorted the subject to the school's health room. Outside this room he explained to the subject that inside it would be dark and all he would see was a square lit up with a stick inside it. He was told that the task was a game, the object of which was to make the stick "go straight up and down, like the walls of a building or a flagpole outside. You will have some buttons which make the stick move." Then the experimenter held a ruler at an angle an asked the subject to make it straight up and dow Invariably the subject immediately repositioned th ruler to vertical.

The subject was blindfolded, led into the com pletely darkened room, and seated in a chair abou seven feet in front of the rod-and-frame apparatu The rod-and-frame apparatus was modeled after th of Witkin and his associates (1962) and consisted an outline square 40 inches (1.02 m) on each si within which was centered a 38-inch (.97 m) ro The rod and frame could be tilted independently and their respective degrees of tilt read off a pro tractor at the back of the apparatus. A variac wa used to limit the brightness of luminescent panel on the rod and frame, so that in a dark room onl the illuminated outline square and rod could be see The degree of tilt of the frame could be controlle manually by the experimenter, whereas the degre of tilt of the rod could be controlled remotely b the subject through push buttons that operated motor. After a brief period of dark adaptation, th subject's blindfold was removed and he was aske to make eight judgments of verticality. On half o the trials the frame was tilted 28 degrees to the right and on half it was tilted 28 degrees to the left, thu generating four combinations of rod-and-frame tilt Each combination was presented twice in random order. After each trial the subject was told to clos his eyes while the experimenter used a flashlight t record the degrees of deviation of the rod setting from vertical and to set the rod and frame at th appropriate tilt for the next trial. The averag degrees of deviation from vertical placement of th rod for the eight trials constituted the measure o field dependence. Finally, the subject was thanked and was returned to his classroom.

Results

Eating Behavior

A $2 \times 2 \times 2$ (Weight × Film × Nut) analysis of variance was performed on log grams of nuts consumed by the children. There were two significant effects that emerged from this analysis. First there was a main effect for nuts, indicating that a greater quantity of nuts without shells were consumed than nuts with shells, $F(1, 39) = 9.15$, $p < .01$. There was also a significant Weight × Nut interaction, $F(1, 39) = 4.94$, $p < .05$, which qualified the nut main effect. Table 1 presents the means involved in the interaction. Both actual grams of nuts consumed and log grams of nuts are represented on this table. It is clear from even a casual perusal of the means presented in this table that the presence or absence of shells on the nuts exerted a much stronger effect on the consumption of obese children

Table 1
Nut Consumption as a Function of Weight and the Presence of Shells

Condition	Normal			Obese		
	Log grams	Grams	*n*	Log grams	Grams	*n*
Shells off	1.06	18.97	11	1.47	43.50	13
Shells on	.94	11.07	9	.67	6.29	14

than on the consumption of normal weight children. Simple effects analysis shows that for obese subjects the presence of shells very significantly depressed consumption in comparison to the shells off condition, $F(1, 39) = 16.94$, $p < .001$, whereas for normal weight subjects the presence of a shell had an insignificant effect on nut consumption, $F(1, 39) < 1$. The type of film (i.e., boring or interesting) had no influence on eating behavior, either as a main effect or in interaction with the other independent variables.[1]

In summary, then, these data provide a very convincing case for the early entry of externality effects in the eating of the obese. The obtained results of this study compare very favorably with those of the Schachter and Friedman (1974) study and the McArthur and Burstein study (1975). A more articulated comparison of these data and McArthur and Burstein's nut-eating data will be taken up in the discussion.

Field Dependence: Judgments of Verticality

The 2 × 2 (Weight × Sex) analysis of variance on subjects' RFT judgments revealed no effects whatever for either weight, sex, or the Weight × Sex interaction. In fact, all *F*s were less than 1.

Nonetheless, inspection of within-cell correlations revealed a consistent pattern of results. In spite of the lack of significant difference in field dependence between obese and normal subjects, the correlation between percentage overweight and RFT assessed field dependence for the total sample of 54 was significant, $r(52) = .33$, $p < .02$. However, the major source of this correlation was within the obese sample itself, $r(25) = .50$, $p < .01$, and it was not significant for the normal weight subjects, $r(18) = .14$, $p < .20$. Hence, only for obese children was field dependence quantitatively related to degree of overweight, and field dependence appears unrelated to the qualitative distinction between overweight and normal weight.

The Relationship Between Eating Behavior and Judgments of Verticality

The negative correlation previously reported by McArthur and Burstein between

[1] Given that the sample of children used in the study varied on age and sex dimensions, it was important to discern whether either of these variables accounted for the obtained results. First of all, the sex distribution approximated a 50–50 split within the Shells × Weight experimental cells. Each of the three cells containing an odd number of subjects (see Table 1) was comprised of one more female than male subject, whereas the remaining cell contained the same number of males and females. Similarly, the ages of subjects were reasonably homogeneously variant within cell; the range of age standard deviations in the four Weight × Shells experimental cells was .91 to 1.37. Despite this equivalence of distribution by sex and age, it was felt necessary to rule out statistically the effects of both age and sex as qualifiers of our results. Accordingly, three analyses were run. The first of these analyses was a 2 × 2 × 2 (Weight × Shells × Film) analysis of covariance using age as a covariate and log grams of nuts consumed as the dependent measure. The Weight × Shells interaction and the shells main effect remained significant, Fs(1, 38) = 6.44 and 11.40, respectively. In addition, no other main effects or interactions emerged as significant. Secondly, a similar three-way analysis of covariance was performed using age *and* sex as multiple covariates. The results of this analysis were quite congruent with the first covariance analysis, Weight × Shells, $F(1, 37) = 6.27$, $p < .02$; Shells, $F(1, 37) = 11.15$, $p < .002$, with no additional effects significant. Third, a Weight × Shells × Sex analysis of variance was performed to test directly the impact of sex on the results. It was found that sex as a main effect yielded an $F < 1$ and that all three interactions of the experimental variables with sex yielded Fs < 1. Finally, reapportioning the sample by sex did not appreciably alter the Fs for the Weight × Shells interaction, $F(1, 39) = 4.49$, $p < .04$, nor for the shells main effects, $F(1, 39) = 12.32$, $p < .001$. A similar absence of age and sex effects was found in the case of the time estimation and field dependence measures. Neither sex nor age used as covariates nor sex used as an independent variable changed the pattern of reported results. In view of the lack of significant variance accounted for by age and sex of the child, collapsing the sample on these variables appeared warranted.

Table 2
Mean Time Estimations (in Minutes) of the Boring Versus Interesting Film by Obese and Normal Children

Group	Film	
	Interesting	Boring
Obese	19.79	16.02
Normal	24.80	13.43

subjects' field dependence on the RFT and log grams of nuts with shells consumed was not found with this sample of children. For all subjects run in the shells on condition, including those excluded from subject groups in the analysis of variance, this correlation was insignificantly positive, $r(25) = .08$. The correlation failed to approach significance, or even to take a negative sign, for any group of subjects—females, $r(14) = .04$; males, $r(13) = .15$; obese subjects 20% overweight, $r(14) = .27$, and normal weight subjects, $r(9) = .05$. The correlation between field dependence on the RFT and log grams of nuts without shells consumed was also insignificantly positive for all subjects run in the shells off condition, $r(29) = .28$, as well as for each group of subjects—females, $r(14) = .05$; males, $r(15) = .38$; obese subjects 20% overweight, $r(13) = .23$; and normal weight subjects, $r(11) = .22$. In summary, no evidence whatever was found to support the prediction that eating behavior would be related to field dependence scores on the RFT.

Time Estimation

A $2 \times 2 \times 2$ (Weight × Film × Nut) analysis of variance was performed on subjects' estimations of the duration of the film. The type of film (boring or interesting) yielded a highly significant main effect, $F(1, 39) = 13.25$, $p < .001$. Subjects estimated a longer duration for the interesting film ($M = 22.29$ minutes) than for the boring film ($M = 14.73$ minutes), even though the real duration of both films was the same (10 minutes).

A marginally significant Weight × Film interaction was also found, $F(1, 39) = 3.35$, $p < .10$. Although the marginality of this effect warrants some interpretive caution, it may be seen from Table 2 that the difference in time estimations for type of film appears somewhat more pronounced for normal subjects than for obese subjects. In fact, the difference was highly significant for normal subjects, $F(1, 39) = 12.77$, $p < .001$, but insignificant for obese subjects, $F(1, 39) = 1.98$, $p > .10$. Hence there is a tentative suggestion that the time judgments of obese subjects were *less* affected by interest cues than were those of normal weight subjects.

Film Ratings

The children's ratings of the interestingness of the two films were assessed in two 2(Weight) × 2(Shells) × 2(Film) analyses of variance. The first analysis was performed upon subjects' responses to the question: "How much do you think other kids would like this movie?" Whereas the second analysis was done on the children's responses to the question: "How interesting was this movie to you?" Responses to the questions were given on a 1–6 scale of liking and interest level. For both questions, the main effect for type of film was highly significant, $Fs(1, 39) = 49.84$ and 40.79; $ps < .001$. Subjects indicated that "other kids" would like the movie used in the interesting condition "pretty much" ($M = 5.32$) and would dislike the movie used in the boring condition "a little" ($M = 3.05$). The subjects themselves found the movie used in the interesting condition to be midway between "pretty interesting" and "very interesting" ($M = 5.50$) and the movie used in the boring condition to be "a little interesting" ($M = 3.82$). These results indicate that the intended differentiation between the boring and interesting movie was successfully manipulated. Although all other effects and interactions for the "other liking" measure yielded $Fs < 1$, there was a surprising and interesting Weight × Film × Nut interaction, $F(1, 39) = 4.49$, $p < .05$, for the subjects' report of their own interest in the films. Inspection of the cell means (see Table 3) reveals that all groups rated the interesting film nearly the same (no significant differences). However, there was a significant Weight × Nut simple interaction for the boring film, $F(1, 39) = 6.47$, $p < .025$. The presence or absence of shells on the peanuts

affected the ratings of the boring film differently for obese subjects than for normal subjects. Looking at simple main effects, the difference did not reach significance for normal subjects, $F(1, 39) = 1.86$, $p > .20$, but it did for obese subjects, $F(1, 39) = 5.24$, $p < .05$. That is, the obese subjects rated the boring film as significantly more interesting when it was viewed in the presence of nuts without shells than when it was viewed in the presence of nuts with shells (shells off $M = 4.33$ or "a little interesting," whereas shells on $M = 3.21$, or "a little boring"). Thus, it appears that the presence of available and attractive food enhances obese children's interest in rather mundane non-food-related stimuli. It should be noted that no other effects from the $2 \times 2 \times 2$ analysis on self-rated interest even approached significance.

Discussion

Externality in Eating and Generalized Externality

The results of the present studies concerning the externality theory strongly support its hypotheses about food-related behavior but fail to give strong support to the hypothesis that food-related externality has its origins in a general external style for obese populations. Table 4 and Figure 1 present a detailed comparison of the McArthur–Burstein experiment with university students and the attempted replication with children. With regard to the eating behavior dimension of the experiments, it is evident that the overall magnitude and pattern of the original results are closely matched in the replication. However, in the McArthur–Burstein experiment, the interaction of weight with type of nut was found to occur because the presence of a shell reduced the likelihood of eating any nuts more for obese than for normal individuals. In the replication, the results were mainly due to the tendency for the shell to reduce the quantity of nuts consumed more for obese than for normal-weight eaters. Although this difference may merely be due to differences in the kind of nut used—almonds are hard to shell, peanuts easy—it is worth mentioning as the only real distinction between the eating behavior found in the two experiments.

Table 3
Cell Means for the Question, "Was This Movie Interesting to You?"

Film	Normal		Obese	
	Shells off	Shells on	Shells off	Shells on
Interesting	5.75	5.40	5.43	5.43
Boring	3.50	4.25	4.33	3.21

However, the respective results of the field-dependence dimension of the experiments are far less congruent. In the McArthur-Burstein experiment, there was a strong relationship between degree of overweight and field dependence for all female subjects but no such relationship for male subjects. By contrast, in the replication there was a strong relationship between overweight and field dependence for all obese subjects but no such relationship for normal weight subjects. Similarly, the strong relationship between field-dependent eating and field-dependent perception found in the original experiment was totally lacking

Table 4
A Comparison of the McArthur–Burstein Experiment With University Students and an Attempted Replication With Children

Item	McArthur–Burstein *r*	Replication *r*
Correlations between percent overweight and judgments of verticality		
All subjects	.62	.33
Obese females	.60	.45
Obese males	−.06	.57
Normal females	.41	.25
Normal males	.04	.13
Correlations between nut consumption and judgments of verticality		
Shells on	−.58	.08
Shells off	.25	.28
Correlations between percent overweight and nut consumption		
Shells on	−.45	−.42
Shells off	.09	.18

in the replication. Furthermore, the lack of any difference between obese and normal children in field dependence casts doubt on the hypothesis that externality is an obesity-disposing and long-standing generalized trait. Instead, it appears that developmentally early evidences of externality in the obese are confined to the eating circumstances.

Using the field-dependence results with children as evidence against the generality of externality as a trait may seem a bit thin. However, the results of the time estimation phase of the experiment are also clearly incompatible with the generalized trait formulation. In a very recent review (Leon & Roth, 1977) of externality theory and the evidence supporting it, it has been claimed that the studies on time estimation constitute the "greatest support for the external control theory" (p. 129) among non-food-related behaviors. Although these studies are somewhat contradictory methodologically, they do consistently report that the time estimation of the obese is more manipulable by external cues than is the time estimation of normals. However, in the present study with children, obese subjects' time estimation was somewhat *less* affected by the boring–interesting distinction than was normal subjects' time estimation. It would not seem that the failure to obtain the predicted Weight × Film Interest interaction could be accounted for by the irrelevance of the time estimation task for children. The direction of the significant main effect for film interest, $F(1, 39) = 13.25$, $p < .001$, is quite compatible with obtained results for this measure that have issued from studies of adult cognitive process. Thus the children in this study react to interest cues in time estimation quite similarly to adults. Yet, these interest-based distortions in time estimation were not found to be moderated by weight status in the manner suggested by a generalized externality hypothesis. This evidence, along with the field-dependence evidence, suggests that externality in response to food cues precedes other manifestations of externality in the obese child.

In contrast to the obese children's relatively low responsiveness to interest cues in the time estimation task, they exhibited an unanticipated responsiveness to food cues in the film ratings. For the obese children, the presence or absence of shells on the nuts appears to have influenced their affective experience of a relatively uninteresting movie. This result suggests that rather small variations in incidental food cues may have a considerable effect on how the obese child experiences an unrelated, fairly neutral stimulus. Although externality in obese children may be confined mainly to food cues, these cues may affect the obese child's experience and behavior in ways not closely related to eating.

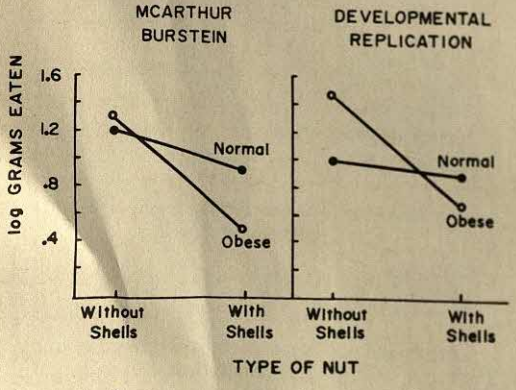


Figure 1. Log grams eaten by type of nut for McArthur–Burstein experiment and developmental replication.

Conclusions and Implications

The above-reported results lead to the summary conclusion that obese children as young as 7–12 years of age already show evidence of the externality in response to salient food cues that had been previously documented in adult obese samples. Yet, it does not appear that this external response style has clearly derived from general externality in this childhood sample.

There are at least two ways to view the discontinuities in the results of this study and prior studies of generalized externality in samples of obese young adults. One could plausibly maintain that the phenomena accompanying and accounting for childhood obesity are distinct from those that characterize obesity in adults. Given this perspective, the results of the current study would lead one to conclude that two different segments of

the obese population (i.e., children and young adults) have a common tendency to react differentially to salient versus nonsalient food-related cues. Further, the disparity between the evidence for non-food-related externality from the current study and past adult inquiries would suggest that both general externality and its relationship to overeating are distinctive characteristics of the adult obese.

Although such a descriptive distinction between two populations of obese individuals is interesting in its own right, previously cited studies documenting the population continuity between obese children and obese young adults (cf. Abraham & Nordsieck, 1970; Heald & Hollander, 1965; Rimm & Rim, 1976) would suggest that the differences found between the current child and past adult obese samples have implications for understanding the development of an obese behavioral adjustment.

From this vantage point it would appear that the causal priority of general externality in the overeating of the obese suggested by Schachter's perspective and directly proposed by Rodin and Slochower (1976) does not bear up well developmentally. The current data from a comparison sample of obese children indicate that the early instances of externality in juvenile-onset obese individuals are largely confined to the eating circumstance.

As an alternative to the proposal that implies a causal sequence that represents obesity resulting from overeating, and overeating resulting from stylistic externality, one might propose that the externality evident in the eating responses of the current sample of children is a function of their obesity and the later development of pervasive externality a consequence of generalization of eating style to other stimuli. Such a proposal makes sense when one takes account of the socialization practices that are likely consequences of a child's obesity. Owing to the deviant status of obesity in middle-class culture, childhood obesity is likely to arouse concern and resulting attempts to mediate weight loss by a variety of food restrictions. This kind of socialization strategy is likely to impose a norm of high restraint of eating responses on the child while simultaneously wresting control over the eating environment away from the child. Such directive parental constraint over a child's behavior has been shown to result in external control adjustments in a number of areas (e.g., Aronfreed, 1964; Baumrind, 1971; Becker, 1964). Aronfreed, for example, notes that strong restrictive control in a given behavioral domain both results in a poor internalization of controls and a sensitization to the stimuli in that domain of behavior. We would propose that the obese individual's food externality may be an instance of such a process of sensitization.

Although a proposal such as the one just presented is a digression from the data base of this study, it is compatible with our pattern of findings and does bear some linkage to at least one emerging framework in the study of obesity. Herman and his colleagues (Herman & Mack, 1975; Herman & Polivy, 1975; Hibscher & Herman, 1977) have proposed that the critical source of externality in the obese individual is his high level of restraint. Because the obese individual is deviant from a desired norm of physical attractiveness, he actively restrains food intake. Such restraint paradoxically renders food cues more salient and results in the stimulus-bound eating characterized by Schachter and others. Herman's results tend to confirm this viewpoint in that they demonstrate (Hibscher & Herman, 1977) that even normal weight restrainers are more likely to evidence external behavioral characteristics than are so-called unrestrained eaters. This might suggest that tracking the developmental origins of dietary self-restraint in the obese person's socialization history might be a fruitful direction for subsequent cross-sectional and longitudinal inquiries into the sources of obese externality.

References

Abraham, S., Collins, G., & Nordsieck, M. Relationship of childhood weight status to morbidity in adults. *Public Health Report*, 1971, *86*, 273.

Abraham, S., & Nordsieck, M. Relationship of excess weight in children and adults. *Public Health Report*, 1970, *85*, 263–273.

Aronfreed, J. The origins of self-criticism. *Psychological Review*, 1964, *71*, 183–218.

Baumrind, D. Current patterns of parental authority. *Developmental Psychology Monograph*, 1971, *4*(1, Pt. 2).

Becker, W. C. Consequences of different kinds of parental discipline. In M. Hoffman & L. Hoffman (Eds.), *Review of child development research*. New York: Russell Sage, 1964.

Faterson, W. F., & Witkin, H. A. Longitudinal study of the development of the body concept. *Developmental Psychology*, 1970, *2*, 429–438.

Garn, S. M., Clark, D. C., & Guire, K. E. Growth, body composition, and development of obese and lean children. In M. Winick (Ed.), *Childhood obesity*. New York: Wiley, 1975.

Grinker, J. Behavioral and metabolic consequences of weight reduction. *Journal of the American Dietetic Association*, 1973, *62*, 30–34.

Heald, E. P., & Hollander, R. J. The relationship between obesity in early adolescence and early growth. *Journal of Pediatrics*, 1965, *67*, 35.

Herman, C. P., & Mack, D. Restrained and unrestrained eating. *Journal of Personality*, 1975, *43*, 647–660.

Herman, C. P., & Polivy, J. Anxiety, restraint, and eating behavior. *Journal of Abnormal Psychology*, 1975, *84*, 666–672.

Hibscher, J. A., & Herman, C. P. Obesity, dieting, and the expression of "obese" characteristics. *Journal of Comparative and Physiological Psychology*, 1977, *91*, 374–380.

Johnson, W. G. *The effect of prior task and food visibility on the food-directed instrumental performance of the obese*. Unpublished doctoral dissertation, Catholic University of America, 1970.

Karp, S. H., & Pardes, H. Psychological differentiation (field dependence) in obese women. *Psychosomatic Medicine*, 1965, *37*, A97–A102.

Knittle, J. L. Obesity in childhood: A problem of adipose tissue cellular development. *Pediatrics*, 1972, *81*, 1048.

Leon, G. R., & Roth, L. Obesity: Psychological causes, correlations, and speculations. *Psychological Bulletin*, 1977, *84*, 117–139.

McArthur, L. Z., & Burstein, B. Field dependent eating and perception as a function of weight and sex. *Journal of Personality*, 1975, *43*, 402–420.

Mullins, A. G. The prognosis in juvenile obesity. *Archives of Disease in Childhood*, 1958, *33*, 307–314.

National Center for Health Statistics. *Height and weight of children—United States* (Series 2, No. 104). Rockville, Md.: United States Department of Health, Education, and Welfare, 1970.

National Center for Health Statistics. *Height and weight of youths 12–17 years—United States* (Series 2, No. 124). Rockville, Md.: United States Department of Health, Education, and Welfare, 1970.

Nisbett, R. E., & Gurwitz, S. B. Weight, sex, and the eating behavior of human newborns. *Journal of Comparative and Physiological Psychology*, 1970, *73*, 245–253.

Pliner, P. L. Effects of cue salience on the behavior of obese and normal subjects. *Journal of Abnormal Psychology*, 1973, *82*, 226–232.

Rimm, I. J., & Rim, A. A. Association between juvenile onset obesity and severe adult obesity in 73,532 women. *American Journal of Public Health*, 1976, *6*, 479.

Rodin, J. Effects of distraction on the performance of obese and normal subjects. *Journal of Comparative and Physiological Psychology*, 1973, *83*, 68–78.

Rodin, J. Causes and consequences of time perception differences in overweight and normal weight people. *Journal of Personality and Social Psychology*, 1975, *31*, 898–904.

Rodin, J., Herman, C. P., & Schachter, S. Obesity and various tests of external sensitivity. In S. Schachter & J. Rodin (Eds.), *Obese humans and rats*. Washington, D.C.: Erlbaum/Halsted, 1974.

Rodin, J., & Slochower, J. Externality in the nonobese: Effects of environmental responsiveness on weight. *Journal of Personality and Social Psychology*, 1976, *33*, 338–344.

Rose, H. E., & Mayer, J. Activity, caloric intake, and the energy balance of infants. *Pediatrics*, 1968, *41*, 18.

Schachter, S. Some extraordinary facts about obese humans and rats. *American Psychologist*, 1971, *26*, 129–144.

Schachter, S., & Friedman, L. N. The effects of work and cue prominence on eating behavior. In S. Schachter & J. Rodin, (Eds.), *Obese humans and rats*. Washington, D.C.: Erlbaum/Halsted, 1974.

Schachter, S., & Rodin, J. (Eds.). *Obese humans and rats*. Washington, D.C.: Erlbaum/Halsted, 1974.

Stunkard, A. J., & Burt, V. Obesity and body image: II. Age at onset of disturbances in the body image. *American Journal of Psychiatry*, 1967, *123*, 1443–1447.

Witkin, H. A., Dyk, R. B., Faterson, G. E., Goodenough, D. R., & Karp, S. A. *Psychological differentiation*. New York: Wiley, 1962.

Received March 8, 1979 ■

ournal of Personality and Social Psychology
979, Vol. 37, No. 12, 2297-2306

Avoidance of the Handicapped: An Attributional Ambiguity Analysis

Melvin L. Snyder, Robert E. Kleck, and Angelo Strenta
Dartmouth College

Steven J. Mentzer
Medical School, University of Minnesota

We demonstrated a general strategy for detecting motives that people wish to conceal. The strategy consists of having people choose between two alternatives, one of which happens to satisfy the motive. By counterbalancing which one does so, it is possible to distill the motive by examining the pattern of choices that people make. The motive used in the demonstration is the desire we believe most people have to avoid the physically handicapped. Because they do not wish to reveal this desire, we predicted that they would be more likely to act on it if they could appear to choose on some other basis. In two studies we found that people avoided the handicapped more often if the decision to do so was also a decision between two movies and avoidance of the handicapped could masquerade as a movie preference.

This paper illustrates a general strategy for detecting motives that people wish to conceal. The strategy involves asking people to choose between two alternatives, one of which accidentally happens to satisfy the motive that we suspect is present but hidden. For instance, we think that most people wish to avoid contact with the physically handicapped but do not want to admit it. If we give a person a choice between sitting next to a handicapped person or sitting beside a normal one, he may choose the handicapped so as to conceal his desire to avoid. However, if we ask a person to choose between two movies, one of which apparently by accident happens to entail sitting next to a handicapped person, the other next to a normal, he can avoid the handicapped while appearing to exercise a preference for a movie. By having enough people make such decisions and by varying which movie is associated with the handicapped person, or more generally, by varying which of two alternatives happens to satisfy the suspected motive, we can see whether people consistently make choices that satisfy it. Thus, we would look for more frequent avoidance of the handicapped when there is a choice between the two movies than when there is not.

This strategy for detecting hidden motives was derived from correspondent inference theory (Jones & Davis, 1965). Jones and Davis (1965) are concerned with how the decisions that an actor makes inform us about his motives. We learn about the actor by looking at the effects his decision produces. Jones and Davis (1965) classify the effects of making decisions into two categories: common effects (those that result regardless of which alternative is selected) and noncommon effects (those brought about by selecting one alternative but not the other). Assume for example, that Sam has a choice between two roommates. Both like classical music but one likes to cook, whereas the other likes long philosophical discussions; in this case, classical music is a common effect, and good food and bull sessions are noncommon effects. Decisions

The authors thank William Bernstein, Mark Davis, Edward E. Jones, William N. Morris, and William Swann for useful comments on a draft of this article. This research was supported in part by National Institute of Mental Health Grant MH 29446 to Robert E. Kleck.

Requests for reprints should be sent to Melvin L. Snyder, Psychology Department, Dartmouth College, Hanover, New Hampshire 03755.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3712-2297$00.75

tell us nothing about a person's intentions with respect to common effects. In this case, Sam's decision can tell us nothing about his feelings toward classical music. However, decisions have the potential to inform us about whether noncommon effects are intended. They are most informative when the number of noncommon effects is low. In Sam's choice of roommates, there are only two noncommon effects: good food and bull sessions. If he chooses the roommate who likes long discussions, we are tempted to conclude that Sam likes bull sessions. But there is still some ambiguity. Perhaps Sam is trying to lose weight and is avoiding the good cook rather than seeking the philosopher. We could be more confident about Sam's motive if there was only one noncommon effect, for example, if the only difference between the roommates was that one liked long discussions and the other did not. The general rule is that the more noncommon effects there are, the more ambiguous the motives behind a decision.

Our strategy for detecting concealed motives can be understood in terms of correspondent inference theory. Again, we will use the handicapped example. When the choice is simply between sitting with a handicapped person or with a normal one, the handicap is the most salient noncommon effect. A decision to avoid strongly suggests the actor's intention. However, when the choice is between movies that are accidentally associated with sitting next to one person or the other, the two movies as well as the handicap are now noncommon effects, and the decision that happens to avoid the handicapped leaves the motive ambiguous. The actor can choose to avoid the handicapped person and claim to be exercising a movie preference because objectively there is no way to tell from a single decision. However, by having many people make such a decision and by counterbalancing which other noncommon effect, that is, which movie, is paired with the handicapped, we can see whether people do more frequently choose to avoid the handicapped person when doing so is one of several noncommon effects of the decision rather than the only noncommon effect.

This strategy can also be understood in very similar fashion by application of Kelley (1972) discounting principle. The principl states that "the role of a given cause in pr ducing a given effect is discounted if oth plausible causes are also present" (p. 10 For example, we are less confident that t choice of a person to date is based on t person's intelligence if the person possess other features such as physical attractivene that are plausible causes for the choice. making the choice between handicapped a normal persons also a choice between movie we allow discounting of the role that t handicap plays in the decision. With t introduction of other plausible causes, t motives behind a decision are obscured. W the researchers, can discover the real moti by employing another of Kelley's principles– the covariation principle: "An effect is attri buted to the one of its possible causes wit which, over time, it covaries" (Kelley, 197 p. 3). We can see whether movie choic covaries with avoidance of the handicappe or, more generally, whether behavior covarie with satisfaction of the unacceptable motiv

To demonstrate the utility of the gener strategy that we have outlined, we decide to examine the motive we used in the exampl above. That is, we decided to test whethe people are motivated to avoid the handicappe but are unwilling to acknowledge tha motive. Being around the handicapped seem to make people feel uncomfortable. (Kleck 1966, 1968). There are a variety of reason why this might be so, for example, fear of th unknown, uncertainty about how to respon or perhaps the reminder that the handicap provides of our own mortal physicality (Becker, 1973). At the same time, few handi capped people do anything to deserve thei fate, and we may believe that we should treat the handicapped with kindness (Kleck, Ono, & Hastorf, 1966). The motive to avoid is not socially acceptable and may not be personally acceptable either.

To test for the presence of this unacceptable motive, our general strategy calls for giving people a choice between two alternatives, one of which accidentally happens to satisfy the motive. The alternatives we used were the same as in our example above— movies.

Experiment 1

Method

Subjects

Subjects were 21 males and 4 females enrolled in Dartmouth College. They were recruited through a newspaper ad and through posters offering $2.00 to participate for half an hour in an experiment on film evaluation. One subject was deleted from the different movie condition because he knew a confederate. After the deletion there were 12 subjects in each condition.

Confederates

Confederates were three college-age males. Each session required the services of two of them, one to play the role of a handicapped person, the other to play the normal. None of the confederates were handicapped. The confederate assigned to the handicapped condition for a particular session wore a metal leg brace.

The Setting

A partition divided the wall at the far end of a large room. The furniture arrangement on one side of the partition was the mirror image of the other side. On each side was a television monitor on a table. The monitor faced out from the corner made by the partition and the wall, at about a 45° orientation from each. A few feet away, there were two chairs facing the monitor. The one closer to the partition was empty. On one side of the partition sat the normal confederate. On the other side sat the confederate with the metal brace. In addition, Canadian crutches, which are made of metal, were leaning against the table that held the monitor.

Procedure

The subject came into the room at the opposite end from the monitors and confederates, and he or she sat parallel to the far wall and in line with the partition.

Cover story. The experimenter said that he was interested in people's reactions to some comedy films from the days of silent movies. They would be watching a film and then evaluating it. To make sure the subject had a chance to notice both confederates, the experimenter left the room for approximately 60 sec, saying he had to get the film description(s).

Same movie condition. In the same movie condition, the experimenter told subjects that usually he gave people a choice between two movies but that one of the videotape machines was broken and therefore he could show only one movie. He then told the name of the movie and gave the subject a brief description of it, which he said the subject should read while he went into the next room to rewind the videotape. The experimenter returned after 2 minutes. When the subject finished reading, his attention was directed to a table located at the end of the partition. It was a small triangular table with the apex pointing toward the subject in line with the partition edge. The table was equidistant from the two confederates and from the two empty chairs. The subject was told to pick up a background questionnaire from the table, take a seat in front of either monitor, and fill out the questionnaire; the movie would begin shortly. After the movie, the real subject and the normal confederate were taken from the room, and the real subject was put in another room by himself and was asked to fill out a movie evaluation form. Then the experimenter came in, probed for suspicion, and gave a debriefing.

Different movie condition. The only difference from the same movie condition is that there was no mention of a broken videotape machine. Subjects were told that they had a choice of two movies and were given a description of each. They were told which movie was to be shown on which side of the partition. Signs taped to the edge of the table below each monitor repeated this information. They were told to sit by the monitor showing the film they would like to see. The two movies were described as follows:

> *Slapstick.* This film covers the great era of visual comedy and the clowns who made it great. Included in the film are some of the top comics of the 1920s. Charlie Chase, Monty Banks, Fatty Arbuckle, Larry Semon, Andy Clyde, and others appear.
>
> *Sad Clowns.* Charlie Chaplin, Buster Keaton, and Harry Langdon, Hollywood's comedy greats, all had widely differing styles and techniques, but a common ability to mix laughter and tears.

The following were counterbalanced: which movie was associated with the handicapped person, which side of the partition the handicapped person was on, which confederate was handicapped. The experimenter was blind as to which side of the partition the handicapped person was on. The confederates' behavior prior to the subject's choice of which side to sit on was standardized. Neither confederate looked up. Each appeared preoccupied with the background questionnaire.

Results

The first two columns of Table 1 present the choice data for the first experiment. In the same movie condition, 58% of the subjects sat with the handicapped confederate. In the different movie condition, however, in which subjects have an excuse for avoidance,

Table 1
Seating Choices

Choice	Experiment 1		Experiment 2		
	Same movie	Different movie	Same movie	Different movie	Social comparison
Handicapped	7	2	11	6	5
Normal	5	10	1	6	7

only 17% sat with the handicapped. Thus, as predicted, there is greater avoidance of the handicapped in the different movie condition than in the same movie condition. This difference between conditions is significant, $\chi^2(1) = 4.44$, $p < .05$.

For purposes of assessing alternative explanations, it is useful to ask for each condition whether results differ from the 50–50 split one would expect by chance. In the same movie condition, the slight preference for the handicapped person (58%) does not differ from the null hypothesis of 50%, $\chi^2(1) = 0.33$, *ns*. In the different movie condition, however, the 83% who chose the normal person and avoided the handicapped person is significantly greater than chance, $\chi^2(1) = 5.33$, $p < .05$. This pattern permits an alternative interpretation of the results in terms of social comparison (Festinger, 1954). The argument is as follows: In the same movie condition the subjects choose seats at random; in the different movie condition subjects believe that the two confederates have exercised a choice between the two movies. Because social comparison tendencies are stronger with a similar, that is, normal, other than with a dissimilar other, subjects are more influenced by the normal's choice. Thus, rather than avoiding the handicapped person, the subject is choosing a movie he believes to be superior on the basis of its selection by the normal. To test this explanation, we ran a second experiment, repeating the two original conditions, adding a third condition, and making a few other changes.

Experiment 2

Method

The physical setting was maintained for the tw old conditions, but the new condition involved so modifications. Two partitions were used to divi the far wall into three sections, each with a tel vision monitor. The two outer sections were occupie one by the handicapped confederate, the other b the normal. Signs were tacked to the partitions i indicate which movie each was going to see; th were never going to see the same movie. The midd section was always left for the subject. There we two additional signs on the table in front of th subject, and he was told to take the sign with th title of the movie he wished to see and place it o top of the middle TV monitor and sit down. In th condition, there is never any possibility of sittin next to the handicapped person, and there is n choice of seat, only a choice of movie. Avoidanc of the handicapped, therefore, cannot be a motiv for movie choice. Although it makes no sense to loo for avoidance in this condition, it does make sens to look for social comparison processes; that is, i the social comparison explanation for the results o the first experiment is correct, we would expect subjects to prefer the movie chosen by the normal Because the new condition required a differen physical setup in the same room, to run the study efficiently and orderly without doing great violenc to assumptions of random assignment, we alternated running first four subjects in the original arrange ment (the same and different movie conditions) and then two in the new social comparison condition Within each physical setup, the various counter balanced conditions were run in truly random fashion. And of course, within the original setup assignment to the same or different conditions was also random.

Besides the new condition, there were some addi tional changes. Subjects were again Dartmouth under graduates, 25 males and 14 females, but this time they were recruited from introductory psychology and were given extra credit for participation instead of money. Three subjects were eliminated: a female in the same movie condition who was a friend of the experimenter of the first study and had dis cussed it with him; a male in the different movie condition who was suspicious of the confederates; and a female in the different movie condition who was so preoccupied with a book that she read it whenever she could—even during the movie—and consequently never noticed the handicap. The con federates were two female students enrolled in a local high school. When the subject appeared, each confederate briefly looked up and then returned to working on the questionnaire. This time the ques tionnaire contained the Nowlis (1965) Mood Adjec tive Checklist (MACL). The MACL was given again after the movie, along with an evaluation form. The rationale for giving the MACL was that mood may be related to how people react to comedy.

Results

The third and fourth columns of Table 1 present the choice results for the second experiment. As in the first experiment, a comparison of the two original conditions shows fewer people choosing to sit next to the handicapped person in the different movie condition (50%) than in the same movie condition (92%), $\chi^2(1) = 5.04$, $p < .05$. Again, there is greater avoidance of the handicapped when there is an excuse for doing so. The manipulation of number of movies, that is, noncommon effects or plausible causes, affects avoidance in the same way as in the first study and to the same extent. In the second study, for some reason, across these two conditions people more often chose the handicapped person than they did in the first study, $\chi^2(1) = 5.37$, $p < .05$. This could be the result of a variety of factors, such as different subjects, female confederates, younger confederates.

An interesting possibility is that the confederates' looking up briefly when the subject arrived resulted in a more frequent choice of the handicapped. An explanation is available in terms of self-awareness theory (Wicklund, 1975). Self-awareness is enhanced by directing attention to self using mirrors, videotapes, audiences, and so forth. It seems plausible that eye contact enhances self-awareness. One of the consequences of self-awareness is greater conformity to salient standards. In this case the standard is probably being kind to the handicapped (Kleck et al., 1966), which is accomplished by sitting next to him or her. In this way eye contact may have led to greater affiliation with the handicapped. Snyder, Grether, and Keller (1974) report data consistent with this line of reasoning, which they interpret in a similar way. They found that hitchhikers could improve their chances of getting a ride by staring at approaching drivers.

Whatever the explanation may be, the difference between experiments is useful. The second experiment shows a significant preference for the handicapped person in the same movie condition, $\chi^2(1) = 8.33$, $p < .01$. And together, the two experiments show that the movie choice variation has the same effect at different places on the scale of net tendencies to approach or avoid the handicapped.

Now let us turn to the new condition, in which avoidance is irrelevant because the subject has no choice of seats but in which social comparison effects are possible. If in the different movie condition people sit with the normal not in order to avoid the handicapped but rather to see the movie that they believe the normal selected, they should prefer the movie the normal selected in this new condition. There is only a slight preference for the normal's movie. Seven subjects rather than the six expected by chance choose the normal's movie and this result is, of course, well within the range of chance expectation, $\chi^2(1) = 0.33$, *ns*. There does not seem to be a social comparison effect. This is not too surprising if we keep in mind that the normal confederate was a high school student discriminably younger and therefore different from the subject. Further evidence against a social comparison explanation of the main results is the significant preference for the handicapped person in the same movie condition. Every subject but one did so. Social comparison theory would suggest that a more similar other, that is, the normal, would be preferred to inform the subject as to the appropriate response to the movie. Finally, although the *ns* are too small for formal analysis, social comparison theory suggests that the results should be stronger for subjects who are the same sex as the confederates (Zanna, Goethals, & Hill, 1975). If anything, the results tended slightly in the other direction: a little stronger for females in the first study when the confederates were male and a little stronger for males in the second study when the confederates were female. In short, there were no obvious consistent sex differences, and certainly none that would suggest a social comparison explanation.

Movie Evaluation

For both experiments subjects evaluated the movie they saw on an eight-item rating form. Each item was a 7-point scale with an adjective at each end. The adjective pairs were: bad/good, weak/strong, passive/active,

Table 2
Movie Evaluation

Choice	Experiment 1		Experiment 2		
	Same movie	Different movie	Same movie	Different movie	Social comparison
Handicapped					
M	40.3	30.0	29.7	38.8	33.9
n	7	2	11	6	5
Normal					
M	34.0	38.6	38.0	34.0	32.7
n	5	10	1	6	7
Combined					
M	37.2	37.7	31.2	36.4	33.3
n	12	12	12	12	12

not at all humorous/humorous, uninteresting/interesting, frivolous/thoughtful, boring/exciting, and unimportant/important. The second adjective in each pair defined the high end of the scale. A single score was derived by summing across the eight adjectives.

Table 2 presents the move evaluation results for both experiments. We thought that evaluations might be high in the different movie condition, as subjects who avoided the handicapped person may deny that they did so by emphasizing that their decision was based on the desirability of the movie. Similarly, those in this condition who sit with the handicapped person may wish to deny that the handicap made a difference, and thus they may also stress the desirability of the movie. In Experiment 1, the average movie rating is slightly and nonsignificantly higher in the different movie condition, $t(22) = .15$. In the second experiment, compared to subjects in the same movie condition, different movie condition subjects do rate the movie more highly, $t(33) = 2.08$, $p < .05$. However, this comparison fails to control for self-selection and dissonance reduction with respect to the movie choice in the different movie condition. To control for these processes, we need to use the social comparison condition as a baseline. When we do so, results are in the same direction but short of significance, $t(33) = 1.37$, $p < .18$. The overall F for the evaluation data is also short of significance, $F(2, 33) = 2.23$, $p < .12$.

Any semblance of order in the movie evaluation data breaks down when we separate subjects according to their movie or seat choice (see Table 2). In retrospect, processes of dissonance reduction, self-selection, and justification of avoidance of the handicapped may have been joined by the additional factors of arousal about being seated next to a handicapped person, the ease of making a decision, and confederate sex to create noise. We are not inclined to offer a precise interpretation of the movie evaluation data.

In both studies, the experimenter talked with the subject about the reasons for choosing a movie or a seat. This was done sufficiently systematically in the second study to report the following: In the same movie condition, of the 11 who sat with the handicapped person, 3 mentioned the handicap in their explanation, for example, 1 said "because people usually avoid them" and 2 said "because they did not want to avoid the brace." One indicated a preference for blondes. Another mentioned a preference for the right side and added that there was no particular reason. The other 5 gave only the latter reason, or rather nonreason, as did the 1 subject who chose the normal. They said such things as "it was random," "no reason," or "I just chose."

In the different movie condition, as expected, most justified their decision on the basis of a movie preference: five of six who sat with the handicapped confederate and also five of six who sat with the normal. One who sat with the handicapped said there was no reason. One who sat with the normal said he chose by a mental coin toss.

In the social comparison condition, of the five choosing the handicapped person's film, two said it was a film preference, and three said they just chose. Among the seven choosing the normal's film, six said it was a film preference, one said "I just chose."

To sum up, 3 of the 11 subjects who sat with the handicapped person in the same movie condition acknowledged that the handicap played a role, but the handicap went unmentioned by the others. Subjects are not particularly willing to admit to affiliation on

the basis of a handicap and appear unwilling to admit to avoidance on that basis.

Anxiety

The anxiety measure was the sum of responses to "clutched up," "tense," "fearful," "nervous," "on edge," and "jittery." The score for each adjective could range from 0 to 3, and thus total anxiety scores could range from 0 to 18. Table 3 presents the results for both the first administration of the Mood Adjective Checklist—Score 1, which came right after the subject chose where to sit or which movie to see, and the second administration (Score 2), which came after the movie. For this measure the three conditions do not differ on Score 1, $F(2, 33) = 1.42$, $p < .26$, nor on Score 2, $F(2, 33) = 1.59$, $p = .22$. However, within the different movie condition, subjects who sat with the handicapped person reported more anxiety right after the choice (Score 1) than did those who sat with the normal, $t(10) = 3.18$, $p < .01$. If this occurred simply because being with the handicapped raises anxiety, we would expect similarly high levels of anxiety (Score 1) for the 11 subjects in the same movie condition who sat with the handicapped. For these subjects, anxiety levels are lower, almost significantly so, $t(15) = 1.92$, $p < .08$. Another finding of interest is the high level of anxiety among subjects in the social comparison condition who chose the normal's movie compared to those in the different movie condition who did so; for Score 1, $t(11) = 3.24$, $p < .01$. For Score 2, $t(11) = 1.13$, ns.

To summarize the anxiety results, experimental conditions by themselves have no significant effect on either score. However, both those in the different movie condition who sat next to the handicapped and those in the social comparison condition who chose the movie apparently selected by the normal scored relatively high on anxiety, significantly higher than the zero level of anxiety displayed by those in the different movie condition who sat with the normal. By the time the movie was over, however, these differences were no longer significant, largely because the two high anxious subgroups become less anxious. The decline in anxiety is not significant for social comparison subjects who chose the normal's movie, $t(6) = 1.58$, $p < .17$. It is significant for different movie subjects who sat next to the handicapped person, $t(5) = 3.24$, $p < .03$.

Table 3
Experiment 2: Anxiety Scores

Choice	Same movie	Different movie	Social comparison
Handicapped			
Score 1	2.36	5.50	3.20
Score 2	1.82	.67	2.80
n	11	6	5
Normal			
Score 1	0.00	0.00	5.57
Score 2	0.00	.33	2.71
n	1	6	7
Combined			
Score 1	2.17	2.75	4.58
Score 2	1.67	.50	2.75
n	12	12	12

Discussion

Our goal was to illustrate a general strategy for detecting motives that people wish to conceal. We suspected that people desire to avoid the handicapped but do not wish to admit it. In a kind of bootstrap operation, we have demonstrated the general strategy by using it to reveal this motive. The strategy is to ask people to choose between two alternatives, one of which accidentally happens to satisfy the suspected motive. In terms of attribution theory, having the decision to avoid or to affiliate with the handicapped also be a decision between two other alternatives is to add noncommon effects of the decision (Jones & Davis, 1965) or plausible causes for the decision (Kelley, 1972). Increasing noncommon effects or plausible causes creates ambiguity about the reasons for behavior. The person may act on an unacceptable motive while appearing to select on some other basis.

Concretely, we had subjects choose between sitting with a handicapped person or a normal, or choose between two movies, one of

which entailed sitting next to a handicapped person, the other of which entailed sitting next to a normal. In the latter case, we predicted greater avoidance of the handicapped because, objectively, motivation is ambiguous and avoidance of the handicapped can masquerade as a motive preference. In both experiments we found, as predicted, greater avoidance when the choice between people was also a choice between movies.

Anxiety scores immediately after the decision were high for the six subjects in the different movie condition who sat next to the handicapped person. But it is questionable whether simple proximity to the handicapped is the explanation, as the 11 subjects who sat next to the handicapped confederate in the same movie condition had much lower scores —a difference that statistically falls short of significance. Perhaps after sitting down they realize the ease with which they could have avoided the handicapped person—by choosing the other movie—and they begin to suspect that they possess some sort of morbid curiosity.

The other group with a high anxiety score right after the decision is the seven people in the social comparison group who chose the normal's movie. Our attempt at explaining this result is rather involved, and as we find it less than compelling, we will dispense with it here. Our final comment on the anxiety scores is that whatever differences exist right after the decision have washed out by the time the movie is over.

Similarly, the movie evaluation data are sufficiently complex to preclude a simple interpretation. One thing that might have happened but did not—at least not in any strikingly obvious fashion—is subjects in the different movie condition did not justify avoiding the handicapped person by rating the movie highly. But consider subjects' explanations for the choice during debriefing of the second experiment: Even though movie ratings were not high, subjects in the different movie condition said that their decision was based on a film preference. No one said it was to avoid the handicapped. There were only three people who mentioned the handicap in explaining their decision. They sat next to the handicapped person in the same movie condition and in debriefing expressed concern about not wanting to avoid the handicapped person.

These results still leave open the following question. When the subject does avoid the handicapped, who is the subject trying to fool: the handicapped person, the experimenter, or himself? If the phenomenon of rationalized avoidance occurs in the absence of the experimenter, imagine what the experience of the handicapped person may be like in a world so complex that it is much more like our different movie condition than like our same movie condition. The data suggest that the handicapped person may be repeatedly rebuffed in social encounters by people who give what may seem to them to be reasonable excuses. The handicapped person can then distill the attitudes of others toward him in the same way that we can in the current study. This casts doubt on whether people really fool the handicapped when they give an excuse for avoiding.

The handicapped are probably not the only group of people to receive such treatment. For one, Gaertner and Dovidio (1977) present data suggesting that blacks are comparably treated. They used the bystander intervention paradigm (Darley & Latané, 1968). Subjects were led to believe that a confederate was either black or white. Then it was made to appear that the confederate was in need of help because a stack of chairs had fallen on him. In addition to varying race of victim, the experimenters also varied the number of bystanders present. Either the subject was the only potential helper, or she was led to believe that she was one of three potential helpers. There was an interaction between race of victim and number of potential helpers on helping. When the subject was alone, the black was helped slightly and nonsignificantly more than the white was. When there were three potential helpers, however, the black victim was helped only half as often as the white one, a result that statistically was a healthy trend though a bit short of significance. The interpretation the authors offered for these results is that people do not want to help the black victim but will only act on

hat desire when they have an excuse such as he assumption that others had already helped r the thought that somebody else would be a better helper. We concur with this interpretation. We would argue that the three-helper condition is much like our different movie condition. It permits the person to act on an unacceptable motive because it provides other plausible causes for doing so. Once again, ambiguity about the motive behind behavior permits the person to fulfill an objectionable desire.

In using our motive detection strategy, one problem is how to select the two choice alternatives that constitute the added noncommon effects. A balance must be struck. On the one hand, if they are too trivial, they may lack plausibility as a basis for a decision. For instance, would we have gotten avoidance of the handicapped if we had had noticeably different, but only slightly different, chairs on the two sides of the room? On the other hand, if the alternatives are too important in and of themselves, they may overwhelm the motive to avoid the handicapped. If the film will last 3 hours and one is a western and the other is a satire, the person who hates westerns and loves satires may choose on the basis of a real movie preference even if the choice entails the discomfort of sitting next to the handicapped.

If motives can only be distilled across many decisions, is this strategy of any value in assessing the hidden motives of a single individual? We believe so. The problem is to get the person to make several decisions without arousing suspicion about what one is looking for. If the decisions are between several different kinds of alternatives and if they take place at several different times in several different contexts, the problem of suspicion may be overcome. Alternatively, instead of asking for several decisions, one could ask for only a few, but each might have several alternatives, only one of which satisfies the hidden motive. If we could avoid arousing suspicion, we could ask for a single decision among twenty alternatives and randomly select the one that satisfies the hidden motive. Selection of that alternative would be significant at the .05 level.

We believe that the layperson often uses informal versions of this strategy to get information without creating offense. Two women inspect the menu posted outside a French restaurant, trying to decide whether to enter. The wealthier one, suspecting that the other thinks it is too expensive but is too shy to say so, asks whether the other is more in the mood for Italian cuisine than for French. When the other replies that yes, she is more in the mood for Italian food, the first will conclude that the French restaurant was too expensive for her friend, but her friend is saved the embarrassment of admitting it. And there are times when the person with the hidden motive will make the excuse herself. For example, without prompting she may state that she is not in the mood for French cuisine. In short, the creation of causal ambiguity can smooth social interaction and permit participants to save face by providing socially acceptable rationalizations for behavior.

Several studies suggest that people who fear failure on a task will take steps to provide themselves with an excuse for failure by not trying very hard (Frankel & Snyder, 1978; Snyder, Smoller, Strenta, & Frankel, Note 1) or by ingesting a drug that they believe will inhibit performance (Berglas & Jones, 1978). Low effort and performance-inhibiting drugs are plausible causes for failure, and their presence creates ambiguity about whether failure should be attributed to them or to low ability.

We are pursuing the general theme that people create attributional ambiguity as in the studies just mentioned and take advantage of it as in the different movie condition of the current research. Here are some other possibilities under consideration. Poor performance in the laboratory by the depressed (e.g., Klein, Fencil-Morse, & Seligman, 1976) and by those who believe in external control of their fate (Hiroto, 1974) may be the result of a low effort strategy designed to create ambiguity about the reasons for failure. We have already shown this to be so for poor performance in the so-called learned helplessness paradigm. In this paradigm, subjects are first given unsolvable problems and then a

second task that can be done. Subjects perform worse on the second task than those initially given solvable problems or no problems. Working in this paradigm, we gave subjects ready-made excuses for failure on the second task by describing the task as difficult (Frankel & Snyder, 1978) or by playing music alleged to be distracting (Snyder et al., Note 1). Their performance improved, presumably because the excuse we provided allowed them to try without forcing them to blame failure on lack of ability.

We can also ask whether procrastination is sometimes a means of providing an excuse of lack of time for possible failure. Does the involvement in multiple and diverse achievement activities that characterizes the Type A or coronary-prone behavior pattern allow such a person to lay failure in a single activity to having spread himself too thin? Is social comparison with similar others avoided or obscured when one fears a negative conclusion? Is affiliation under stress sometimes a means of creating ambiguity about responsibility by diffusing it among members of a group? We think the time is ripe to explore how people exploit rational principles of attribution (e.g., correspondent inference theory, the discounting principle) to serve such needs as self-esteem and self-presentation through the creation and judicious use of attributional ambiguity.

Reference Note

1. Snyder, M. L., Smoller, B., Strenta, A., & Frankel, A. *A comparison of egotism, negativity and learned helplessness as explanations for poor performance after unsolvable problems.* Unpublished manuscript, 1979. (Available from Melvin L. Snyder, Psychology Department, Dartmouth College, Hanover, New Hampshire 03755.)

References

Becker, E. *The denial of death.* New York: Free Press, 1973.

Berglas, S., & Jones, E. E. Drug choice as an internalization strategy in response to noncontingent success. *Journal of Personality and Social Psychology*, 1978, *36*, 405–417.

Darley, J., & Latané, B. Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, 1968, *8*, 377–383.

Festinger, L. A theory of social comparison processes. *Human Relations*, 1954, *7*, 117–140.

Frankel, A., & Snyder, M. L. Poor performance following unsolvable problems: Learned helplessness or egotism? *Journal of Personality and Social Psychology*, 1978, *36*, 1415–1423.

Gaertner, S. L., & Dovidio, J. F. The subtlety of white racism, arousal, and helping behavior. *Journal of Personality and Social Psychology*, 1977, *35*, 691–707.

Hiroto, D. S. Locus of control and learned helplessness. *Journal of Experimental Psychology*, 1974, *102*, 187–193.

Jones, E. E., & Davis, K. E. From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2). New York: Academic Press, 1965.

Kelley, H. H. Attribution in social interaction. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior.* Morristown, N.J.: General Learning Press, 1972.

Kleck, R. E. Emotional arousal in interactions with stigmatized persons. *Psychological Reports*, 1966, *19*, 12–26.

Kleck, R. E. Physical stigma and nonverbal cues emitted in face-to-face interaction. *Human Relations*, 1968, *21*, 19–28.

Kleck, R. E., Ono, H., & Hastorf, A. H. The effect of physical deviance upon face-to-face interaction. *Human Relations*, 1966, *19*, 425–436.

Klein, D. C., Fencil-Morse, E., & Seligman, M. E. P. Learned helplessness, depression, and the attribution of failure. *Journal of Personality and Social Psychology*, 1976, *33*, 508–516.

Nowlis, V. Research with the Mood Adjective Checklist. In S. S. Tomkins & C. E. Izard (Eds.), *Affect, cognition, and personality.* New York: Springer, 1965.

Snyder, M., Grether, J., & Keller, K. Staring and compliance: A field experiment on hitch-hiking. *Journal of Applied Social Psychology*, 1974, *4*, 165–170.

Wicklund, R. A. Objective self-awareness. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 8). New York: Academic Press, 1975.

Zanna, M. P., Goethals, G. R., & Hill, J. F. Evaluating a sex-related ability: Social comparison with similar others and with standard setters. *Journal of Experimental Social Psychology*, 1975, *11*, 86–93.

Received March 21, 1979 ■

rnal of Personality and Social Psychology
79, Vol. 37, No. 12, 2307-2317

Does Aggression Cause a Preference for Viewing Media Violence?

Allan Fenigstein
Kenyon College

Virtually all of the research concerned with media violence has attempted to determine whether the viewing of violence is associated with or causes aggression. Little experimental research has been directed at understanding why persons view media violence. The present research experimentally tested the hypotheses that physical aggression and fantasy aggression would lead to a preference for viewing violence. In Experiment 1, undergraduate men and women were induced to express aggressive, nonaggressive, or no fantasies and were then given an opportunity to select film clips for viewing. The films chosen by men contained more violence than those chosen by women. In addition, aggressive fantasies in males, compared to nonaggressive fantasies, increased the preference for viewing violence. Experiment 2, using only males, replicated the results of the first study and also found that men who were given an opportunity to aggress physically, compared to those who had no such opportunity, were more likely to choose to view films containing violent content. These results provide an additional perspective on the relationship between the observation of violence and the expression of aggression by suggesting that the causal effects are bidirectional: Just as the viewing of violence may increase aggression, so, too, aggressive behavior may increase the preference for viewing violence.

Psychologists have shown a strong interest n the relationship between the mass media nd behavior, especially with respect to ob-erved and expressed aggression. Extensive orrelational research using a number of ifferent indices of violent programming and ggressive behavior has consistently shown hat exposure to violence and antisocial be-avior are positively related (e.g., Dominick Greenberg, 1971; Lefkowitz, Eron, Walder, Huesmann, 1971; McIntyre & Teevan, 971; and McLeod, Atkin, & Chaffee, 1971). One causal explanation of this relationship hat has received considerable experimental pport has been that the observation of iolence leads to aggressive behavior. Since the early 1960s, an impressive number of laboratory and field experiments have investigated the consequences, for aggressive behavior by the observer, of observing violence (e.g., Bandura, Ross, & Ross, 1961; Berkowitz, 1965; Berkowitz & Geen, 1966; Leifer & Roberts, 1971; Liebert & Baron, 1972; Murray, 1973; Stein & Friedrich, 1971; Steuer, Applefield, & Smith, 1971). Although these experiments were conducted under several diverse theoretical frameworks such as social learning theory and drive theory, and in spite of some disconfirming evidence (e.g., Feshbach & Singer, 1971; Kaplan & Singer, 1976; and Milgram & Shotland, 1973), there is a remarkable degree of convergence in the experimental literature suggesting that the observation of violence contributes to aggression (Geen, 1976).

In view of the predominance of violence in the mass media (Baker & Ball, 1969; Clark & Blankenburg, 1971; Gerbner, 1971) and its potential for harm, it would seem important to know why persons seek out violent

Portions of this article were presented at the nnual Convention of the American Psychological ssociation, Toronto, August 1978. This research as conducted while the author was on leave at the niversity of Virginia.

Requests for reprints should be sent to Allan nigstein, Department of Psychology, Kenyon ollege, Gambier, Ohio 43022.

Copyright 1979 by the American Psychological Association, Inc. 0022-3514/79/3712-2307$00.75

material for viewing. Although research on media violence and aggression has been prolific, virtually no experimental work has been directed toward understanding the determinants of the preference for viewing violence (Kaplan & Singer, 1976; Weiss, 1969). The present studies attempt to identify some factors that may affect one's exposure to media violence. Specifically, it is proposed that aggression on the part of the viewer leads to a heightened interest in viewing violence. A number of previous studies have found individual differences in the preference for viewing violence that are consistent with this proposal (Diener & DeFour, 1978; Greenberg & Gordon, 1971). Diener and DeFour, for example, found a positive relationship for men between dispositional aggression and liking for program violence. One causal explanation for these findings is that aggressive persons seek out violent programs; presumably, aggressive thoughts or actions set in motion certain psychological processes, such as the need to understand one's behavior, which then cause persons to seek out portrayals of violence in the media.

The hypothesis that aggression is a causal antecedent of the viewing of violence may also be seen as an alternative interpretation of the well-established correlation between the observation of violence and the occurrence of aggression. As noted earlier, a causal relationship positing that the viewing of violence facilitates aggressive behavior has received considerable experimental support. This evidence, however, does not provide a complete explanation of the aggression–television relationship; a substantial portion of the variance in this relationship remains unaccounted for, and additional explanatory mechanisms are warranted (Eron, Lefkowitz, Huesmann, & Walder, 1972). Furthermore, the existence of a unidirectional causal relationship does not rule out the plausibility of a bidirectional relationship. It is quite possible that aggressive behavior would result in congruent viewing preferences, while at the same time media exposure would be shaping subsequent behaviors. Although both explanations are feasible, only one—that viewing violence increases aggression—has been systematically subjected to experimental verification. The present research contends that a better and more complete understanding of the relationship between aggression and the visual mass media can be achieved by examining the hypothesis that aggression may also increase selective exposure to filmed violence.

There are two *correlational* studies that have used statistical techniques to determine if a preference for violent TV programs is caused by aggression. Chaffee and McLeod (1973), using a systematic process analysis, found no relationship between reports of aggressive behavior and preference for televised violence. However, these investigators found only a weak correlation between their measure of preference and actual viewing of violent programs. Thus, the validity of the preference measure itself is somewhat open to question. Lefkowitz et al. (1971), using a cross-lagged panel technique, found no relationship between aggressive behavior at age 9 and the viewing of violence on TV 10 years later, again suggesting that the preference for viewing violence is not caused by prior aggression. However, this study failed to rule out any immediate effects of aggression on the preference for viewing violence, and the reliability of the measures used is also open to question (see Eysenck & Nias, 1978). In addition to the specific drawbacks of these two studies, correlational research in general cannot control all relevant variables, and the possibility of interference by third-order variables rules out any definitive conclusions regarding causal relationships. Even if the data generated by these studies were completely reliable and valid, the fact remains that *only two* studies (both correlational) exist that address the question of the effects of aggression on the preference for viewing violence, and neither of these studies has examined the problem through direct, controlled experimentation.

The first study reported here focused specifically on fantasy aggression as a causal antecedent of the preference for mass media violence. The fact that much of the violence on television is a visually presented fantasy suggests that *aggressive fantasies* may be particularly relevant to an understanding of TV

viewing preferences, and previous research has shown that the two variables are related (Feshbach, 1976). One explanation for this relationship may be found in Feshbach's (1961) work showing that watching violence on TV produced an increase in fantasy aggression. However, the present research suggests a *different causal sequence*: the presence of aggressive fantasies may lead to a preference for viewing violence. Thus, it was predicted that subjects who were induced to express an aggressive fantasy would have a greater interest in viewing violence than those expressing nonaggressive fantasies.

Experiment 1

Method

Subjects. The subjects were 45 female and 42 male undergraduates who participated as part of a course requirement in introductory psychology.

Procedure. In the context of a two-part experiment described as investigating "the effects of verbal and visual stimuli on thought processes," subjects were induced to express either an aggressive or a nonaggressive fantasy. In addition there was a group that was not given any opportunity to fantasize. All subjects were subsequently asked to select film clips for viewing. Subjects in the fantasy conditions were first presented with one of two sets of 24 words and were asked to construct a story using these words. For one group of fantasy subjects, 10 of the words on the list had aggressive connotations (e.g., hurt, insult, knife, anger); for the other group, matching nonaggressive words were used (e.g., help, praise, pen, joy). The remaining 14 words were constant for both groups (e.g., park, couple, dog, light). Instructions were to describe an event, its antecedents, and its outcome, and to describe the thoughts and feelings of the characters in the story. Subjects were given 10 minutes to construct and take notes on their stories, and then the story was tape-recorded for later analysis. The no fantasy group participated in a brief free-recall paradigm during this part of the study.

All subjects participated in the next part of the study, which was conducted by a second experimenter in a different room. Subjects were told that they would be shown a series of film clips and that subsequently certain (unspecified) cognitive processes would be examined. In this instance, subjects were to have the opportunity to select 10 films from a total pool of 26. Each clip was supposedly about 1 to 2 minutes long and was described by a brief phrase, for example, "rioting at a rock concert," "a downhill ski run," "a fist fight," "a student taking a test." Each film description had been previously rated by 14 raters, 7 men and 7 women, on three dimensions: aggression, interest, and action. There were no gender differences on the ratings, suggesting that it is unlikely that a sex bias was operating in the perception of the films. Interrater agreement was high, averaging 84.9%, indicating that the perceived aggressive content, degree of interest, and amount of action were reasonably consistent across viewers.[1] As an example, on a scale of 0 (contains none at all) to 3 (contains a great deal of), the film described by the phrase "the Mardi Gras festival" received average ratings of 1.38 on aggression, 2.61 on interest, and 2.76 on action; the film described as "a person getting hit by a brick" had average ratings of 2.90 on aggression, 0.84 on interest, and 2.15 on action. The total aggression, interest, and action scores of the 10 films chosen constituted the major dependent variables.[2]

These three dimensions were assessed in an attempt to remove, through covariance techniques, a persistent confound that pervades research on media violence: specifically, that a violent program usually contains a good deal of action and that violence often occurs in a context of other interesting story elements; thus it is difficult to measure the effect of violence alone. This confound was confirmed by pretest scores indicating that aggressive content was significantly correlated ($p < .01$) with both interest ($r = .54$) and action ($r = .81$). By measuring and statistically controlling for the amount of action and interest of each film chosen, it was possible to obtain a relatively unconfounded indication of the preference for viewing films with aggressive content.

Following the choice of films, all subjects were given a questionnaire that assessed their expectations of the films they chose. Subjects were asked to rate the chosen films, as a group, on 7-point bipolar scales measuring interest, humor, violence, action, educational value, and aesthetic value. Experimental subjects only were also given a second questionnaire, which asked them to indicate the motives (e.g., achievement, aggression, dominance, nurturance) and emotions (e.g., pleasure, fear, surprise, disgust, anger) that had been expressed in their earlier "word stories" on a scale of 1 (not expressed at all) to 5 (expressed very strongly). Finally, all subjects were debriefed.

Results

In view of the possibility of gender differences in aggression (see Maccoby & Jacklin,

[1] Interrater agreement was assessed by dichotomizing the aggression rating continuum into low aggression scores (0 & 1) and high aggression scores (2 & 3). The percentage of agreement on each film description was then calculated: for example, if 11 of the 14 raters gave a high score, then the agreement on that film was 78.57%. The percentage of agreement for each film was then used to calculate an overall mean agreement of 84.9%.

[2] All film descriptions, scores, and measures are available upon request from the author.

Table 1
Mean Aggression Scores of Films Chosen

Gender	Fantasy condition		
	Non-aggressive	None	Aggressive
Women			
M	1.24	1.07	1.09
n	13	22	10
Men			
M	1.28	1.49	1.71
n	10	20	12

Note. Scores range from 0 (low aggression) to 3 (high aggression).

1974), sex was used as a variable in the analyses.

Success of the fantasy manipulation. The word-story method used in the present study had no prior validation as a technique for inducing fantasies. Thus, it was necessary to determine whether the fantasy groups differed in terms of the stories they told. When 2 (Nonaggressive, Aggressive Word-Story) × 2 (Male, Female) analyses of variance were performed on the subjects' perceptions of the motives and emotions expressed in their stories, sex had no effect, nor were there any significant interactions. However, subjects using aggressive words perceived their stories as expressing significantly more fear ($p < .004$), disgust ($p < .002$), anger ($p < .001$), and hatred ($p < .001$), and significantly less pleasure ($p < .004$), happiness ($p < .001$) and love ($p < .04$) than those using nonaggressive words. Similarly, aggressive-word subjects felt their stories had indicated greater needs for aggression ($p < .001$) and dominance ($p < .001$) and lesser needs for nurturance ($p < .01$) and affiliation ($p < .02$) than the nonaggressive-word subjects.

In addition, preliminary content analyses of the fantasies leave little doubt that the word-story technique is a successful means of inducing specific fantasies. In view of this success, it may be well to point out that this method is considerably easier to administer and score and, perhaps more important, may offer greater experimental control over the nature of the fantasy induced, than other well established fantasy instruments such as the Thematic Apperception Test (TAT).

Aggressive content of chosen films. The total aggression scores of the 10 films chosen is presented in Table 1. A 3 (Aggressive, Nonaggressive, or No Word-Story) × 2 (Male, Female) analysis of variance on these scores revealed a main effect for sex, $F(1, 81) = 39.5$, $p < .001$, and a significant Sex × Fantasy interaction, $F(2, 81) = 6.6$, $p < .002$. Films chosen by men contained significantly more aggression than those chosen by women. Planned comparisons showed that when induced to express aggressive fantasies, men had a far stronger preference for observing violence than women, $t(20) = 5.56$, $p < .001$. Also, in the absence of any induced fantasy, men were more likely than women to prefer films containing violence, $t(40) = 4.49$, $p < .001$.

Of more interest to the hypotheses of the present study is the finding that aggressive fantasies, compared to nonaggressive fantasies, increased the amount of violence in the films chosen by males, $t(20) = 4.0$, $p < .001$, but had no effect on the violent content of films chosen by women.

Interest and action levels of chosen films. The films chosen by males contained more action, $F(1, 81) = 16.2$, $p < .001$, and were more interesting, $F(1, 82) = 5.37$, $p < .03$, than those chosen by women. However, these preferences for action and interest in films chosen were not affected by the type of fantasy induced, nor were there any significant interactions between the sex and fantasy variables on these dimensions.

Covariation. In view of both the sex effects for action and interest content of the films and the previously reported intercorrelations between aggression, interest, and action content of the films, a 3 × 2 covariance analysis was performed on the total aggression scores of the films chosen, using action and interest scores as covariables. This analysis yielded a significant effect for sex, $F(1, 79) = 8.00$, $p < .006$, and a significant Sex × Fantasy interaction, $F(2, 79) = 5.38$, $p < .008$, suggesting that the preferences for violent film content operate independently of the films' action and interest content.

Discussion

Gender effects. It may be argued that aggression in contemporary society is essentially a masculine behavior, and the present study found that this gender difference apparently generalizes to film preferences. Several reasons for this difference may be suggested. Socialization strongly affects the expression and acceptability of aggression (Bandura, 1973), resulting in a greater frequency of physically aggressive acts on the part of males than females. This increased familiarity with and exposure to aggressive scenes may heighten the attractiveness of aggressive stimuli (Zajonc, 1968). Men's preferences for viewing violence may also be partially understood in terms of the appropriateness of the role models and behavior patterns displayed in such media presentations. Given these socialization expectations, the results of the present study may be offered in partial support of the validity of the film preference measure.

The sex differences that emerged in this study may also be used to address the issue of demand characteristics. It may be argued that the words given to the aggressive fantasy group provided cues as to the major hypothesis of the study. But since there was no evidence that these cues led to differences in the aggressive content of the stories told by men and women, it is unlikely that these same cues could account for the subsequent gender differences in the violent content of the films chosen. That is, a demand characteristics interpretation cannot explain why men and women responded differently in one part of the experiment, but not in another, to the supposed demand cues.

In addition, the debriefing sessions yielded little indication of suspicion. The cover story proved to be effective in that very few subjects made any explicit connection between the fantasy part of the experiment and the film choices. No mention was ever made of the aggressive nature of the word stimuli, and fantasy subjects invariably perceived their stories as being of their own creation. Finally, accurate knowledge of the hypothesis under investigation was never spontaneously offered by any subject. Thus, there is no evidence to support a strong demand effect.

In view of some recent evidence suggesting that sex differences in aggression are decreasing (see Baron, 1977), the finding that preferences for viewing violence was stronger for men than for women may be considered somewhat puzzling. It should be noted, however, that the experimental research showing a decline in sex differences deals primarily with angry aggressive behavior, whereas the present research is concerned with aesthetic preferences for viewing violence; thus it may be that sex differences are diminishing in the domain of instrumental behavior but continue to exist in the area of expressive behavior. In response to these issues, additional data was gathered (as part of another experiment unrelated to the present hypotheses) on the viewing preferences of 21 women and 20 men, assessed with the same measures used in the first study. The gender differences of Experiment 1 were clearly replicated: The films that men chose to view contained more violence than those chosen by women, $F(1, 37) = 50.9$, $p < .001$, again providing evidence for the existence of sex differences in the preference for viewing violence.

Fantasy effects. The present findings support the hypothesis that the expression of aggressive, compared to nonaggressive, thoughts and images would result in corresponding viewing preferences, but this was the case only for men. For women, the degree of preference for violent films was relatively constant across all conditions, and generally lower than men's preferences; this may best be attributed to socialization pressures that produce a consistent indifference or antipathy on the part of women toward violence in the media.

It is clear that men's film preferences were affected by the nature of their preceding fantasy, but the underlying mechanisms are open to speculation. Singer (1966) has shown that the expression of specific fantasies may arouse corresponding motives. Thus, subjects in the aggressive fantasy condition, compared to those in the nonaggressive fantasy condition, may have experienced a heightened motivation toward aggression. It may be argued

that persons who are motivated to aggress show an increased preference for violent media programs under several different circumstances: Given that the observation of violence facilitates aggressive behavior (Geen, 1976), a person in a prepotent state of aggression may seek out media violence; and if there is no expectation or opportunity for aggression, then media violence may be sought out for its symbolic catharsis value (Geen, 1976), as a means of vicariously satisfying the aggressive drive (Pytkowicz, Wagner, & Sarason, 1967; Rosenbaum & DeCharms, 1960).

Alternatively, the underlying explanation may involve cognitive rather than behavioral processes: Fantasy aggression may lead to increased exposure to violent stimuli for purposes of social comparison (Festinger, 1954). That is, one may wish to compare his own visual image of an aggressive event with that provided by the media. Clearly further research is needed to identify the nature of the processes that underlie the relationship between aggressive fantasies and exposure to media violence. But despite the absence of any clear explanatory mechanisms, the significance of the basic phenomenon should not be overlooked: Aggressive fantasies in males have been shown to increase their preference for viewing violence. Thus, an important antecedent of the preference for viewing violence has been uncovered. In addition, the relationship between TV violence and aggression may be viewed from a new perspective: Not only does television violence increase aggressive behavior, but aggression, in the form of fantasies, also leads to a heightened preference for filmed violence.

Experiment 2

The results of the first study left unanswered the question of whether the tendency to observe violence is affected by overt aggressive *behavior*. Earlier correlational studies demonstrating a relationship between observed violence and aggression were concerned specifically with behavioral measures of aggression (e.g., Dominick & Greenberg, 1971; McIntyre & Teevan, 1971). As suggested earlier, a reasonable explanation for this relationship is that aggressive behavior increases the preference for viewing violence. Several mediating processes may be proposed. Persons who behave aggressively may selectively expose themselves to media violence because of a *need to justify* their prior behavior (Aronson, 1976; Brock & Buss, 1962; Glass, 1964). Watching others act aggressively may provide that justification by intimating that aggression is an appropriate, acceptable behavior (Bandura, 1973; Geen, 1976). In addition, aggression may arouse the need to understand one's aggressive behavior, and that need may be satisfied through *social comparison* processes (Festinger, 1954), leading to the observation of other's aggressive behavior on film. In this context, it is interesting to note that exposure to sexually explicit stimuli is also believed to be a function of social comparison needs; that is, it provides a way of comparing ourselves with others (Byrne, 1976).

Thus, the second study hypothesized that physical aggression would increase the preference for viewing violent film stimuli. Fantasy aggression was also included as a variable in this study because of the desirability of replicating the findings of Experiment 1.

Method

Subjects. The subjects were 64 male undergraduates who participated as part of a course requirement in introductory psychology.

Overview. There were three parts to the study. Subjects first had the opportunity to express either an aggressive or nonaggressive fantasy; a third group did not undergo the fantasy induction. In the second part, half the subjects delivered supposedly painful noises to another person; the other subjects did not use noises. Finally, all subjects indicated their film preferences.

Procedure. Prior to the experiment, informed consent concerning the use of noise was obtained from all subjects, and they were informed of their right to withdraw at any time without penalty. Subjects were scheduled individually and participated with another male subject (actually a confederate). The experiment was described as being interested in the effects of "verbal, visual, and evaluative stimuli on thought processes." It was explained that the participants would work together in some parts of the experiment and separately in other parts. The

st part involved individual participation, and the bject was asked to remain in the room while the nfederate was sent elsewhere. This part of the periment was described to the subject as examin- g the "effects of verbal stimuli on thought pro- esses," and was intended to induce aggressive and onaggressive fantasies, using the same word-story ntasy induction technique as in Experiment 1. gain, subjects were divided into (a) an aggressive ord group, (b) a nonaggressive word group, and c) a no fantasy group that participated in a brief ow-recall paradigm during this part of the study.

After completing the first part, all subjects were ent to a second room, where another experimenter blind to the subjects' fantasy condition) conducted he next part of the study. During this time, the ubject and confederate participated together. For ubjects in the aggression condition, this portion of the experiment was described as looking at how "evaluative feedback in the form of unpleasant oises affects creative thinking ability." Subjects were supposedly to deliver noxious noises to the confed- rate as a means of evaluating the creativity of his esponses in a free-association task (cf. Konečni, 975). The subject read from a list of words, and the nfederate was expected to give a "one word esponse that is as creative as possible." It was explained that it was completely up to the subject o "decide how creative each response is." This was o be communicated through the use of 10 buttons n a Buss (1961) type aggression apparatus that an deliver 10 different levels of "static noise" to he confederate through earphones. The subject was o provide one evaluation for each response. If the nfederate's response was not creative, the subject would indicate that by giving a high level of noise; f the response was very creative, a low level of oise would be chosen. The subject was reminded hat although the noises may be "painful, they are ompletely controlled and are in no way dangerous." (Of course, no noises were actually used.)

The subject was then told that "we have a reason- able understanding of the effects of very high and very low levels of noise, so what would be of most value now would be to examine the effects of moderate noise levels, say Buttons 4 through 7. Although this restricts the range of your evaluations to some degree, you still have the freedom to use any button in that intermediate range." (It must be emphasized that in this study aggression was not a dependent variable, but rather an independent variable. The experiment was not interested in mea- suring aggression; rather it was attempting to *control* aggressive behavior. By manipulating the subjects' range of responses, it was insured that all subjects in this condition would believe that they were de- livering a moderate amount of noxious stimulation to another and thus would be acting aggressively [Buss, 1961]. Pilot work had indicated that these instructions were successful in controlling the sub- jects' aggressive responses to an average noise level of about 5).

Subjects in the no aggression condition were told that this second part of the experiment examined how "another's evaluation affects creative thinking ability." They were given essentially the same in- structions concerning the evaluation of the confed- erates' creative responses and used the same apparatus as aggression subjects did, but control subjects were told that their evaluations were indicated on the confederate's light panel that corresponded to the subject's button panel. Confederates in this case did not wear earphones, and there was no mention of the use of noise. As before, subjects were informed that Button 1 indicated a highly creative response and Button 10 a noncreative response. Again, it was explained that the experiment was interested in moderate (4 to 7) levels of evaluation. In both condi- tions, confederates surreptitiously recorded the levels used by the subject, and the experimenter briefly questioned each subject concerning his per- ception of the button's effects. This information was again elicited during the final debriefing session. In all cases, subjects in the noise condition indicated awareness of the aversiveness of their responses, whereas no such information was offered by subjects in the no noise condition.

After the second part of the study had been com- pleted, the subject alone was sent to a third experi- menter, who was blind as to the subject's experi- mental condition. As in Experiment 1, the subject was told that this part of the study was interested in the "effects of visual stimuli on thought pro- cesses." The procedure for choice of films used in the first study was followed here. Following the choice of films, all subjects filled out questionnaires assessing their expectations of the chosen films, and exerimental subjects also indicated the motives and emotions expressed in their fantasies. After completion of the experiment, participants were fully debriefed. Again, there was little indication of suspicion, and no subject was able accurately to ascertain the actual hypothesis of the study.

Results

Fantasy manipulation. The success of the fantasy induction was determined by perform- ing 2 (Nonaggressive, Aggressive Word-Story) × 2 (No Noise, Noise) analyses of variance on the subjects' perceptions of the motives and emotions expressed in their stories. The noise variable had no effect, and there were no significant meaningful inter- actions. There were strong effects, however, for the word-story manipulation, and these were virtually identical with the results of the first study. Subjects using aggressive words perceived their stories as expressing signifi- cantly more fear ($p < .004$), disgust ($p < .001$), anger ($p < .001$), and hatred ($p < .001$), and significantly less pleasure ($p <$

Table 2
Mean Aggression Scores of Films Chosen

Stimulus	Fantasy condition: Non-aggressive	None	Aggressive
No noise			
M	1.28	1.41	1.48
n	12	8	12
Noise			
M	1.63	1.78	1.80
n	12	8	12

Note. Scores range from 0 (low aggression) to 3 (high aggression).

.001), happiness, ($p < .001$) and love ($p < .001$). Again, subjects in the aggressive fantasy condition felt that their stories had indicated greater needs for aggression ($p < .001$) and dominance ($p < .001$) and lesser needs for nurturance ($p < .001$) and affiliation ($p < .002$) than subjects in the nonaggressive fantasy condition believed their stories had shown. Thus, the success of the fantasy manipulation was well established.

Film preferences. The major dependent variable was the total aggression scores of the films chosen (see Table 2). A 3 (Aggressive, Nonaggressive, or No Word-Story) × 2 (No Noise, Noise) analysis of variance yielded an effect for noise, $F(1, 58) = 21.0$, $p < .001$, and a marginally significant main effect for fantasy, $F(2, 58) = 2.46$, $p < .10$; there was no interaction between the aggression and fantasy variables on this measure. Subjects who had had an opportunity to aggress physically using noise clearly had a stronger preference for viewing violent films than did subjects who had no opportunity to aggress; planned comparisons showed that this effect was consistently significant across all fantasy conditions.

To assess the effects of the fantasies on film preferences more clearly, an additional analysis of variance was performed involving only those groups that had expressed fantasies. This analysis indicated that there was a greater preference for viewing violence among the aggressive fantasy subjects than among the nonaggressive fantasy subjects, $F(1, 44) = 4.18$, $p < .05$. Planned comparisons found that in the no noise condition (which is comparable to Experiment 1), there was a greater amount of violence in the films chosen by men who expressed an aggressive fantasy than in the films selected by those in the nonaggressive fantasy condition, $t(22) = 1.91$, $p < .05$.

Covariance analyses. Films chosen by subjects who had physically aggressed involved more action, $F(1, 58) = 15.4$, $p < .001$, and were more interesting, $F(1, 58) = 6.53$, $p < .02$, than did those chosen by the nonaggression group. Fantasy had no effect on preferences for action or interest, nor were there any interactions. These aggression effects suggested the need for a 3 × 2 covariance analysis, using interest and action scores as the covariates, in order to determine whether preferences for viewing violent films were independent of the films' action and interest content. The results of the analysis indicated that this was the case: There was now a significant effect for fantasy, $F(2, 56) = 3.26$, $p < .05$, as well as an effect for physical aggression, $F(1, 56) = 5.70$, $p < .02$.

General Discussion

Empirical Findings

It was hypothesized that both physical aggression and fantasy aggression would lead to an increased preference for viewing violence. These predictions were confirmed for men. Experiment 2 effectively replicated the first study: Both found that for males, the expression of an aggressive fantasy, compared to a nonaggressive fantasy, led to a greater preference for viewing violence. In the second study, this effect was sustained despite the fact that another time-consuming and impactful manipulation interceded between the fantasy condition and the dependent measure. Experiment 2 also found that persons who physically aggressed toward another, using noxious noise, were far more likely to choose to view violence than were those who had not aggressed.

Although the present studies represent the first experimental demonstration that aggressive behavior leads to the viewing of violence,

his causal hypothesis is completely consistent with previous correlational work demonstrating a positive relationship between TV violence and aggression. Several other studies, conducted in both the laboratory and the field, may also be interpreted in terms of the major hypotheses of the present research. Doob and Macdonald (1979) found that people in high crime areas watched more violent TV than those in low crime areas. Similarly, Boyanowsky, Newtson, and Walster (1974) found that attendance at a violent movie increased markedly following a brutal local murder, whereas no such increase in attendance occurred for a nonviolent film. A possible explanation for both these findings is that violent crime results in a heightened preoccupation with aggressive thoughts and fantasies, which then leads persons to seek out violence in the media. Feshbach, Stiles, and Bitter (1967) showed that witnessing acts of aggression may have reinforcing effects, but only for those who have been insulted. Again, this is consistent with the notion that persons seek out media portrayals of violence following an episode in which they experienced aggressive thoughts or feelings. Finally, it has been reported that low-income and black youths enjoy watching filmed violence more than middle-class youths do (Greenberg & Gordon, 1971); and dispositionally aggressive men have a greater preference for viewing violence than do those who are nonaggressive (Diener & DeFour, 1978). Presumably, aggression in one's own life results in congruent viewing preferences. Taken together with the present research, these studies provide considerable evidence for a theory of viewer behavior: Persons watch violence in the media because of both long-term and situationally induced aggressive thoughts and actions.

Implications

The problem of violence in the mass media has attracted a great deal of public, governmental, and scientific attention. That attention has been primarily focused on the *consequences* of viewing violence in the mass media, with little research having been directed toward the *antecedents* of the viewing of violence. Yet, in a sense, the two are inseparable: in order to understand the effects of the mass media, it is necessary to understand why viewers watch the media; and conversely, to understand viewing behavior, the effects of the mass media must be known. Cognizant of the fact that watching violence facilitated aggression, the present research assumed that aggression, both physical and fantasy, was related to the viewing of violence. The results suggest that the assumptions were warranted: The preference for watching violence may be attributed, in part, to aggression-related behaviors on the part of the viewer.

The present research also provides an additional perspective on the relationship between the observation of violence and the expression of aggression: aggressive behaviors and fantasies are not only a *result* of viewing violence, as established by previous research (e.g., Eron, et al., 1972; Feshbach, 1961), but may also be a *cause* of the viewing of violence. Just as what one watches in the media influences one's behavior, so too the way one behaves influences what one watches. This bidirectional model of influence argues for a cyclical pattern of effects such that aggressive thoughts and actions increase one's preference for viewing violence in the mass media, which in turn increases aggressive behavior.

Finally, the research helps to counter some of the criticism that has been directed at the form of mass media research. Because previous experimental research on television violence and aggression was interested in the effects of watching violence, the typical research paradigm denied the participant the *choice* of what to watch; subjects were simply exposed to predetermined programs, and effects were measured. This procedure, although standard (e.g., Friedrich & Stein, 1973; Liebert & Baron, 1972), may be considered ecologically invalid in that it disregards tendencies or motives that may determine the choice of program and may interact with the program content in affecting subsequent behavior. The present research partially overcomes these criticisms by investigating and identifying some of the variables

that affect what the viewer selects to watch, thus directing attention to the reciprocal nature of the relationship between aggressive behavior and mass media violence.

References

Aronson, E. *The social animal*. San Francisco: Freeman, 1976.

Baker, R. K., & Ball, S. J. *Violence and the media: A staff report to the National Commission on the Causes and Prevention of Violence*. Washington, D.C.: U.S. Government Printing Office, 1969.

Bandura, A. *Aggression: A social learning analysis*. Englewood Cliffs, N.J.: Prentice-Hall, 1973.

Bandura, A., Ross, D., & Ross, S. A. Transmission of aggression through imitation of aggressive models. *Journal of Abnormal and Social Psychology*, 1961, *63*, 575–582.

Baron, R. A. *Human aggression*. New York: Plenum Press, 1977.

Berkowitz, L. Some aspects of observed aggression. *Journal of Personality and Social Psychology*, 1965, *2*, 359–369.

Berkowitz, L., & Geen, R. G. Film violence and the cue value of available targets. *Journal of Personality and Social Psychology*, 1966, *3*, 525–530.

Boyanowsky, E. O., Newtson, D., & Walster, E. Film preferences following a murder. *Communications Research*, 1974, *1*, 32–43.

Brock, T. C., & Buss, A. H. Dissonance, aggression, and evaluation of pain. *Journal of Abnormal and Social Psychology*, 1962, *65*, 197–202.

Buss, A. H. *The psychology of aggression*. New York: Wiley, 1961.

Byrne, D. Sexual imagery. In J. Money & H. Musaph (Eds.), *Handbook of sexology*. Amsterdam: Excerpta Medica, 1976.

Chaffee, S. H., & McLeod, J. M. Adolescents, parents, and televised violence. Cited in R. Liebert, J. M. Neale, & E. S. Davidson. *The early window: Effects of television on children and youth*. New York: Pergamon Press, 1973.

Clark, D. G., & Blankenburg, W. B. Trends in violent content in selected mass media. In G. A. Comstock & E. A. Rubinstein (Eds.), *Television and social behavior* (Vol. 1). Washington, D.C.: U.S. Government Printing Office, 1971.

Diener, E., & DeFour D. Does television violence enhance program popularity? *Journal of Personality and Social Psychology*, 1978, *36*, 333–341.

Dominick, J. R., & Greenberg, B. S. Attitudes toward violence: The interaction of television exposure, family attitudes, and social class. In G. A. Comstock & E. A. Rubinstein (Eds.), *Television and social behavior* (Vol. 3). Washington, D.C.: U.S. Government Printing Office, 1971.

Doob, A. N., & Macdonald, G. E. Television viewing and fear of victimization: Is the relationship causal? *Journal of Personality and Social Psychology*, 1979, *37*, 170–179.

Eron, L. D., Lefkowitz, M. M., Huesmann, L. R., & Walder, L. D. Does television violence cause aggression? *American Psychologist*, 1972, *27*, 253–263.

Eysenck, H. J., & Nias, D. K. B. *Sex, violence, and the media*. New York: Harper & Row, 1978.

Feshbach, S. The stimulating versus cathartic effects of a vicarious aggressive activity. *Journal of Abnormal and Social Psychology*, 1961, *63*, 381–385.

Feshbach, S. The role of fantasy in the response to television. *Journal of Social Issues*, 1976, *32*, 71–85.

Feshbach, S., & Singer, R. D. *Television and aggression*. San Francisco: Jossey-Bass, 1971.

Feshbach, S., Stiles, W. B., & Bitter, E. Reinforcing effects of witnessing aggression. *Journal of Research in Personality*, 1967, *2*, 133–139.

Festinger, L. A theory of social comparison processes. *Human Relations*, 1954, *7*, 117–140.

Friedrich, L. K., & Stein, A. H. Aggressive and prosocial television programs and the natural behavior of preschool children. *Monograph of the Society for Research in Child Development*, 1973, *38*, 1–64.

Geen, R. G. Observing violence in the mass media: Implications of basic research. In R. G. Geen & E. O'Neal (Eds.), *Perspectives on aggression*. New York: Academic Press, 1976.

Gerbner, G. Violence in television drama: Trends and symbolic functions. In G. A. Comstock & E. A. Rubinstein (Eds.), *Television and social behavior* (Vol. 1). Washington, D.C.: U.S. Government Printing Office, 1971.

Glass, D. C. Changes in liking as a means of reducing cognitive discrepancies between self-esteem and aggression. *Journal of Personality*, 1964, *32*, 531–549.

Greenberg, B. S., & Gordon, T. F. Social class and racial differences in children's perceptions of televised violence. In G. A. Comstock, E. A. Rubinstein, & J. P. Murray (Eds.), *Television and social behavior*. Washington, D.C.: U.S. Government Printing Office, 1971.

Kaplan, R. M., & Singer, R. D. Television violence and viewer aggression: A reexamination of the evidence. *Journal of Social Issues*, 1976, *32*, 35–70.

Konečni, V. J. Annoyance, type, and duration of post-annoyance activity, and aggression: The "cathartic effect." *Journal of Experimental Psychology: General*, 1975, *104*, 76–102.

Lefkowitz, M. M., Eron, L. D., Walder, L. D., & Huesmann, L. R. Television violence and child aggression: A follow-up study. In G. A. Comstock & E. A. Rubinstein (Eds.), *Television and social behavior* (Vol. 3). Washington, D.C.: U.S. Government Printing Office, 1971.

Leifer, A. D., & Roberts, D. F. Children's responses to television violence. In J. P. Murray, E. A. Rubinstein, & G. A. Comstock (Eds.), *Television and social behavior* (Vol. 2). Washington, D.C.: U.S. Government Printing Office, 1971.

Liebert, R. M., & Baron, R. A. Some immediate effects of televised violence on children's behavior. *Developmental Psychology*, 1972, *6*, 469–475.

McIntyre, J., & Teevan, J. Television violence and deviant behavior. In G. A. Comstock & E. A. Rubinstein (Eds.), *Television and social behavior* (Vol. 3). Washington, D.C.: U.S. Government Printing Office, 1971.

McLeod, J. M., Atkin, C. K., & Chaffee, S. H. Adolescents, parents, and television use. In G. A. Comstock & E. A. Rubinstein (Eds.), *Television and social behavior* (Vol. 3). Washington, D.C.: U.S. Government Printing Office, 1971.

Maccoby, E. E., & Jacklin, C. N. *The psychology of sex differences*. Stanford, Calif.: Stanford University Press, 1974.

Milgram, S., & Shotland, R. L. *Television and antisocial behavior: Field experiments*. New York: Academic Press, 1973.

Murray, J. P. Television and violence: Implications of the Surgeon-General's research program. *American Psychologist*, 1973, *28*, 472–478.

Pytkowicz, A., Wagner, N., & Sarason, I. G. An experimental study of the reduction of hostility through fantasy. *Journal of Personality and Social Psychology*, 1967, *5*, 295–303.

Rosenbaum, M., & DeCharms, R. Direct and vicarious reduction of hostility. *Journal of Abnormal and Social Psychology*, 1960, *60*, 105–111.

Singer, J. L. *Daydreaming*. New York: Random House, 1966.

Stein, A. H., & Friedrich, L. K. Television content and your children's behavior. In J. P. Murray, E. A. Rubinstein, & C. A. Comstock (Eds.), *Television and social behavior* (Vol. 2). Washington, D.C.: U.S. Government Printing Office, 1971.

Steuer, F. B., Applefield, J. M., & Smith, R. Televised aggression and the interpersonal aggression of preschool children. *Journal of Experimental Child Psychology*, 1971, *11*, 442–447.

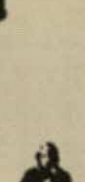
Weiss, W. Effects of the mass media of communication. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (Vol. 5). Reading, Mass.: Addison-Wesley, 1969.

Zajonc, R. Attitudinal effects of mere exposure. *Journal of Personality and Social Psychology Monograph*, 1968, *9*(1, Pt. 2).

Received March 29, 1979 ■

GRADUATE STUDY IN PSYCHOLOGY 1980-1981

Prospective psychology graduate students and college counselors will find *Graduate Study in Psychology for 1980-1981* an indispensable resource. This 13th annual edition published by the American Psychological Association provides 656 pages of up-to-date, specific information on more than 550 graduate programs in the United States and Canada. Listed for each institution are application procedures, admission requirements, tuition, financial assistance, internships, and minority considerations. General information on applying to graduate school is included to help with that important decision about graduate study. Price: $7.



To order, write to:
American Psychological Association, Order Dept.
1200 Seventeenth Street, NW
Washington, D.C. 20036

Orders of $25 or less must be prepaid.

JOURNAL OF
Personality AND Social Psychology

Volume 37 1979

Editor, Issues 1–6

ANTHONY G. GREENWALD
Ohio State University

Topic Editors

IRWIN ALTMAN
University of Utah

ROBERT A. BARON
Purdue University

DARYL BEM
Stanford University

EUGENE BURNSTEIN
University of Michigan

ROBERT B. CIALDINI
Arizona State University

REID HASTIE
Harvard University

DOUGLAS N. JACKSON
University of Western Ontario
London, Canada

LESLIE Z. MCARTHUR
Brandeis University

Acting Editor, Issues 7–12

CLYDE HENDRICK
University of Miami

Action Editors

Irwin Altman
University of Utah

Charles S. Carver
University of Miami

Richard J. Harris
University of New Mexico

John H. Harvey
Vanderbilt University

William Ickes
University of Missouri—
St. Louis

George Levinger
University of Massachusetts—
Amherst

Brendan Gail Rule
University of Alberta,
Edmonton, Canada

David R. Shaffer
University of Georgia

Cookie White Stephan
New Mexico State University

Barbara Strudler Wallston
George Peabody College for
Teachers

Published monthly by the
American Psychological Association, Inc.
1200 Seventeenth Street, N.W., Washington, D.C. 20036

Copyright © 1979 by the American Psychological Association, Inc.

Consulting Editors, Issues 1–6

Icek Ajzen, *University of Massachusetts—Amherst*
Robert S. Baron, *University of Iowa*
Gordon Bear, *Ramapo College of New Jersey*
Ellen Berscheid, *University of Minnesota*
Michael Birnbaum, *University of Illinois at Urbana-Champaign*
Philip Brickman, *Institute for Social Research*
J. Merrill Carlsmith, *Stanford University*
John S. Carroll, *Loyola University of Chicago*
Jerry L. Cohen, *University of Illinois at Urbana-Champaign*
Sheldon Cohen, *University of Oregon*
Rick Crandall, *Texas Christian University*
William D. Crano, *Michigan State University*
Robyn M. Dawes, *University of Oregon*
Kay K. Deaux, *Purdue University*
Marshall Dermer, *University of Wisconsin—Milwaukee*
Edward Diener, *University of Illinois at Urbana-Champaign*
Anthony Doob, *University of Toronto, Canada*
Susan V. Eisen, *McLean Hospital, Belmont, Mass.*
Baruch Fischhoff, *Decision Research, Eugene, Ore.*
Irene H. Frieze, *University of Pittsburgh*
David Funder, *Stanford University*
David L. Hamilton, *University of California, Santa Barbara*
Ranald D. Hansen, *Oakland University*
Richard J. Harris, *University of New Mexico*
Edward Helmes, *University of Western Ontario, Canada*
Donald S. Hiroto, *VA Hospital, Brentwood, Los Angeles, Calif.*
Edward E. Jones, *Princeton University*
Martin F. Kaplan, *Northern Illinois University*
Douglas T. Kenrick, *Montana State University*
Norbert L. Kerr, *University of California, San Diego*
Samuel Komorita, *University of Illinois at Urbana-Champaign*
Vladimir J. Konečni, *University of California, San Diego*
Robert M. Krauss, *Columbia University*
Ellen J. Langer, *Harvard University*
Mark R. Lepper, *Stanford University*
Melvin Manis, *University of Michigan (Ann Arbor)*
Hazel Markus, *University of Michigan (Ann Arbor)*
Margaret Matlin, *SUNY College at Geneseo*
Dale T. Miller, *University of British Columbia, Canada*
Thomas M. Ostrom, *Ohio State University*
Paul B. Paulus, *University of Texas at Arlington*
Richard Petty, *University of Missouri—Columbia*
Harry T. Reis, *University of Rochester*
Judith Rodin, *Yale University*
Seymour Rosenberg, *Livingston College, Rutgers*
Ralph L. Rosnow, *Temple University*
Lee Ross, *Stanford University*
Michael Ross, *University of Waterloo, Canada*
Diane Ruble, *Princeton University*
Michael J. Saks, *Nat'l Center for State Courts, Williamsburg, Va.*
Michael F. Scheier, *Carnegie-Mellon University*
Barry R. Schlenker, *University of Florida*
Philip Shaver, *New York University*
Steven J. Sherman, *Indiana University*
Jerome E. Singer, *University of the Health Services, Bethesda, Md.*
Eliot R. Smith, *University of California, Riverside*
Mark Snyder, *University of Minnesota*
Melvin L. Snyder, *University of Texas at Austin*
Lawrence Stricker, *Educational Testing Service*
Shelley E. Taylor, *Harvard University*
Abraham Tesser, *University of Georgia*
Bernard Weiner, *University of California, Los Angeles*
Gary L. Wells, *University of Alberta, Canada*
Stephen Worchel, *University of Virginia*
Robert S. Wyer, *University of Illinois at Urbana-Champaign*
Mark Zanna, *University of Waterloo, Canada*
Miron Zuckerman, *University of Rochester*

Holly Wong, *Assistant to the Editor*

Consulting Editors, Issues 7–12

Robert S. Baron, *University of Iowa*
Paul A. Bell, *Colorado State University*
Ellen Berscheid, *University of Minnesota (Minneapolis)*
Michael H. Birnbaum, *University of Illinois at Urbana-Champaign*
Jack Brehm, *University of Kansas*
Philip Brickman, *Northwestern University*
J. Merrill Carlsmith, *Stanford University*
Paul C. Cozby, *California State University, Fullerton*
Rick Crandall, *Texas Christian University*
James M. Dabbs, Jr., *Georgia State University*
John M. Darley, *Princeton University*
Andrew R. Davidson, *University of Washington*
James H. Davis, *University of Illinois at Urbana-Champaign*
Kay Deaux, *Purdue University*
Marshall Dermer, *University of Wisconsin—Milwaukee*
Edward Donnerstein, *Iowa State University*
Ebbe B. Ebbesen, *University of California, San Diego*
Yakov M. Epstein, *Rutgers—The State University*
Irene H. Frieze, *University of Pittsburgh*
J. Martin Giesen, *Mississippi State University*
Lewis R. Goldberg, *University of Oregon*
William G. Graziano, *University of Georgia*
Alan E. Gross, *University of Maryland (College Park)*
Nancy Hirschberg, *University of Illinois at Chicago Circle*
Chester A. Insko, *University of North Carolina at Chapel Hill*
Alice M. Isen, *University of Maryland, Baltimore County*
Martin F. Kaplan, *Northern Illinois University*
Samuel S. Komorita, *University of Illinois at Urbana-Champaign*
Vladimir J. Konečni, *University of California, San Diego*
Ellen Langer, *Harvard University*
Mark R. Lepper, *Stanford University*
Darwyn E. Linder, *Arizona State University*
Dale T. Miller, *University of British Columbia, Vancouver, Canada*
Judson Mills, *University of Maryland (College Park)*
Warren T. Norman, *University of Michigan (Ann Arbor)*
Michael Ross, *University of Waterloo, Waterloo, Canada*
Michael F. Scheier, *Carnegie-Mellon University*
Barry R. Schlenker, *University of Florida*
Kelly G. Shaver, *College of William & Mary*
Phillip Shaver, *New York University*
Marvin E. Shaw, *University of Florida*
Steven J. Sherman, *Indiana University*
Harold Sigall, *University of Maryland (College Park)*
Mark Snyder, *University of Minnesota (Minneapolis)*
Melvin L. Snyder, *University of Texas at Austin*
Shelley E. Taylor, *Harvard University*
Stephen G. West, *Florida State University*
Steve Worchel, *University of Virginia*
Mark P. Zana, *University of Waterloo, Waterloo, Canada*

Sally Miller, *Assistant to the Editor*

The Acting Editor acknowledges with thanks the advice and consultation provided by the following persons, who served recently as ad hoc reviewers of manuscripts. The number of manuscripts is in parentheses, if more than one.

Lyn Y. Abramson (4)
Gerald R. Adams (3)
John R. Aiello
William L. Allbutt
Henry Alker
Norman H. Anderson (3)
Richard Archer
Robert Arkin (4)
David J. Armor
John A. Arrowood
Richard D. Ashmore
William Austin (4)
James R. Averill (4)

Gerald G. Bachman
Roger Bakeman (3)
A. Harvey Baker (2)
Reuben M. Baron
Robert A. Baron (3)
Robert S. Baron
John Barresi (2)
Ronald K. Barrett
Andrew Baum (3)
Gordon Bear
Lee A. Becker
Paul A. Bell (3)
Peter L. Benson (2)
Thomas J. Berndt (2)
Sandy Bernstein
Ellen S. Berscheid (4)
Michael H. Birnbaum (3)
Brian F. Blake
Fletcher A. Blanchard
Paul H. Blaney
James J. Blascovich
Thomas Blass (2)
Jack Block
Jeanne H. Block
Brian Bolton
Beth BonTempo
Eugene Borgida
Jack W. Brehm (3)
Sharon S. Brehm
Marilynn B. Brewer
Philip Bickman (2)
John C. Brigham (3)
Timothy C. Brock
Gene Brody
Thomas Brothen
James Bryan
Brenda K. Bryant
Jeff B. Bryson
Robert Buckhout
Daphne Blunt Bugental
Gary K. Burger (2)
Peter J. Burke
Clarke A. Burnham
Eugene Burnstein
Lynn E. Bush II
Donn Byrne (3)

Lawrence G. Calhoun
Donald Campbell
Joseph J. Campos
Nancy Cantor (2)
G. Craig Caputo
Donal E. Carlston
Jerome M. Chertkoff
Robert B. Cialdini (3)
Russell D. Clark, III
Jerry L. Cohen
Ronald L. Cohen
Sheldon Cohen (2)
Stanley H. Cohen
Barry E. Collins
John Condry
Judy Corder-Bolz
Philip R. Costanzo
Paul C. Cozby (3)
Rick Crandall (4)
William D. Crano
Faye Crosby
Herbert J. Cross
John D. Cunningham (2)

James M. Dabbs, Jr. (2)
Anthony R. D'Augelli
Andrew R. Davidson (2)
Christine V. Davidson (2)
Emily S. Davidson (2)
Deborah Davis
James H. Davis
Keith E. Davis (2)
Robyn M. Dawes
Kay Deaux (10)
Richard De Charms
Edward L. Deci
Theodore Dembroski
Bella M. DePaulo
Valerian J. Derlega
Marshall Dermer
Brenda McEvoy DeVellis
Edward Diener (2)
Richard A. Dienstbier
Robert L. Dipboye
Gordon J. DiRenzo
Rod K. Dishman
Barbara Snell Dohrenwend
Ed Donnerstein (2)
Leslie L. Downing
Joseph DuCette
Starkey Duncan

Alice H. Eagly (2)
Ebbe B. Ebbesen (3)
Hillel J. Einhorn
Nancy Eisenberg-Berg
Richard Eisler
Phoebe Ellsworth (2)
Donald Elman (3)
Alan C. Elms
Michael Enzle (4)
Seymour Epstein
Yakov M. Epstein (2)
Franz R. Epting
James K. Esser (2)
Dorcas Evans
Gary W. Evans
Ralph V. Exline

Russell H. Fazio (2)
N. T. Feather
Richard Feinberg (2)
Robert H. L. Feldman
Robert S. Feldman (2)
Fred E. Fiedler
Baruch Fischhoff (2)
Jeffrey D. Fisher (2)
William A. Fisher
Susan T. Fiske (3)
Uriel G. Foa
Robert Folger (2)
Jonathan L. Freedman
Howard S. Friedman (2)
Irene H. Frieze (8)
William J. Froming

Samuel L. Gaertner
Howard Garland
Gerald T. Gardner
Harold B. Gerard
Frederick X. Gibbons
Martin Giesen (2)
George R. Goethals
Therese E. Goetz
Lewis R. Goldberg (2)
Jeffrey H. Goldstein (2)
Kenneth M. Goldstein
Jacqueline D. Goodchilds
Cheryl C. Graesser (3)
William G. Graziano (5)
Jerald Greenberg (3)
Alan E. Gross (4)
Charles L. Gruder
Joan E. Grusec (2)
Barbara A. Gutek

Norma Haan
Karl L. Hakmiller
Judith A. Hall (3)
David L. Hamilton (4)
V. Lee Hamilton (2)
Ranald D. Hansen (2)
Barbara Hartman Hanusa (2)
Stephen G. Harkins (2)
Ben Harris
Albert A. Harrison
James Harrison
James J. Hart
Willard W. Hartup
Michael D. Harvey
Reid Hastie
Albert H. Hastorf
Fred Heilizer
Robert L. Helmreich (2)
Donald E. Henson
C. Peter Herman
Dick Heslin
Charles T. Hill
Kenneth A. Hill
Samuel Himmelfarb
Steve Hinkle
Richard Hirschman
Martin L. Hoffman
James H. Hogge
Steven D. Hollon
Pauline Houlden (2)
James House
B. Kent Houston
Ted L. Huston (2)
Janet Shipley Hyde (2)
Chester A. Insko (5)
Alice M. Isen (7)
Carroll E. Izard (3)

Melvyn Jaffa
Ronnie Janoff-Bulman
Paula B. Johnson
Thomas Johnson
Shawn Johnston
Russell A. Jones (3)

Arnold Kahn (4)
Daniel Kahneman
David E. Kanouse
Kalman J. Kaplan
Martin F. Kaplan (3)
Stuart A. Karabenick (2)
Rachel Karniol
Martin Katahn
Irwin Katz
H. H. Kelley
Jeffrey A. Kelly
David A. Kenny
Douglas T. Kenrick (2)
Robert F. Kidd (5)
Louise H. Kidder (5)
Donald Kinder
Bert T. King
Eric S. Knowles
Philip L. Knowles
Daniel Kohneman
Samuel S. Komorita (3)
David S. Kranta (2)
Robert E. Kraut
Jaro Křivohlavý
Judith E. Krulewitz
Edward Krupot
Michael Kuhlman
Donald Kuiken

John I. Lacey
Marianne LaFrance (2)
Annette M. La Greca
James D. Laird (2)
Michael E. Lamb
Irving Lane
Ellen Langer (4)
Judith H. Langlois
James R. Larson, Jr.
Stephen A. LaTour
R. Michael Latta
Herbert M. Lefcourt (2)
Gerald J. Leger
Robert K. Leik
Ellen Lenney
Christopher Leone (3)
Mark R. Lepper (3)
Howard Leventhal (5)
George Levinger
Leon H. Levy
E. Allan Lind
Darwyn E. Linder
Rhoda Lindner (2)
Svenn Lindskold
John H. Lingle
Richard Lippa
Charles G. Lord

James Luginbuhl

Jacqueline Macaulay (3)
Daniel B. Madsen
Vivian Parker Makosky (2)
Neil M. Malamuth
Melvin Manis
Jeanne Marecek (3)
Hazel Markus
John C. Masters
Joseph Matarrazo
Karen Matthews (3)
Kevin D. McCaul (2)
David C. McClelland
William J. McGuire
Steven P. McNeel
Karla S. McPherson
Frederic J. Medway (2)
David M. Messick (2)
Charles E. Miller
Dale T. Miller (2)
Judson Mills (4)
Herbert L. Mirels (2)
Thomas C. Monson (5)
Bert S. Moore
Stanley A. Mulaik
Thomas H. Murray

Jack R. Nation
Charlan Nemeth (2)
Kent L. Norman
Warren T. Norman (2)
Rudy V. Nydegger (4)

Virginia E. O'Leary (8)
Anthony B. Olejnik
Christopher Olson
Patrick M. O'Malley
Jacob L. Orlofsky (3)
Thomas M. Ostrom

S. Mark Pancer (2)
Jacquelynne E. Parsons
Miles L. Patterson
Paul B. Paulus (2)
James W. Pennebaker (2)

Louis A. Penner
Albert Pepitone
Letitia Anne Peplau
Daniel Perlman
Sidney I. Perloe
David Perry (2)
Richard E. Petty (4)
Jane Allyn Piliavin (3)
Thane S. Pittman
Alexander Pollatsek
John Pryor

Lenore Sawyer Radloff
D. W. Rajecki
Amnon Rapoport
Marvina C. Rich
Deborah C. Richardson (2)
Joan Riedle
Marc Riess
Judith Rodin
Ronald W. Rogers
T. B. Rogers
Sidney Rosen (2)
Morris Rosenberg
Seymour Rosenberg (2)
Howard M. Rosenfeld
David Rosenfield
Robert Rosenthal
Lee D. Ross (2)
Michael A. Ross (2)
Betty B. Rossman
Susan Roth
Julian B. Rotter
William E. Roweton
Zick Rubin
Diane N. Ruble
James A. Russell
Nancy Felipe Russo
Joseph F. Rychlak (2)
Richard M. Ryckman (3)

William Samuel
John W. Santrock
Leonard Saxe
Candace Garrett Schau
Michael F. Scheier (2)

R. Steven Schiavo
Barry R. Schlenker (4)
Frank W. Schneider
William Schneiderman
Janet Ward Schofield
John Schopler
Samuel Schultz II
Richard Schulz
Shalom Schwartz
Paul Schwieger
Martin E. P. Seligman
Lisa A. Serbin (2)
Jack A. Shaffer
Juliet Popper Shaffer
James Shanteau
Kelley Shaver
Phillip Shaver (3)
Jerry I. Shaw (2)
Marvin E. Shaw
Carolyn Sherif
Steven J. Sherman (2)
Jim I. Shields
Harold Sigall
Roberta G. Simmons
Dean Keith Simonton (2)
Jerome E. Singer (3)
Yolanda S. Slocum
Eliot R. Smith
Roberta Ann Smith
Ronald E. Smith
William P. Smith (2)
Mark Snyder (6)
Melvin L. Snyder (4)
Joy Stapp (5)
Erwin Staub
Larry Staw
Ivan Steiner (2)
Walter G. Stephan (4)
Daniel Stokols (3)
Michael D. Storms
Fred L. Strodtbeck
Stanley Sue
Jerry Suls (2)
Eric Sundstrom (3)
William B. Swann, Jr.

Robert B. Tallarico
Elizabeth Decker Tanke
Dalmas A. Taylor
Shelley E. Taylor
Howard Tennen (2)
Abraham Tesser (6)
Mark H. Thelen
Susan Thomas
Donna E. Thompson (2)
Julie A. Tucker

Robin R. Vallacher (2)
Neil Vidmar
Amiram Vinokur

Kenneth A. Wallston (4)
Rebecca Warner (2)
Gifford Weary (5)
Daniel M. Wegner (4)
John C. Weidman
Bernard Weiner
Shirley Weitz (2)
Gary L. Wells (6)
Carol Werner (2)
Stephen G. West
Gregory L. White
Kathleen M. White
Robert A. Wicklund
Cathy Spatz Widom
Morton Wiener
Jerry S. Wiggins
David A. Wilder
Edwin P. Willems
Raymond A. Winbush
David G. Winter
Thomas M. Wolf
Stephen Worchel (3)
Judith Worell (2)
Lawrence S. Wrightsman (3)

Ben Yandell

Dolf Zillmann (2)
Fred M. Zimring

APA Journal Staff

Anita DeVivo, *Executive Editor*
Ann I. Mahoney, *Manager, Journal Production*

Marjorie M. Henry, *Production Supervisor*
Robert J. Hayward, *Advertising Representative*
Juanita Brodie, *Subscription Manager*

Author Index to Volume 37

Key to Pagination

Issue No.	Month	Pages	Issue No.	Month	Pages
[illegible]	January	1–146	7	July	1097–1249
[illegible]	February	147–314	8	August	1251–1439
[illegible]	March	315–469	9	September	1441–1630
[illegible]	April	471–635	10	October	1631–1931
[illegible]	May	637–810	11	November	1933–2136
[illegible]	June	811–1096	12	December	2137–2317

ARTICLES

Adewole, Akanbi. *See* Insko, Chester A.
Ajzen, Icek, Dalto, Carol Ann, and Blyth, Daniel P. Consistency and bias in the attribution of attitudes 1871
Allen, Robert B. *See* Ebbesen, Ebbe B.
Amabile, Teresa M. Effects of external evaluation on artistic creativity 221
Anderson, Craig A. *See* Carlsmith, J. Merrill.
Anderson, Erling A. *See* Dermer, Marshall.
Anderson, Norman H. Indeterminate theory: Reply to Gollob 950
Anderson, Norman H. *See* Farkas, Arthur J.
Apple, William, Streeter, Lynn A., and Krauss, Robert M. Effects of pitch and speech rate on personal attributions 715
Appelbaum, Mark I. *See* Gray-Little, Bernadette.
Arkkelin, Daniel, Oakley, Thomas, and Mynatt, Clifford. Effects of controllable versus uncontrollable factors on responsibility attributions: A single-subject approach 110
Arnkoff, Diane B. *See* Newcombe, Nora.
Austin, William. Sex differences in bystander intervention in a theft 2110

Babad, Elisha Y. Personality correlates of susceptibility to biasing information 195
Bachman, Jerald G. *See* O'Malley, Patrick M.
Baer, Robert. *See* Miller, Arthur G.
Bagozzi, Richard P., and Burnkrant, Robert E. Attitude organization and the attitude–behavior relationship 913
Baker, A. Harvey, Mishara, Brian L., Kostin, Irene W., and Parker, Laurence. Menstrual cycle affects kinesthetic aftereffect, an index of personality and perceptual style 234
Barden, R. Christopher. *See* Masters, John C.
Barker, Roger G. Settings of a professional lifetime 2137
Bassili, John N. Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face 2049
Baumann, Donald J. *See* Kenrick, Douglas T.
Baumeister, Roy F., Cooper, Joel, and Skib, Bryant A. Inferior performance as a selective response to expectancy: Taking a dive to make a point 424
Baumgardner, Michael H. *See* Lingle, John H.
Beaman, Arthur L., Klentz, Bonnel, Diener, Edward, and Svanum, Soren. Self-awareness and transgression in children: Two field studies 1835
Beck, Pearl. *See* Langer, Ellen J.
Becker, Lawrence J. *See* Brickman, Philip.
Bem, Daryl J., and Lord, Charles G. Template matching: A proposal for probing the ecological validity of experimental settings in social psychology 833
Bem, Sandra Lipsitz. Theory and measurement of androgyny: A reply to the Pedhazur–Tetenbaum and Locksley–Colten critiques 1047
Benassi, Victor A., Sweeney, Paul D., and Drevno, Gregg E. Mind over matter: Perceived success at psychokinesis 1377
Bentler, P. M., and Huba, G. J. Simple minitheories of love 124
Berger, Seymour M., Carli, Linda L., Hammersla, Kathy S., Karshmer, Judith F., and Sanchez, M. Estela. Motoric and symbolic mediation in observational learning 735
Berman, John J. *See* Kahle, Lynn R.

Bernstein, Sandy. *See* Richardson, Deborah Capasso.
Bernstein, William M., Stephan, Walter G., and Davis, Mark H. Explaining attributions for achievement: A path analytic approach 1810
Birnbaum, Michael H., and Mellers, Barbara A. One-mediator model of exposure effects is still viable 1090
Birnbaum, Michael H., and Mellers, Barbara A. Stimulus recognition may mediate exposure effects 391
Birnbaum, Michael H., and Stegner, Steven E. Source credibility in social judgment: Bias, expertise, and the judge's point of view 48
Blaney, Paul H. *See* Carver, Charles S.
Blankenship, Virginia. *See* Kuhl, Julius.
Block, Jack, Weiss, Daniel S., and Thorne, Avril. How relevant is a semantic similarity interpretation of personality ratings? 1055
Blyth, Daniel P. *See* Ajzen, Icek.
Boggiano, Ann K., and Ruble, Diane N. Competence and the overjustification effect: A developmental study 1462
Brickman, Philip, Becker, Lawrence J., and Castle, Sidney. Making trust easier and harder through two forms of sequential interaction 515
Brinberg, David. *See* Jaccard, James.
Brockner, Joel. The effects of self-esteem, success–failure, and self-consciousness on task performance 1732
Brockner, Joel. Self-esteem, self-consciousness, and task performance: Replications, extensions, and possible explanations 447
Brounstein, Paul, Ostrove, Nancy, and Mills, Judson. Divergence of private evaluations of alternatives prior to a choice 1957
Brown, Donald. *See* Leventhal, Howard.
Brunner, Lawrence J. Smiles can be back channels 728
Brunner, Lawrence J. *See* Duncan, Starkey, Jr.
Brunson, Bradford I. *See* Matthews, Karen A.
Buckert, Ursula, Meyer, Wulf-Uwe, and Schmalt, Heinz-Dieter. Effects of difficulty and diagnosticity on choice among tasks in relation to achievement motivation and perceived ability 1172
Burnkrant, Robert E. *See* Bagozzi, Richard P.
Burnstein, Eugene. *See* Sentis, Keith P.
Buss, Allan R. On the relationship between causes and reasons 1458
Bussey, Kay. *See* Perry, David G.

Cacioppo, John T. Effects of exogenous changes in heart rate on facilitation of thought and resistance to persuasion 489
Cacioppi, John T. *See* Petty, Richard E.
Cacioppo, John T., and Petty, Richard E. Attitudes and cognitive response: An electrophysiological approach 2181
Cacioppo, John T., and Petty, Richard E. Effects of message repetition and position on cognitive response, recall, and persuasion 97
Callahan-Levy, Charlene M., and Messé, Lawrence A. Sex differences in the allocation of pay 433
Capehart, George. *See* Seta, John J.
Carles, Elena M., and Carver, Charles S. Effects of person salience versus role salience on reward allocation in a dyad 2071
Carli, Linda L. *See* Berger, Seymour M.
Carlsmith, J. Merrill, and Anderson, Craig A. Ambient temperature and the occurrence of collective violence: A new analysis 337
Carver, Charles S. A cybernetic model of self-attention processes 1251
Carver, Charles S. *See* Carles, Elena M.
Carver, Charles S. *See* Scheier, Michael F.
Carver, Charles S., Blaney, Paul H., and Scheier, Michael F. Focus of attention, chronic expectancy, and responses to a feared stimulus 1186
Carver, Charles S., Blaney, Paul H., and Scheier, Michael F. Reassertion and giving up: The interactive role of self-directed attention and outcome expectancy 1859
Cassman, Ted. *See* Hastorf, Albert H.
Castle, Sidney. *See* Brickman, Philip.
Chaiken, Shelly. Communicator physical attractiveness and persuasion 1387
Chartier, George M. *See* Gregory, W. Larry.
Cialdini, Robert B. *See* Kenrick, Douglas T.
Clark, Margaret S., and Mills, Judson. Interpersonal attraction in exchange and communal relationships 12

Cohen, Sidney J. *See* Dermer, Marshall.
Colten, Mary Ellen. *See* Locksley, Anne.
Cooper, Harris M. Statistically combining independent studies: A meta-analysis of sex differences in conformity research 131
Cooper, Joel. *See* Baumeister, Roy F.
Cooper, Joel. *See* Goethals, George R.
Corkran, Louise. *See* Folger, Robert.
Costanzo, Philip R., and Woody, Erik Z. Externality as a function of obesity in children: Pervasive style or eating-specific attribute? 2286
Covington, Martin V., and Omelich, Carol L. Are causal attributions causal? A path analysis of the cognitive model of achievement motivation 1487
Crocker, Jennifer. *See* Taylor, Shelley E.
Cunningham, John D., Starr, Philip A., and Kanouse, David E. Self as actor, active observer, and passive observer: Implications for causal attributions 1146
Cunningham, Michael R. Weather, mood, and helping behavior: Quasi experiments with the sunshine Samaritan 1947

Dalal, Ajit K. *See* Singh, Ramadhar.
Dalto, Carol Ann. *See* Ajzen, Icek.
D'Andrade, Roy G. *See* Shweder, Richard A.
Davidson, Andrew R., and Jaccard, James J. Variables that moderate the attitude–behavior relation: Results of a longitudinal survey 1364
Davis, Deborah. *See* Ostrom, Thomas M.
Davis, Deborah, and Perkowitz, William T. Consequences of responsiveness in dyadic interaction: Effects of probability of response and proportion of content-related responses on interpersonal attraction 534
Davis, Mark H. *See* Bernstein, William M.
deCarufel, André. Factors affecting the evaluation of improvement: The role of normative standards and allocator resources 847
deCarufel, André, and Schopler, John. Evaluation of outcome improvement resulting from threats and appeals 662
DeJong, William. An examination of self-perception mediation of the foot-in-the-door effect 2221
Dempsey, Carolyn. *See* Foss, Robert D.
DePaulo, Bella M. *See* Rosenthal, Robert.
DePaulo, Bella M., and Rosenthal, Robert. Telling lies 1713
Dermer, Marshall, Cohen, Sidney J., Jacobsen, Elaine, and Anderson, Erling A. Evaluative judgments of aspects of life as a function of vicarious exposure to hedonic extremes 247
De Soto, Clinton B. *See* Taylor, Ralph B.
Diener, Ed. Deindividuation, self-awareness, and disinhibition 1160
Diener, Ed, and Srull, Thomas K. Self-awareness, psychological perspective, and self-reinforcement in relation to personal and social standards 413
Diener, Edward. *See* Beaman, Arthur L.
Doherty, William J., and Ryder, Robert G. Locus of control, interpersonal trust, and assertive behavior among newlyweds 2212
Doob, Anthony N., and Macdonald, Glenn E. Television viewing and fear of victimization: Is the relationship causal? 170
Downing, Leslie L. *See* Johnson, Robert D.
Dreben, Elizabeth K., Fiske, Susan T., and Hastie, Reid. The independence of evaluative and item information: Impression and recall order effects in behavior-based impression formation 1758
Drevno, Gregg E. *See* Benassi, Victor A.
Duncan, Starkey, Jr., Brunner, Lawrence J., and Fiske, Donald W. Strategy signals in face-to-face interaction 301
Duval, Shelley, Duval, Virginia Hensley, and Neely, Robert. Self-focus, felt responsibility, and helping behavior 1769
Duval, Virginia Hensley. *See* Duval, Shelley.

Ebbesen, Ebbe B., and Allen, Robert B. Cognitive processes in implicit personality trait inferences 471
Eisen, Susan V. Actor–observer differences in information inference and causal attribution 261
Elig, Timothy W., and Frieze, Irene Hanson. Measuring causal attributions for success and failure 621
Elliott, Gregory C. Some effects of deception and level of self-monitoring on planning and reacting to a self-presentation 1282
Ellsworth, Phoebe C. *See* Tourangeau, Roger.

Engquist, Gretchen. *See* Leventhal, Howard.
Entin, Elliot. *See* Natale, Michael.
Epstein, Seymour. The stability of behavior: I. On predicting most of the people much of the time 1097

Farkas, Arthur J., and Anderson, Norman H. Multidimensional input in equity theory 879
Farrow, Dana L. *See* Fodor, Eugene M.
Fazio, Russell H. Motives for social comparison: The construction–validation distinction 1683
Feather, N. T. Value correlates of conservatism 1617
Fehrenbach, Peter A., Miller, David J., and Thelen, Mark H. The importance of consistency of modeling behavior upon imitation: A comparison of single and multiple models 1412
Fenigstein, Allan. Does aggression cause a preference for viewing media violence? 2307
Fenigstein, Allan. Self-consciousness, self-attention, and social interaction 75
Fincham, F., and Jaspars, J. Attribution of responsibility to the self and other in children and adults 1589
Fisher, Jeffrey D. *See* Whitcher, Sheryle J.
Fiske, Donald W. *See* Duncan, Starkey, Jr.
Fiske, Susan T. *See* Dreben, Elizabeth K.
Fiske, Susan T. *See* Taylor, Shelley E.
Fodor, Eugene M., and Farrow, Dana L. The power motive as an influence on use of power 2091
Folger, Robert, Rosenfield, David, Grove, Janet, and Corkran, Louise. Effects of "voice" and peer opinions on responses to inequity 2253
Ford, Martin E. *See* Masters, John C.
Foss, Robert D., and Dempsey, Carolyn B. Blood donation and the foot-in-the-door technique: A limiting case 580
Freedenberg, Terry G. *See* Rajecki, D. W.
Frieze, Irene Hanson. *See* Elig, Timothy W.

Gara, Michael A., and Rosenberg, Seymour. The identification of persons as supersets and subsets in free-response personality descriptions 2161
Geva, Nehemia. *See* Lingle, John H.
Gibbons, Frederick X. *See* Scheier, Michael F.
Goethals, George R., Cooper, Joel, and Naficy, Anahita. Role of foreseen, foreseeable, and unforeseeable behavioral consequences in the arousal of cognitive dissonance 1179
Goethals, George R., and Zanna, Mark P. The role of social comparison in choice shifts 1469
Gollob, Harry F. A reply to Norman H. Anderson's critique of the subject–verb–object approach to social cognition 931
Goodenough, Donald R. *See* Witkin, Herman A.
Gough, Harrison G. A creative personality scale for the adjective check list 1398
Gray-Little, Bernadette, and Mark I. Appelbaum. Instrumentality effects in the assessment of racial differences in self-esteem 1221
Gregory, W. Larry, Chartier, George M., and Wright, Mark H. Learned helplessness and learned effectiveness: Effects of explicit response cues on individuals differing in personal control expectancies 1982
Grove, Janet. *See* Folger, Robert.
Grush, Joseph E., and Yehl, Janet G. Marital roles, sex differences, and interpersonal attraction 116
Gupta, Meenakshi. *See* Singh, Ramadhar.
Gur, Ruben C., and Sackeim, Harold A. Self-deception: A concept in search of a phenomenon 147

Hammersla, Kathy S. *See* Berger, Seymour M.
Harackiewicz, Judith M. The effects of reward contingency and performance feedback on intrinsic motivation 1352
Harkins, Stephen. *See* Latané, Bibb.
Harris, Bruce. *See* Klein, Kitty.
Hastie, Reid. *See* Dreben, Elizabeth K.
Hastie, Reid, and Kumar, Purohit Anand. Person memory: Personality traits as organizing principles in memory for behaviors 25
Hastorf, Albert H., Wildfogel, Jeffrey, and Cassman, Ted. Acknowledgment of handicap as a tactic in social interaction 1790
Harvey, John H., and Tucker, Jalie A. On problems with the cause–reason distinction in attribution theory 1441
Heller, Tamar. *See* Stein, R. Timothy.
Helmes, Edward. *See* Jackson, Douglas N.
Helmreich, Robert L. *See* Spence, Janet T.

Helmreich, Robert L., Spence, Janet T., and Holahan, Carole K. Psychological androgyny and sex role flexibility: A test of two hypotheses 1631
Holahan, Carole K. *See* Helmreich, Robert L.
Holahan, Carole K. *See* Spence, Janet T.
Holmes, David S. *See* McCaul, Kevin D.
Holt, Lewis E. *See* Watts, William A.
Hong, George K. *See* McCallum, Richard.
Huba, G. J. *See* Bentler, P. M.
Hull, Jay G., and Levy, Alan S. The organizational functions of the self: An alternative to the Duval and Wicklund model of self-awareness 756
Huston, Ted L. *See* Shotland, R. Lance.

Imber, Lois G. *See* Langer, Ellan J.
Insko, Chester A. *See* Tashakkori, Abbas.
Insko, Chester A., and Adewole, Akanbi. The role of assumed reciprocation of sentiment and assumed similarity in the production of attraction and agreement effects in *p-o-x* triads 790

Jaccard, James, Knox, Richard, and Brinberg, David. Prediction of behavior from beliefs: An extension and test of a subjective probability model 1239
Jaccard, James J. *See* Davidson, Andrew R.
Jackson, Douglas N. *See* Paunonen, Sampo V.
Jackson, Douglas N., and Helmes, Edward. Personality structure and the circumplex 2278
Jacobsen, Elaine. *See* Dermer, Marshall.
Jacques, Susan. *See* McCauley, Clark.
Jaffe, Joseph. *See* Natale, Michael.
Janoff-Bulman, Ronnie. Characterological versus behavioral self-blame: Inquiries into depression and rape 1798
Jaspars, J. *See* Fincham, F.
Johnson, Gregory A. *See* Kenrick, Douglas T.
Johnson, Kent G. *See* Pancer, S. Mark.
Johnson, Robert D., and Downing, Leslie L. Deindividuation and valence of cues: Effects on prosocial and antisocial behavior 1532
Johnston, Robert E. *See* Kraut, Robert E.
Jones, Edward E., Riggs, Janet Morgan, and Quattrone, George. Observer bias in the attitude attribution paradigm: Effect of time and information order 1230

Kabatoff, Randal A. *See* Pancer, S. Mark.
Kahle, Lynn R., and Berman, John J. Attitudes cause behaviors: A cross-lagged panel analysis .. 315
Kanouse, David E. *See* Cunningham, John D.
Karshmer, Judith F. *See* Berger, Seymour M.
Kassin, Saul M. Consensus information, prediction, and causal attribution: A review of the literature and issues 1966
Kassin, Saul M., and Wrightsman, Lawrence S. On the requirements of proof: The timing of judicial instruction and mock juror verdicts 1877
Kenny, David A., *See* Warner, Rebecca M.
Kenrick, Douglas T., Baumann, Donald J., and Cialdini, Robert B. A step in the socialization of altruism as hedonism: Effects of negative mood on children's generosity under public and private conditions 747
Kenrick, Douglas T., and Johnson, Gregory A. Interpersonal attraction in aversive environments: A problem for the classical conditioning paradigm? 572
Kleck, Robert E. *See* Snyder, Melvin L.
Klein, Helen M., and Willerman, Lee. Psychological masculinity and feminity and typical and maximal dominance expression in women 2059
Klein, Kitty, and Harris, Bruce. Disruptive effects of disconfirmed expectancies about crowding .. 769
Klentz, Bonnel. *See* Beaman, Arthur L.
Knox, Richard. *See* Jaccard, James.
Kobasa, Suzanne C. Stressful life events, personality, and health: An inquiry into hardiness 1
Kostin, Irene W. *See* Baker, A. Harvey.
Krantz, David S. *See* Mills, Richard T.
Krauss, Robert M. *See* Apple, William.
Kraut, Robert E., and Johnston, Robert E. Social and emotional messages of smiling: An ethological approach 1539
Kravitz, David A., and Wyer, Robert S., Jr. The effects of behavioral intentions and consequences on judgments of the actor and other: An S-V-O analysis 1561

Kruglanski, Arie W. Causal explanation, teleological explanation: On radical particularism in attribution theory 1447
Kuhl, Julius, and Blankenship, Virginia. Behavioral change in a constant environment: Shift to more difficult tasks with constant probability of success 551
Kuiper, N. S., and Rogers, T. B. Encoding of personal information: Self–other differences 499
Kumar, Purohit Anand. *See* Hastie, Reid.

Langer, Ellen J., and Imber, Lois G. When practice makes imperfect: Debilitating effects of overlearning 2014
Langer, Ellen, J., Rodin, Judith, Beck, Pearl, Weinman, Cynthia, and Spitzer, Lynn. Environmental determinants of memory improvement in late adulthood 2003
Larson, Rita W. *See* Messé, Lawrence A.
Latané, Bibb, Williams, Kipling, and Harkins, Stephen. Many hands make light the work: The causes and consequences of social loafing 822
Leippe, Michael R. *See* Lingle, John H.
Lepper, Mark R. *See* Lord, Charles G.
Lerman, David. *See* Weiner, Bernard.
Levenson, Robert W. *See* Notarius, Clifford I.
Leventhal, Howard, Brown, Donald, Shacham, Saya, and Engquist, Gretchen. Effects of preparatory information about sensations, threat of pain, and attention on cold pressor distress 688
Levinger, George. *See* Rands, Marylyn.
Levings, Diane E. *See* O'Neill, Patrick.
Levy, Alan S. *See* Hull, Jay G.
Lieb, Robert. *See* Taylor, Ralph B.
Lingle, John H., Geva, Nehemia, Ostrom, Thomas M., Leippe, Michael R., and Baumgardner, Michael H. Thematic effects of person judgments on impression organization 674
Lingle, John H., and Ostrom, Thomas M. Retrieval selectivity in memory-based impression judgments 180
Locksley, Anne, and Colten, Mary Ellen. Psychological androgyny: A case of mistaken identity? 1017
Lord, Charles G. *See* Bem, Daryl J.
Lord, Charles G., Ross, Lee, and Lepper, Mark R. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence 2098

Macdonald, Glenn E. *See* Doob, Anthony N.
Maki, Judith E., Thorngate, Warren B., and McClintock, Charles G. Prediction and perception of social motives 203
Markus, Gregory B. *See* Zanonc, R. B.
Markus, Hazel. *See* Zajonc, R. B.
Marshall, Gary D., and Zimbardo, Philip G. Affective consequences of inadequately explained physiological arousal 970
Martin, Lenny. *See* Seta, John J.
Maslach, Christina. Negative emotional biasing of unexplained arousal 953
Masters, John C., Barden, R. Christopher, and Ford, Martin E. Affective states, expressive behavior, and learning in children 380
Matthews, Karen A., and Brunson, Bradford I. Allocation of attention and the Type A coronary-prone behavior pattern 2081
McCallum, Richard, Rusbult, Caryl E., Hong, George K., Walden, Tedra A., and Schopler, John. Effects of resource availability and importance of behavior on the experience of crowding 1304
McCarthy, Patricia J. *See* Rajecki, D. W.
McCaul, Kevin D., Solomon, Sheldon, and Holmes, David S. Effects of paced respiration and expectations on physiological and psychological responses to threat 564
McCauley, Clark, and Jacques, Susan. The popularity of conspiracy theories of presidential assassination: A Bayesian analysis 637
McClintock, Charles G. *See* Maki, Judith E.
McCord, Joan. Some child-rearing antecedents of criminal behavior in adult men 1477
McMullen, Linda M. *See* Pancer, S. Mark.
Mellers, Barbara A. *See* Birnbaum, Michael H.
Mentzer, Steven J. *See* Snyder, Melvin L.
Messé, Lawrence A. *See* Callahan-Levy, Charlene M.
Messé, Lawrence A., and Sivacek, John M. Predictions of others' responses in a mixed-motive game: Self-justification or false consensus? 602
Messé, Lawrence A., Stollak, Gary E., Larson, Rita W., and Michaels, Gerald Y. Interpersonal consequences of person perception processes in two social contexts 369
Meyer, Wulf-Uwe. *See* Buckert, Ursula.

Michaels, Gerald Y. *See* Messé, Lawrence A.
Miller, Arthur G., Baer, Robert, and Schonberg, Peter. The bias phenomenon in attitude attribution: Actor and observer perspectives 1421
Miller, David J. *See* Fehrenbach, Peter A.
Miller, Frederick D. *See* Smith, Eliot R.
Mills, Judson. *See* Brounstein, Paul.
Mills, Judson. *See* Clark, Margaret S.
Mills, Richard T., and Krantz, David S. Information, choice, and reactions to stress: A field experiment in a blood bank with laboratory analogue 608
Mischel, Walter. *See* Yates, Brian T.
Mishara, Brian L. *See* Baker, A. Harvey.
Moreland, Richard L., and Zajonc, Robert B. Exposure effects may not depend on stimulus recognition 1085
Murnighan, J. Keith, and Szwajkowski, Eugene. Coalition bargaining in four games that include a veto player 1933
Mynatt, Clifford. *See* Arkkelin, Daniel.

Nafcy, Anahita. *See* Goethals, George R.
Nagy, Geraldine F. *See* Shanteau, James.
Natale, Michael, Entin, Elliot, and Jaffe, Joseph. Vocal interruptions in dyadic communication as a function of speech and social anxiety 865
Neely, Robert. *See* Duval, Shelley.
Nerenz, David R. *See* Rajecki, D. W.
Newcombe, Nora, and Arnkoff, Diane B. Effects of speech style and sex of speaker on person perception 1293
Newtson, Darren, and Rindner, Richard J. Variation in behavior perception and ability attribution 1847
Notarius, Clifford I., and Levenson, Robert W. Expressive tendencies and physiological response to stress 1204

Oakley, Thomas. *See* Arkkelin, Daniel.
Oltman, Philip K. *See* Witkin, Herman A.
O'Malley, Patrick M., and Bachman, Jerald G. Self-esteem and education: Sex and cohort comparisons among high school seniors 1153
Omelich, Carol L. *See* Covington, Martin V.
O'Neill, Patrick, and Levings, Diane E. Inducing biased scanning in a group setting to change attitudes toward bilingualism and capital punishment 1432
Ostrom, Thomas M. *See* Lingle, John H.
Ostrom, Thomas M., and Davis, Deborah. Idiosyncratic weighting of trait information in impression formation 2025
Ostrove, Nancy. *See* Brounstein, Paul.

Pancer, S. Mark, McMullen, Linda M., Kabatoff, Randal A., Johnson, Kent G., and Pond, Carole A. Conflict and avoidance in the helping situation 1406
Parker, Laurence. *See* Baker, A. Harvey.
Paunonen, Sampo V., and Jackson, Douglas N. Nonverbal trait inference 1645
Pedhazur, Elazar J., and Tetenbaum, Toby J. Bem Sex Role Inventory: A theoretical and methodological critique 996
Perkowitz, William T. *See* Davis, Deborah.
Perry, David G., and Bussey, Kay. The social learning theory of sex differences: Imitation is alive and well 1699
Petty, Richard E. *See* Cacioppo, John T.
Petty, Richard E., and Cacioppi, John T. Issue involvement can increase or decrease persuasion by enhancing message-relevant cognitive responses 1915
Pittman, Nancy L., and Pittman, Thane S. Effects of amount of helplessness training and internal–external locus of control on mood and performance 39
Pittman, Thane S. *See* Pittman, Nancy L.
Pond, Carole A. *See* Pancer, S. Mark.

Quattrone, George. *See* Jones, Edward E.

Rajecki, D. W., Nerenz, David R., Freedenberg, Terry G., and McCarthy, Patricia J. Components of aggression in chickens and conceptualizations of aggression in general 1902
Rands, Marylyn, and Levinger, George. Implicit theories of relationship: An intergenerational study 645

Richardson, Deborah Capasso, Bernstein, Sandy, and Taylor, Stuart P. The effect of situational contingencies on female retaliative behavior 2044
Riggs. Janet Morgan. *See* Jones, Edward E.
Rindner, Richard J. *See* Newtson, Darren.
Rodin, Judith. *See* Langer, Ellen J.
Rogers, T. B. *See* Kuiper, N. S.
Romer, Daniel. Internalization versus identification in the laboratory: A causal analysis of attitude change 2171
Rosenberg, Seymour. *See* Gara, Michael A.
Rosenfeld, David. *See* Folger, Robert.
Rosenthal, Robert. *See* DePaulo, Bella M.
Rosenthal, Robert, and DePaulo, Bella M. Sex differences in eavesdropping on nonverbal cues 273
Ross, Lee. *See* Lord, Charles G.
Ross, Michael, and Sicoly, Fiore. Egocentric biases in availability and attribution 322
Royce, Joseph R. Toward a viable theory of individual differences 1927
Ruble, Diane N. *See* Boggiano, Ann K.
Rusbult, Caryl E. *See* McCallum, Richard.
Russell, Dan. *See* Weiner, Bernard.
Russell, James A. Affective space is bipolar 345
Ryder, Robert G. *See* Doherty, William J.

Sackeim, Harold A. *See* Gur, Ruben C.
Sanchez, M. Estela. *See* Berger, Seymour M.
Schachter, Stanley, and Singer, Jerome E. Comments on the Maslach and Marshall-Zimbardo experiments 989
Scheier, Michael F. *See* Carver, Charles S.
Scheier, Michael F., Carver, Charles S., and Gibbons, Frederick X. Self-directed attention, awareness of bodily states, and suggestibility 1576
Schmalt, Heinz-Dieter. *See* Buckert, Ursula.
Schoggen, Phil. Roger G. Barker and behavior settings: A commentary 2158
Schonberg, Peter. *See* Miller, Arthur G.
Schopler, John. *See* deCarufel, André.
Schopler, John. *See* McCallum, Richard.
Sentis, Keith P., and Burnstein, Eugene. Remembering schema-consistent information: Effects of a balance schema on recognition memory 2200
Seta, John J., Martin, Lenny, and Capehart, George. Effects of contrast and generalization on the attitude similarity – attraction relationship 462
Shacham, Saya. *See* Leventhal, Howard.
Shanteau, James, and Nagy, Geraldine F. Probability of acceptance in dating choice 522
Shotland, R. Lance, and Huston, Ted L. Emergencies: What are they and do they influence bystanders to intervene? 1822
Shweder, Richard A., and D'Andrade, Roy G. Accurate reflection or systematic distortion? A reply to Block, Weiss, and Thorne 1075
Sicoly, Fiore. *See* Ross, Michael.
Simonton, Dean Keith, Multiple discovery and invention: Zeitgeist, genius, or chance? 1603
Singer, Jerome E. *See* Schachter, Stanley.
Singh, Ramadhar, Gupta, Meenakshi, and Dalal, Ajit K. Cultural difference in attribution of performance: An integration-theoretical analysis 1342
Sivacek, John M. *See* Messé, Lawrence A.
Skib, Bryan A. *See* Baumeister, Roy F.
Smith, Eliot R., and Miller, Frederick D. Attributional information processing: A response time model of causal subtraction 1723
Smith, Eliot R., and Miller, Frederick D. Salience and the cognitive mediation of attribution 2240
Snyder, Melvin L., Kleck, Robert E., Strenta, Angelo, and Mentzer, Steven J. Avoidance of the handicapped: An attributional ambiguity analysis 2297
Solomon, Sheldon. *See* McCaul, Kevin D.
Spence, Janet T. *See* Helmreich, Robert L.
Spence, Janet T., and Helmreich, Robert L. The many faces of androgyny: A reply to Locksley and Colten 1032
Spence, Janet T., Helmreich, Robert L., and Holahan, Carole K. Negative and positive components of psychological masculinity and femininity and their relationships to self-reports of neurotic and acting out behaviors 1673
Spitzer, Lynn. *See* Langer, Ellen J.
Sprinzen, Merle. *See* Taylor, Shelley E.

Srull, Thomas K. *See* Diener, Ed.
Srull, Thomas K., and Wyer, Robert S., Jr. The role of category accessibility in the interpretation of information about persons: Some determinants and implications 1660
Starr, Philip A. *See* Cunningham, John D.
Stegner, Steven E. *See* Birnbaum, Michael H.
Stein, R. Timothy, and Heller, Tamar. An empirical analysis of the correlations between leadership status and participation rates reported in the literature 1993
Stephan, Walter G. *See* Bernstein, William M.
Stollak, Gary E. *See* Messé, Lawrence A.
Storms, Michael D. Sex role identity and its relationships to sex role attributes and sex role stereotypes 1779
Stoto, Michael. *See* Warner, Rebecca M.
Streeter, Lynn A. *See* Apple, William.
Strenta, Angelo. *See* Snyder, Melvin L.
Svanum, Soren. *See* Beaman, Arthur L.
Sweeney, Paul D. *See* Benassi, Victor A.
Szwajkowski, Eugene. *See* Murnighan, J. Keith.

Tashakkori, Abbas, and Insko, Chester A. Interpersonal attraction and the polarity of similar attitudes: A test of three balance models 2262
Taylor, Marylee C. Race, sex, and the expression of self-fulfilling prophecies in a laboratory teaching situation 897
Taylor, Ralph B., De Soto, Clinton B., and Lieb, Robert. Sharing secrets: Disclosure and discretion in dyads and triads 1196
Taylor, Shelley E., Crocker, Jennifer, Fiske, Susan T., Sprinzen, Merle, and Winkler, Joachim D. The generalizability of salience effects 357
Taylor, Stuart P. *See* Richardson, Deborah Capasso.
Tetenbaum, Toby J. *See* Pedhazur, Elazar J.
Tetlock, Philip E. Identifying victims of groupthink from public statements of decision makers 1314
Thelen, Mark H. *See* Fehrenbach, Peter A.
Thorne, Avril. *See* Block, Jack.
Thorngate, Warren B. *See* Maki, Judith E.
Tourangeau, Roger, and Ellsworth, Phoebe C. The role of facial response in the experience of emotion 1519
Trope, Yaacov. Uncertainty-reducing properties of achievement tasks 1505
Tucker, Jalie A. *See* Harvey, John H.

Van Egeren, Lawrence F. Cardiovascular changes during social competition in a mixed-motive game 858

Waldren, Tedra A. *See* McCallum, Richard.
Wall, James A., Jr. The effects of mediator rewards and suggestions upon negotiations 1554
Warner, Rebecca M., Kenny, David A., and Stoto, Michael. A new round robin analysis of variance for social interaction data 1742
Watts, William A., and Holt, Lewis E. Persistence of opinion change induced under conditions of forewarning and distraction 778
Weary, Gifford. Self-serving attributional biases: Perceptual or response distortions? 1418
Weiner, Bernard, Russell, Dan, and Lerman, David. The cognition–emotion process in achievement-related contexts 1211
Weinman, Cynthia. *See* Langer, Ellen J.
Weiss, Daniel S. The effects of systematic variations in information on judges' descriptions of personality 2121
Weiss, Daniel S. *See* Block, Jack.
Whitcher, Sheryle J., and Fisher, Jeffrey D. Multidimensional reaction to therapeutic touch in a hospital setting 87
White, Leonard A. Erotica and aggression: The influence of sexual arousal, positive affect, and negative affect on aggressive behavior 591
Wiggins, Jerry S. A Psychological taxonomy of trait-descriptive terms: The interpersonal domain 395
Wildfogel, Jeffrey. *See* Hastorf, Albert H.
Willerman, Lee. *See* Klein, Helen M.
Williams, Kipling. *See* Latané, Bibb.
Wilson, William Raft. Feeling more than we can know: Exposure effects without learning 811
Winkler, Joachim D. *See* Taylor, Shelley E.

Witkin, Herman A., Goodenough, Donald R., and Oltman, Philip K. Psychological differentiation: Current status 1127
Wollert, Richard W. Expectancy shifts and the expectancy confidence hypothesis 1888
Woody, Erik Z. *See* Costanzo, Philip R.
Wright, Mark H. *See* Gregory, W. Larry.
Wrightsman, Lawrence S. *See* Kassin, Saul M.
Wyer, Robert S., Jr. *See* Kravitz, David A.
Wyer, Robert S., Jr. *See* Srull, Thomas K.

Yates, Brian T., and Mischel, Walter. Young children's preferred attentional strategies for delaying gratification 286
Yehl, Janet G. *See* Grush, Joseph E.

Zajonc, R. B., Markus, Hazel, and Markus, Gregory B. The birth order puzzle 1325
Zajonc, Robert B. *See* Moreland, Richard L.
Zanna, Mark P. *See* Goethals, George R.
Zimbardo, Philip G. *See* Marshall, Gary D.

OTHER

Ad hoc reviewers 146, 300, 423, 810, 1096
Correction to Stephan and Rosenfield 47
Corrections to DePaulo, Rosenthal, Eisenstat, Rogers, and Finkelstein 1965
Corrections to Koretzky, Kohn, and Jeger 727
Corrections to Lynch and Cohen 233
Editor's foreword 930
Editors named for new *Journal of Personality and Social Psychology* 314, 469, 635, 809, 1249, 1439
New JPSP format delayed until April 1446, 1722, 1992, 2261
Instructions for submission of manuscripts 1486, 1768, 2013, 2199
Notice on author alterations 38, 194
Notice to authors of the *Journal of Personality and Social Psychology* 468

New Directions in Psychology

PERSONALITY THEORIES, *Fourth Edition*

Salvatore Maddi, *University of Chicago*

This thoroughly absorbing text compares and contrasts many theoretical viewpoints with a comprehensive, consistent, and original framework for the personality field. Pertinent issues are posed and research studies are examined for their value in resolving issues. **Teacher's Manual.** 1980

UNDERSTANDING SOCIAL PSYCHOLOGY, *Revised Edition*

Stephen Worchel, *University of Virginia* and
Joel Cooper, *Princeton University*

The traditional areas of social psychology are covered in both theory and research along with demonstrations of how social psychology can and cannot be used to understand "real world" incidents. Includes a new chapter on socialization. **Student Study Guide. Teacher's Manual.** 1979

PERSONALITY: Strategies and Issues, *Third Edition*

Robert M. Liebert, *State University of New York at Stony Brook* and
Michael D. Spiegler, *Providence College*

Widely respected and highly readable introduction to major strategies and issues in the psychological study of personality. New discussions of attribution theory and cognitive behavior modifications are included. **Instructor's Manual.** 1978

Examination copies for adoption consideration available on request. Please indicate course title and text presently used.

THE DORSEY PRESS

Homewood, Illinois 60430

Nebraska Symposium on Motivation, 1978
Human Emotion

Edited by
Herbert E. Howe, Jr., and Richard A. Dienstbier

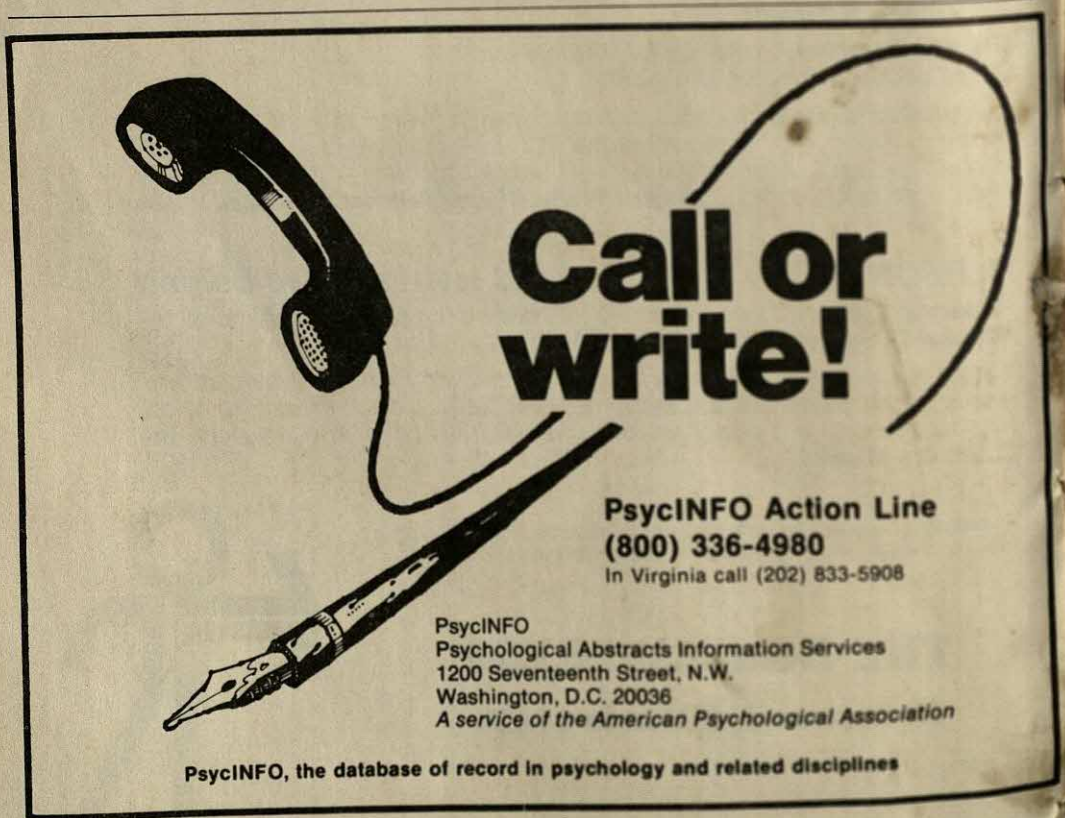
With this volume of the Nebraska Symposium on Motivation, the concept of motivation regains some of the importance that it held in the inception of the series. The authors study the interrelationship between motivation and emotion from various theoretical standpoints, utilizing a wide variety of research data to support their views. Includes "Anger" by James R. Averill, "Behavior Analysis of Motivational and Emotional Interactions in a Programmed Environment" by Joseph V. Brady and Henry H. Emurian, "Psychoneuroendocrine Approaches to the Study of Emotion as Related to Stress and Coping" by Marianne Frankenhaeuser, "Emotions as Motivations: An Evolutionary-Developmental Perspective" by C. E. Izard, "Script Theory: Differential Magnification of Affects" by Sylvan Tomkins, and "Emotion-Attribution Theory: Establishing Roots and Exploring Future Perspectives" by Richard A. Dienstbier.

xxvi, 321 pages.

Cloth **$17.95**
Paper **$ 8.95**

EMOTIONS

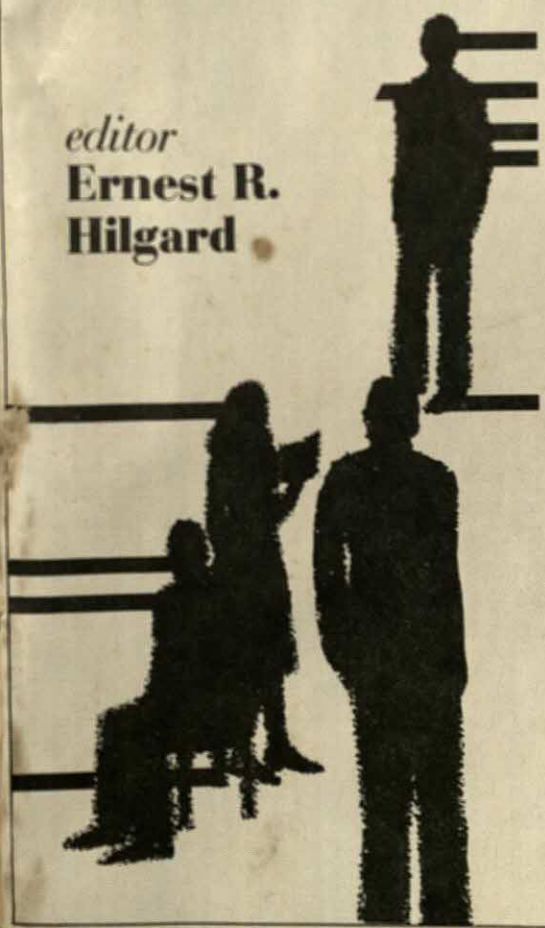
University of Nebraska Press
901 North 17th Street Lincoln 68588



American Psychology in Historical Perspective

1892-1977

editor
Ernest R. Hilgard



Now available in one book — 21 APA presidential addresses in their entirety from some of the most important names in American psychology. Included are classic pieces from James, Cattell, Dewey, Thorndike, Woodworth, and Watson. You'll also find Harlow's "The Nature of Love" and Miller's "Analytical Studies of Drive and Reward."

This book is truly a milestone for historical psychology. It provides both a fascinating chronology of the presidents of the American Psychological Association from 1892 to 1977 and a valuable collection of significant essays.

This new APA publication traces the development of American psychology over four broad periods in psychology's history: the first 25 years (1892-1916), the years of the two world wars (1917-1945), the 20 years after World War II (1946-1967), and the recent past (1968-1977).

For each of these periods, the editor, Ernest R. Hilgard, provides a brief summary of the thinking in psychology at the time, biographies of all the APA presidents with abstracts of their presidential addresses, and the selected presidential addresses in full.

American Psychology in Historical Perspective *may be ordered in hard cover for $18 or in soft cover for $15 by writing to: American Psychological Association, Order Dept., 1200 Seventeenth Street, NW, Washington, D.C. 20036*

Please include full remittance for orders of $25 or less.

The Psychology Major

Training and Employment Strategies

By **Paul J. Woods,** 1979, 384 pages, 6⅛" x 9¼". Price: $10.

Faced with an increasingly competitive job market, students and faculty alike will find *The Psychology Major* invaluable in the selection of courses, work, and field experience that employers seek.

Contributions are presented in four sections. "Preparation and Strategies for Employment" consists of "how-to-do-it" chapters of guidance and advice for those at the undergraduate level. The next section discusses "Job Prospects: Potential and Real."

The psychology graduate's role and job prospects in community agencies and paraprofessional positions is examined separately in "Training for Careers in Community Services, Mental Health, and Public Affairs." The last section, "The Undergraduate Major: Surveys, Models, and Problems" provides overviews of existing programs (including practica, field work, and internships) and areas of innovations and resistance to change in undergraduate training.

In the appendixes, the author presents valuable training and employment strategies such as an annotated bibliography on curriculum and careers, U.S. Civil Service Qualifications Standards for psychological jobs, a listing of Federal Job Information Centers and state personnel offices, and a description of strategies for admission.

This latest work from Paul J. Woods serves as a companion to his earlier *Career Opportunities for Psychologists* (APA, 1976).

To order, write to: American Psychological Association, Order Dept. • 1200 Seventeenth Street, NW Washington, D.C. 20036

Orders of $25 or less must be prepaid.

(Continued from front cover)

Avoidance of the Handicapped: An Attributional Ambiguity Analysis 2297
Melvin L. Snyder, Robert E. Kleck, Angelo Strenta, and Steven J. Mentzer

Does Aggression Cause a Preference for Viewing Media Violence? 2307
Allan Fenigstein

Instructions for Submission of Manuscripts 2199

New JPSP Format Delayed Until April 2261

This issue completes Volume 37 and contains the author index to the volume.

THE PSYCHOBIOLOGY OF THE DEPRESSIVE DISORDER

Implications for the Effects of Stress

Edited by RICHARD A. DEPUE

SECTION HEADINGS: Conceptual Overview of Stress and Disease. Distinctions Within the Topic. The Interaction of Stress and Biologic Variables in the Depressive Disorders. Psychosocial Aspects of Stress and the Depressive ... Models of Stress–Biology In... Theoretical and Research Strategies. Each Chapter Contains References.

1979, 464 pp., $29.50 ISBN: 0-12-211...

PROGRESS IN EXPERIMENTAL PERSONALITY RESEARCH

VOLUME 9

Edited by BRENDAN A. MAHER

CONTENTS: *R. A. Prentky*, CREATIVITY AND PSYCHOPATHOLOGY: A NEUROCOGNITIVE PERSPECTIVE. The Creative and the Insane. Definitions of Central Constructs. Empirical Research on Creativity. A Neurocognitive Integration. Neurophysiological Evidence. Conclusion. Summary. *P. J. Caplan*, BEYOND THE BOX SCORE: A BOUNDARY CONDITION FOR SEX DIFFERENCES IN AGGRESSION AND ACHIEVEMENT STRIVING. Introduction. Hazards of the Box Score Tally. Aggression. Achievement Striving. The Problem of No-Difference Studies. Conclusion. *W. M. Cox*, THE ALCOHOLIC PERSONALITY: A REVIEW OF THE EVIDENCE. Introduction. Empirical Evidence. Conclusions. *V. J. Konečni*, DETERMINANTS OF AESTHETIC PREFERENCE AND EFFECTS OF EXPOSURE TO AESTHETIC STIMULI: SOCIAL, EMOTIONAL, AND COGNITIVE FACTORS. Introduction. Effects of Social, Emotional, and Cognitive Factors on Choice between Melodies Differing in Complexity. Effects of Exposure to Melodies Differing in Complexity on Emotional States (Anger, Fear) and Aggressive Behavior. General Discussion. *Watt et al.*, TOWARD LONGITUDINAL CONCEPTIONS OF PSYCHIATRIC DISORDER. Rationals. Method. Presentation of Results. Discussion. Subject Index. Each Chapter Contains References.

1979, 312 pp., $26.00
ISBN: 0-12-541409-9

ADVANCES IN EXPERIMENTAL SOCIAL PSYCHOLOGY

Volume 12

Edited by LEONARD BERKOWITZ

CONTENTS: STUDIES IN SOCIAL COGNITION. *N. Cantor and W. Mischel*, Prototypes in Person Perception. *D. L. Hamilton*, A Cognitive-Attributional Analysis of Stereotyping. *M. Snyder*, Self-Monitoring Processes. SOCIAL INFLUENCES AND SOCIAL INTERACTIONS. *A. Baum and S. Valins*, Architectural Mediation of Residential Density and Control: Crowding and the Regulation of Social Contact. *J. W. Berry*, A Cultural Ecology of Social Behavior. *D. P. Farrington*, Experiments on Deviance with Special Reference to Dishonesty. *J. P. Murray and S. Kippax*, From the Early Window to the Late Night Show: International Trends in the Study of Television's Impact on Children and Adults. *J. P. Rushton*, Effects of Television and Film Material on the Prosocial Behavior of Viewers.

Each Chapter Contains References.

1979, 384 pp., $24.00 ISBN: 0-12-015212-6
Library Edition: $31.00 Microfiche: $17.00

A complete microfiche (4″ x 6″, 98 frames per fiche, 20x) is inserted in the back of each Library Edition. The Library Editions can be supplied on a Continuation Order basis, and the microfiche can be purchased separately.

Continuation Orders authorize us to ship and bill each volume in a series, or "Advances" type publication automatically, immediately upon publication. This order will remain in effect until cancelled. Specify the volume number or title with which your order is to begin.

Send payment with order and save postage and handling charge.
Prices are subject to change without notice.

ACADEMIC PRESS, INC.

A Subsidiary of Harcourt Brace Jovanovich, Publishers
111 FIFTH AVENUE, NEW YORK, N.Y. 10003
24-28 OVAL ROAD, LONDON NW1 7DX